University of BRISTOL

BILT
Bristol Institute for Learning and Teaching

BILT STUDENT RESEARCH JOURNAL

ISSUE 7

ISSN: 2976-8152

Contents

Select the page number to jump to the corresponding page.

Co-Editors-in-Chief's Forewords - page 7

Deputy Editors' Forewords - page 12

Business Studies and Economics - page 16

Working Across Generations: Perceptions of Young–Senior Co-Entrepreneurship
Author: Khushi Agarwarl - page 18

Does the Prosus–Just Eat Takeaway Acquisition Create Value?
Author: Huaijin Xing - page 42

Pink PP as Cultural Branding: How Luxury Brands Extract Community Meaning into Commercial Value
Author: Shenzimo Zhang - page 58

Research on Green Transformation, Financing Constraints, and Corporate Performance
Author: Wenxu Yu - page 75

Education and Digital Society - page 96

Student Feedback Literacy from a Critical Perspective: A Neurodivergence-ARirming Review
Author: Nathaniel Joyce - page 98

Digital Care and the Shift of Responsibility in AI Intimacy: Evidence from Character.AI and Love and DeepSpace
Author: Zixiao Fu - page 118

Modelling optimal networks: Countering the rational terrorist
Author: Henry Arbiter - page 131

Law - page 155

Is the Paris Agreement Enough? Human Rights and State Obligations in the Climate Crisis
Author: Sosefo Patelisio - page 157

Climate Refugees and International Legal Gaps: Deriving a Minimum Standard of Protection from Existing International Law
Authors: Thi Ngoc Mai Nguyen and Le Bao Ngan Pham - page 175

The Law of State Responsibility is increasingly irrelevant; it cannot respond to the increasing power and influence of Non-State Actors
Author: Sosefo Patelisio - page 193

Life Sciences - page 211

The ERect of Dominance Rank on Anemone Use in a High-Density Captive Population of the Common Clownfish *Amphiprion ocellaris*
Authors: Johnny Magnay and Fazeed Kotta - page 213

Investigating Natural Variation in Photosynthetic Temperature Responses Across *Arabidopsis Thaliana* Ecotypes and *Arabidopsis Lyrata*
Author: Danica Cruz -page 231

From Statistical Associations to Ecological Interactions: Rethinking Microbial Co-occurrence Networks in Soil Ecosystems
Author: Yuxuan Liao - page 255

Mathematical Models and Chemistry - page 276

Numerical modelling of Fresnel and Fraunhofer diRraction using Gaussian quadrature and Monte Carlo integration methods
Author: *Chenhao Zhouhan* - page 278

Temporal Kolmogorov-Arnold Networks (T-KAN) for High-Frequency Limit Order Book Forecasting: ERiciency, Interpretability, and Alpha-Decay
Author: Ahmad Makinde - page 303

Orbital Dynamics of the Earth-Moon-Probe System: A Numerical Study Using RK45
Author: Chenhao Zhouhan - page 317

Assessing the DiRerence in PM10, PM2.5 and NO2 Concentration Between Urban and Rural London and the Impact of Weather Conditions from 2018 to 2021
Author: Yun (Kerry) Shen - page 335

Molecular Biology and Translational Medicine - page 348

Beyond Transcription: The Mechanistic Functions of RNA in Double-Strand Break Repair
Author: Jack Magson - page 350

Do Relaxation and Stress management Interventions have Lasting Benefits for Hypertension: Subgroup Meta-analysis
Author: Laurence Ning - page 366

An Audit of WHO Surgical Checklist Compliance of University Hospitals Bristol and Weston NHS Foundation Trust with a Focus on the Bristol Royal Hospital for Children
Authors: Uliana Zybina, Kiru Uthayanan, Ashley Adjeroh - page 382

Towards Reliable Synthetic Circuit Design: Challenges, Modelling Approaches, and Future Directions -
Author: George Gould - page 396

Philosophy and Humanities - page 414

Interpreting and assessing the validity of Durkheim's Rules under conditions of structural injustice – a standpoint theory approach
Author: Sian Amekuedi - page 416

Susan Moller Okin on Women's Choices: Freedom of Choice or False Consciousness?
Author: Antonia Milla - page 432

More than just 'bad manners' and dropping 'f-bombs': The pharmakon approach and Duterte's ambivalent populist politics classified as a family resemblances
Author: Stefanija Kovacevic - page 442

Slipping Through the Nets of Language: Character and Circulation in *Jacob's Room*
Author: Lily Prole - page 456

How and why did the 'long historiography' of the Reformation in Bristol evolve between the sixteenth and the early twentieth centuries? -
Author: Nathan Roberts - page 469

Policy and Governance - page 482

Political Hate Speech as an electoral strategy in India -
Author: Kongkita Konwar - page 484

*P*aradigm Shifts and Policy Reform: The Role of New Ideas in UK Asylum Policy
Author: Foday Kabia - page 502

The Effect of Women's Political Empowerment on Environmental Performance
Author: Theadora Gallie - page 521

Psychology and Neuroscience - page 544

Serum Hydrogen Sulphide as a Biochemical Correlate of Major Depressive Disorder: A Cross-Sectional Case-Control Study from a Tertiary Care Centre in Sub-Himalayan India
Author: Atrim Das - page 546

Neuroticism, Conscientiousness, and Disease-Relevant Social Stress Processes
Author: Natalie Kah Xuan Koh - page 557

Functional EEG ERects of Short-Term Focused Attention Meditation: Do They Mirror Long-Term Structural Adaptations?
Author: Lucy Haun Eade - page 572

Serum Hydrogen Sulphide and Nitric Oxide Levels in Treated and Drug-Naive Patients with Schizophrenia: A Comparative Cross-Sectional Study
Authors: Atrim Das and Anamika Chakravarty - page 582

Serum Hydrogen Sulphide as a Biochemical Correlate of Major Depressive Disorder: A Cross-Sectional Case-Control Study from a Tertiary Care Centre in Sub-Himalayan India -
Author: Atrim Das - page 586

Co-Editor-in-Chief’s Forewords

It is a pleasure to welcome you to the seventh edition of the BILT Student Research Journal.

This journal represents far more than the collection of papers that follow. It reflects months of careful research, thoughtful discussion, constructive feedback, and collaboration between students who have dedicated their time to producing and refining original work. Every paper published here is the result of a shared commitment to asking meaningful questions and communicating ideas with clarity and rigour.

One of the highlights of this year was the BILT Student Research Conference. Bringing together students from across the University to present their work through talks, posters, and discussions, the conference demonstrated just how much enthusiasm there is for research beyond the classroom. It was particularly encouraging to see many first-time researchers share their work with confidence, engage with feedback, and learn from one another. The event reinforced something that sits at the heart of this journal: research thrives when ideas are shared, challenged, and discussed.

Producing this edition has also been a truly collaborative effort. Our editorial team, comprising Deputy Editors, Subject Editors, Copy Editors, and Peer Reviewers from a wide range of disciplines and academic backgrounds, worked diligently to ensure every submission received thoughtful and constructive consideration. For many, this was a first experience of academic editing and peer review; for others, it was an opportunity to build on existing skills while supporting fellow students. One of the most rewarding aspects of leading the journal has been hearing how this experience has helped participants develop confidence, communication, organisation, and teamwork - qualities that extend well beyond academic publishing.

One of the greatest challenges as Editor-in-Chief has been doing justice to the work submitted. Every manuscript represents a significant investment of time, curiosity, and perseverance, making editorial decisions both rewarding and difficult. Throughout the process, our aim has always been to ensure that each author received careful consideration and constructive feedback. Even for those whose work does not appear in this edition, I hope the experience has encouraged them to continue developing their research and sharing it with others.

Looking ahead, I hope the journal continues to reach an even wider community of student researchers. Outstanding work is taking place across the University every year, and publications such as this provide an important opportunity to recognise those efforts while encouraging others to take their first steps into research. My hope is that future editions continue to build on that spirit of collaboration and create opportunities for even more students to contribute.

Finally, I would like to thank Amy Palmer and Caroline Harvey for their guidance and encouragement throughout the year. My sincere thanks also go to our editorial board and peer reviewers for their dedication and professionalism, to our authors for entrusting us with their work, to the wider BILT team, including staff and Student Fellows, for their continued support, and to the wider University community for its commitment to fostering student research.

Ihope you enjoy reading Issue 7 of the BILT Student Research Journal and that the work presented here reflects the talent and dedication of everyone who has contributed to its publication.

Emma Segal
BILT Student Fellow and Co-Editor-in-Chief

Co-Editor-in-Chief’s Forewords

Every edition of this journal is, at its heart, a reflection of what students can achieve when they are given the space, support, and encouragement to pursue research with curiosity and purpose. The papers that follow are not simply assignments taken one step further; they are the result of months of questioning, refining, challenging assumptions, and revisiting ideas. They represent students who chose not to settle for an answer simply because it was the easiest one to reach, but instead committed themselves to understanding something more deeply. That spirit of curiosity is what makes the BILT Student Research Journal such a privilege to be part of.

This year's edition has been unlike any other. Owing to delays earlier in the year, the journal entered its main production phase later than in previous editions, following the BILT Student Research Conference in February. What might have seemed an impossible timeline quickly became an opportunity to see what a dedicated team could accomplish when united by a shared purpose. Looking back now, I am immensely proud of everything that followed.

The conference itself remains one of the defining moments of the year. It brought together students, academics, and postgraduate researchers from across the University to celebrate undergraduate research. We heard honest and inspiring reflections from our panel of academics and postgraduate researchers, watched nine students present their work to a live audience, explored eight research posters, shared ideas during a lively networking session, and discussed the Student Research Journal itself through an open panel with students. There was an unmistakable sense that research was not something reserved for established academics, but something students were actively shaping for themselves. It was in that room that I first realised what this community was capable of.

That same sense of purpose carried us through the months that followed. This edition contains 36 published papers, each one progressing through two rounds of editorial review and two rounds of peer review. Behind every article lies countless hours of reading, discussion, constructive criticism, revision, and careful editing. Every manuscript was challenged, strengthened, and refined through a process that relied not only on academic rigour but on the generosity and willingness of students

to invest their own time and expertise in helping someone else's work become stronger. That kind of collaboration cannot be measured by publication numbers alone. It is the foundation upon which this journal is built.

I joined the editorial team later than many of my colleagues, stepping into a project that was already well underway. It would have been easy to feel as though I had arrived too late to make much of a contribution. Instead, I found a team that welcomed new ideas, trusted one another's judgement, and approached every challenge with patience, professionalism, and a shared determination to do things well. Over the course of this year, I have seen first-hand how much can be achieved when talented people work towards a common goal. The compressed timeline made the work more demanding, but it also made every success feel that much more meaningful.

It has been a privilege to serve as Co-Editor-in-Chief alongside Emma, whose insight, dedication, and calm judgement have been invaluable throughout every stage of the publication process. I’m grateful to our three Deputy Editors, whose leadership, organisation, and unwavering commitment helped guide this edition from submission to publication. My sincere thanks also go to every Subject Editor, Copy Editor, and Peer Reviewer. The care and attention you invested in every manuscript, often alongside your own studies and commitments, reflects the very best of what academic collaboration can be.

I would also like to thank Amy Palmer and Caroline Harvey for their guidance, encouragement, and unwavering support throughout the year, as well as the wider BILT team and Student Fellows whose support behind the scenes made this journal possible. Finally, to every student author who entrusted us with your research, thank you. Publishing your work requires confidence, resilience, and a willingness to engage with constructive criticism. Your enthusiasm for discovery and your commitment to improving your work made this edition what it is.

If this year has reinforced one thing for me, it is that meaningful research is never the product of one person alone. Every paper in this journal reflects a conversation: between authors and reviewers, between students and mentors, and between questions and the evidence used to answer them. Those conversations are what

strengthen ideas, build confidence, and create a research community that continues to grow.

I hope that, as you read this seventh edition of the *BILT Student Research Journal*, you see not only the quality of the research it contains, but also the collaboration, perseverance, and dedication that made it possible. It has been an honour to be part of that process, and I could not be prouder of what this team has achieved.

Megan Stannett

BILT Student Fellow and Co-Editor-in-Chief

Deputy Editors Forewords

Faculty of Arts, Law and Social Sciences

As Deputy Editor for the Student Research Journal, I am delighted to introduce this year's Faculty of Arts, Law and Social Sciences sections. The articles featured here showcase the breadth, creativity, and intellectual curiosity of student research across a diverse range of disciplines. Drawing on a variety of methodologies and perspectives, the contributors engage with pressing cultural, historical, social, and political questions, while also demonstrating the value of critical and interdisciplinary inquiry. This section brings together research that explores the complexity of human experience through literature, history, languages, law, politics, sociology, and related fields. From close analyses of texts to investigations of social issues, these contributions highlight the many ways in which research in the arts, law, and social sciences deepens our understanding of the world and the communities we inhabit. The Student Research Journal continues to provide an important platform for students to share their work, develop their academic voices, and engage with a wider scholarly audience. I would like to thank all contributors, editors, reviewers, and members of the BILT team for their dedication and support throughout the publication process. I hope you enjoy reading these articles and find inspiration in the originality and enthusiasm of their authors.

Danlu Chen

Deputy-Editor-in-Chief of Arts, Law and Social Sciences

Meet Danlu Chen

Danlu Chen is a PhD candidate in Comparative Literatures and Cultures at the University of Bristol. Her doctoral research investigates sensory experiences and perception in nineteenth-century French and British short-form sensation fiction, drawing on sensory studies, comparative approaches, and cross-disciplinary methodologies. She has presented her research at national and international conferences and has experience in academic publishing. As ASSL Deputy Editor for the Student Research Journal, she is committed to fostering an inclusive and engaging scholarly community.

Faculty of Health and Life Sciences

Welcome to the Life Sciences sections of BILT's 7th Student Research Journal. I would like to start by thanking all our subject editors and peer reviewers. The process of getting to a published journal edition relies heavily on the considerable hours they put in reviewing and evaluating the submissions we receive. I have been incredibly impressed by their commitment and attention to detail this year and I do hope they will continue to work for the journal next year. Secondly, I want to thank all the students that made life sciences submissions. This year there is a great breadth of topics across biology, biochemistry, medicine, psychology, and neuroscience, reflecting a great diversity of interest and expertise that students at the University of Bristol have. It's not easy to have the confidence to put your work forwards for scrutiny and challenge, but I hope everyone who submitted work to us has found it a useful experience for developing their scientific and writing skills, and I am hoping that by reading this journal, you the reader might be convinced to submit your work for future editions.

Will Wheatley

Deputy-Editor-in-Chief of Health and Life Sciences

Meet Will Wheatley

William is a PhD student in the department of Psychology and Neuroscience. His background and training covers both disciplines, having studied Experimental Psychology for his Bachelor's degree, and Neuroscience in his Master's degree. Currently, in his PhD research, he studies how the brain regulates autonomic functions and behaviour during torpor – an adaptive physiological state adopted by a range of species in response to adverse environmental conditions. William was Deputy Subject Editor for Psychology in last year's student research journal and has greatly enjoyed taking on the role of Deputy Editor this year.

Faculty of Science and Engineering

As a Deputy Editor- in - chief I had an amazing experience of working with different students where played important roles of peer reviewer and subject editors in order for this journal to publish. Thank to BILT for providing all the students this opportunity for us to experience and expand of knowledge in the field of reseach and how it works. Science and technology helps in encouraging and transforming practical skills into advancing fields in everyday life. I hope this journal edition (6th Student research journal) inspires more students to do research and expand their horizon. Hope everyone enjoys the journal as I did.

Sanjana Sundararaman

Deputy-Editor-in-Chief of Science and Engineering

Meet Sanjana Sundararaman

Sanjana Sundararaman is currently a masters student in mathematics education at the University of Bristol. She has completed her undergraduate in BSC (hons) statistical data science. Having released research paper in the field of data science and machine learning. Her personal research interests range from the intersection of education , mathematics and data science. It was an enlightening experience for her to read and review for the field of science and engineering and gain great amount of experience and knowledge in this field. She hopes to continue her research journey in different fields of mathematics , education and data science.

Business and Economics

Meet the Subject Editor

Solomon Nderitu
Business and Economics

Solomon Nderitu is an MSc Economics student at the University of Bristol and a Research Assistant in the School of Economics. His research interests lie in firm dynamics, productivity, financial intermediation, and the role of resource allocation in shaping aggregate economic performance. He is particularly interested in how financial frictions influence the process of creative destruction through their effects on firm selection, business dynamism, and the allocation of capital and labour across firms and industries. Drawing on a background in accounting and financial analysis, he uses firm-level data to study how these forces shape long-run productivity growth and economic performance.

Working Across Generations: Perceptions of Young–Senior Co-Entrepreneurship

Author: Khushi Agarwal

ORCID: 0009-0005-3091-7551

Programme: MScManagement (Entrepreneurship & Innovation)

Affiliation: University of Bristol Business School

Email: to24450@alumni.bristol.ac.uk

Abstract

This study explores the feasibility of intergenerational co-entrepreneurship between young adults (20–34) and retired professionals (60+), globally. It is motivated by the increasing participation of both groups in entrepreneurship, and the growing interest in collaboration across generations, often presented as a complementary model combining youthful innovation with senior experience. However, limited research explores how these groups understand the realities of co-founding a venture. To address this, the study draws on stakeholder theory and treats both groups as internal stakeholders. It adopts a qualitative, exploratory approach using 20 semi-structured interviews, analysed thematically to examine participants' motivations, expectations, and concerns. The findings show that while intergenerational (young-senior) co-entrepreneurship offers significant potential, its complementarity is not automatic. Its feasibility depends on how partners distribute authority, build trust, align decision-making pace and risk orientation, and manage capacity constraints. Additionally, participants highlighted value beyond financial outcomes, including social purpose, knowledge exchange, and personal development. Overall, the study reframes intergenerational co-entrepreneurship as a relational, contingent model and extends stakeholder theory by highlighting internal power dynamics within entrepreneurial teams. The findings offer practical insights for entrepreneurs, support organisations, and policymakers designing sustainable intergenerational ventures.

Keywords: Intergenerational co-entrepreneurship, Young–senior co-entrepreneurship, Stakeholder theory, Thematic analysis, Venture co-founding.

The author would like to thank her dissertation supervisor, Stratos Ramoglou, for his guidance and support throughout this research. Gratitude is also extended to all participants who generously shared their time and insights.

Introduction

Entrepreneurship is increasingly pursued across generations. (Li et al., 2020; Murnieks and Cardon, 2023). For young adults, it is often linked to autonomy, innovation, and meaningful work, while shaped by insecure labour markets, unequal access to finance, and limited professional networks (Sitoula, 2015; Elkayaly and Fahim, 2021). Reflected in the AXA UK (2024) report that 60% of young people aspire to be self-employed by age 30, indicating strong and growing entrepreneurial ambition.

Entrepreneurship has also increased among seniors, often associating it with longer working lives, ageing populations, and the desire to remain active after retirement (Mouraviev and Avramenko, 2020; Management Editorial Team, 2023; Kanama et al., 2025). In the UK, around 991,000 individuals aged 60+ were self-employed in 2023 (Martin, 2024). While seniors bring experience, knowledge, and networks, they may also face ageism, health constraints, and digital challenges (Martin and Welsch, 2019; Dent, 2020; Boerner, Fritz and Frick, 2025).

The research in these domains highlight potential complementarity. Younger entrepreneurs are often seen as energetic, creative, and digitally skilled, whereas older adults are associated with experience, credibility, and strategic judgment (Perez-Encinas et al., 2021a; Perez-Encinas et al., 2021b; Watkins and Omilion-Hodges, 2025). This has led to growing interest in intergenerational collaboration. Studies and real-world cases suggest that age-diverse teams can create value. For example, John and Mark Cronin's enterprise demonstrates how a father's expertise can enhance his son's vision, providing a model for family-based intergenerational collaboration (Berger, 2018; Cronin and Cronin, 2021). Additional cases, such as OneClick.chat, founded by Dillon Myers and Alan Gibson, and Chip Conley's involvement at Airbnb, illustrate how younger and older partners may contribute distinct but valuable forms of knowledge (Hilgers, 2016; Hannon, 2019; Conley, 2018).

However, studies on intergenerational co-founding remains limited. Research on youth and senior entrepreneurship is largely separate, while intergenerational studies often focus on mentoring or workplace interaction rather than shared venture creation (Sitoula, 2015; Dent, 2020; Watkins and Omilion-Hodges, 2025). Even when discussed, it typically presented success cases or outcomes, with little focus on

partner coordination (Hannon, 2019; Perez-Encinas et al., 2021a). This is important, as theoretical complementarity may not translate easily into real collaboration.

Existing literature also points to potential tensions. For instance, younger founders may lack experience, stability, and legitimacy, while older partners may prefer slower communication, structured routines, and lower-risk strategies (Polat and Yılmaz, 2020; Elkayaly and Fahim, 2021; Mustafa, 2024; Manoj, 2025). Further complicated by stereotypes such as, older adults being seen as resistant to change and younger individuals as impulsive. Thus, influencing both internal relationships and external evaluations (Lyons and Kuron, 2014; Stanford Report, 2024; Boerner, Fritz and Frick, 2025). However, the co-founding perceptions still remains underexplored.

Therefore, the article examines how young adults (20–34) and retired professionals (60+) perceive the feasibility and value of co-founding a venture. The study uses stakeholder theory to examine young–senior co-entrepreneurship as an internal stakeholder relationship, rather than only as a combination of age-based traits or entrepreneurial motivations. Stakeholder theory has commonly been used to understand the wider network of individuals and groups connected to a firm, including customers, suppliers, employees, investors and communities, and to challenge the narrow view that business should prioritise shareholders alone (Freeman, 1984; Freeman et al., 2010). Building on this, Ramoglou, Zyglidopoulos and Papadopoulou (2023) argue that entrepreneurial opportunities are actualised not only through market demand, but also through the involvement of stakeholders at centre-stage and non- economic considerations. This lens is useful here because a young–senior venture does not become feasible simply because two individuals possess different resources or motivations; it depends on whether the co-founders, as internal stakeholders, can align their interests, expectations, values and forms of contribution within the venture. While motivation, human capital, entrepreneurial identity and life-course perspectives help explain why an individual may enter entrepreneurship (Elder, 1998; Shane, Locke and Collins, 2003; Fauchart and Gruber, 2011; Unger *et al*., 2011), they are less able to capture the co-founder dynamics. Adopting a global perspective, the study captures diverse contexts of ageing and entrepreneurship, offering insights for entrepreneurs, support organisations, and policymakers.

Accordingly, this research contributes to the underexplored field of intergenerational co-entrepreneurship, here termed young–senior co-entrepreneurship, by reconceptualising co-founders as internal stakeholders and examining whether such ventures create value beyond financial outcomes, including learning, social contribution, personal development and post-retirement engagement.

The remainder of this paper is structured as follows: Section 2 outlines the research design and methodology. Section 3 presents the key findings. Section 4 discusses these findings in relation to existing literature and stakeholder theory. Section 5 concludes with the main insights, contributions, and directions for future research.

Methodology

Research Design and Sample

This study adopts a qualitative, exploratory design to examine how younger and older adults perceive intergenerational co-entrepreneurship. A qualitative approach is appropriate as the research focuses on experiences, expectations, and relational dynamics that cannot be meaningfully captured through quantitative measures (Bell, Harley and Bryman, 2022).

Data was collected through 20 semi-structured interviews with two groups: young adults (20–34) and retired professionals (60+). The youth range reflects GEM's 18–34 category (Schøtt, Kew and Cheraghi, 2015), excluding those under 20 as decision- making maturity stabilises after this stage (Cauffman and Steinberg, 2000). While this strengthens the reliability of responses, it excludes younger aspirants whose views may differ. Similarly, the 60+ group captures the post-retirement stage (Statista, 2023; see Appendix 1 for the Statista graph), avoiding the tendency to merge seniors with mid-career adults, which can obscure distinct motivations (Kautonen, 2013).

Participants were selected using purposive sampling to include individuals with exposure to entrepreneurship and intergenerational interaction. This ensured responses were grounded in experience rather than abstract opinion. The sample included participants from India, the United Kingdom, the United States, China, Germany, Thailand, and Georgia, offering diverse perspectives. Details of

participants, including anonymised codes, age, country, sampling method, and interview length, are presented in Table 1 to ensure transparency while maintaining confidentiality.

Table 1 Participant Identifier Table

Younger Adult Participants (Aged 20–34)				
Code Name	Age	Country of Origin	Sampling Technique	Interview Length
Y1	25	India	Self - selection	52 minutes
Y2	24	Germany	Self - selection	32 minutes
Y3	23	India	Purposive	40 minutes
Y4	26	Thailand	Purposive	37 minutes
Y5	24	United Kingdom	Purposive	50 minutes
Y6	26	Georgia	Purposive	52 minutes
Y7	21	India	Purposive	32 minutes
Y8	23	Thailand	Purposive	50 minutes
Y9	24	China	Purposive	31 minutes
Y10	33	India	Purposive	58 minutes
Retired Professional Participants (Aged 60 and above)				
Code Name	Age	Country of Origin	Sampling Technique	Interview Length
R1	60	United Kingdom	Self - selection	45 minutes
R2	63	United Kingdom	Self - selection	36 minutes
R3	60	India	Purposive	31 minutes
R4	69	India	Purposive	30 minutes
R5	77	India	Purposive	35 minutes
R6	72	India	Purposive	40 minutes
R7	61	India	Purposive	32 minutes
R8	62	United States	Self - selection + Quota	44 minutes
R9	62	United Kingdom	Purposive + Quota	45 minutes
R10	67	India	Purposive	30 minutes

Data Collection and Analysis

Data was collected through semi-structured interviews conducted online or in person, depending on availability. This allowed consistency while giving participants flexibility to elaborate. Interviews lasted 30–60 minutes, were recorded with consent, and transcribed for analysis. Conducted between July and August 2025, the study adopts a cross-sectional design to capture current perceptions rather than changes over time (Saunders, Lewis and Thornhill, 2023).

The interview guide focused on themes such as trust, role expectations, decision-making, and perceived value in intergenerational co-entrepreneurship. Open-ended

questions encouraged reflection, with follow-up prompts used where necessary (see Appendix 3).

Data was analysed using thematic analysis following Braun and Clarke (2006). Each transcript was read several times to build familiarity before being manually analysed in Microsoft Word by highlighting key excerpts and adding initial codes via comments, which were then grouped into sub-themes and refined into broader themes, as shown in Appendix 2. Coding was mainly inductive, allowing patterns to emerge from interview responses, and was guided by concepts like value co-creation and stakeholder dynamics (Freeman et al., 2010), which served as sensitising lenses to guide interpretation without restriction.

Ethical procedures were followed throughout. Participation was voluntary, informed consent was obtained, and all data was anonymised.

This methodology has limitations. Purposive sampling and a small sample limit generalisability and may introduce bias. Although participants came from multiple countries, cultural differences were not examined in depth. Translation from Hindi to English, single-researcher analysis and uneven gender representation may also have shaped the findings.

Findings

This section presents findings from interviews with young adults (20–34) and retired professionals (60+). Participants are coded as R (retired professionals) and Y (young adults).

Understanding and Self-Perception of Entrepreneurship

Participants described entrepreneurship differently across generations. Retirees framed it practically, linking it to self-employment and financial security:

“It simply means being self-employed and it could be from small to big level” — R3

Whereas, younger participants emphasised ownership, responsibility, and value creation:

“You’re also responsible for everything… if you create an impact, if you create value, you might as well create it for yourself.” – Y2

Differences also appeared in perceived capacity. Retirees highlighted limits related to health and lifestyle, while younger individuals stressed stamina:

“We can work for 15 or 16 hours at a stretch… we are at the peak of our energy,” — Y1

However, some younger participants also recognised constraints around experience and financial readiness. Motivations reflected these differences: younger participants focused on independence and fulfilment, while retirees emphasised continued income and applying prior experience. Both groups saw preparation as essential, though younger participants linked it to work experience and retirees to clear planning.

Motivators for Intergenerational Collaboration

Participants consistently highlighted the value of combining generational strengths. Retirees emphasised the energy and ideas of young people, while younger participants valued the experience and strategic perspective of older partners. As one noted,

“Two different brains from two different cultures, and age diversity… could contribute more rather than… the same age group, same experience.” – R6

while another reflected,

“Idon’t see myself having that level of vision… I’m good with planning.” – Y1

Collaboration was also viewed as a space for mutual learning. Younger participants expected guidance alongside opportunities to contribute digital skills, while older participants described the exchange as reciprocal:

“Both sides have learned from me. I’ve learned also from them… we learned together.” — R2

Shared values further shaped collaboration. Participants stressed that alignment in purpose, particularly beyond financial outcomes, was central to sustaining partnerships:

“You have to create job opportunities… make contributions to society.” — Y9

Intergenerational ventures were also seenas generating broader societal benefits, including knowledge exchange, innovation,and continued engagement for retirees.

Barriers and Tensions in Collaboration

Despite recognising benefits, participantsidentified tensions around working style, risk and capacity. Retirees often viewedyounger partners as impatient, while younger participants perceived older collaborators as slower:

“Younger generation… work fast… but older people… thinkmultiple times before doingthings…youneed to findthe middleground.” — Y4

Risk orientation reflected this divide. Older participants emphasised caution and risk aversion, whereas younger individuals appeared more open to experimentation and failure, supporting research linking youth entrepreneurship with greater risk-taking and older entrepreneurship with more cautious strategies (Perez-Encinas et al., 2021a). As one retiree observed:

“When you get older, you get morecautious, more risk averse. And the thing about youngsters is that they are just like,‘I’m just going to do it. If it goes wrong, we’ll do something else.” — R2

Stereotypes shaped expectations but did not eliminate willingness to collaborate. Practical constraints, particularly time and energy, also emerged:

“you can maybe work 18 hours a day… sometimes in an older generation… you can’t expect that.” —Y10

These findings suggest that capacity is not only self-perceived but also a structural constraint influencing collaboration.

Collaboration Dynamics

Participants described collaboration around roles, decision-making, and communication. Role division often followed generational lines, with younger participants focusing on execution and older partners on strategy and finance:

“I would like to do the marketing part by myself… I would pass the finance management and the sourcing to the older partner” — Y2

“I would trust the older individual with the decision making… I’ll probably take the execution part.” — Y10

Older participants similarly emphasised planning and oversight:

“Ideation… strategic planning… operations… I can take these, and I would pass social media marketing to the younger partner.” — R8

Decision-making preferences differed, with retirees favouring structured processes and younger participants prioritising efficiency. Trust and communication also emerged as central.

“Trust. … trust is something which builds a core foundation. … communication, it needs to be crystal clear, communication gaps can dent your business in the long run.” —Y1

While some noted that age differences could create misunderstandings, others emphasised that trust outweighed age:

“I wouldn’t consider age a factor at all,really; what matters is suitability and trustworthiness, not their date of birth.” — R3

Overall, collaboration dynamics depended on role clarity, shared authority and transparent communication.

Enablers and Conditions for Successful Ventures

Participants identified enabling conditions related to skills, mindset and environment. Six younger participants emphasised interpersonal skills such as communication, empathy and patience:

“Communication… andempathy… you understand where they come from, and you don’ttakesomethingasapersonal offence.” — Y6

Retired participants stressed structure, clarity, and accountability:

“writing things down and being able to remember them and thendiscuss them in the group and say, we said this, and we went wrong… CRMs… recording our conversation with clients… what actually was said rather than what you remember wassaid.” — R2

Mentorship and reciprocal learning were also valued. Additionally, R8 critiqued the education system, arguing that *“there's too much of an emphasis on business studies… critical thinking skills have become lost,”* suggesting that interpersonal and reflective skills are essential for sustaining such ventures.

Beyond individual skills, participants highlighted supportive environments:

“doesn't matter what the age group is… it totally depends on the workspace environment,” — Y7

While R8 argued *“companies don't grow without diversity of thought,”* underlining that organisational climate and value alignment matter more than age alone.

Overall, successful intergenerational ventures rely on a balance of relational skills, structured practices, and supportive environments.

Therefore, the findings suggest that while intergenerational collaboration offers clear complementary strengths, its effectiveness depends on how differences in roles, pace, and expectations are managed. The following section interprets these findings in relation to existing literature and stakeholder theory.

Discussion

The findingschallenge the assumption that intergenerational entrepreneurship operates asastraightforward combination of complementarities and barriers. Instead, collaborationis shaped by ongoing negotiation across roles, trust, tempo, and capacity.

Role Differentiation and Uneven Authority

The issue in intergenerational co-founding is not role differentiation itself, since venture requires division of work, but the basis on which roles and authority are assigned. In the findings, roles were often divided through age-based assumptions, with younger partners closer to execution and older partners closer to strategy and decision-making. While this reflects the complementary strengths noted in earlier research (Ensari, 2017; Perez-Encinas et al., 2021a), it also shows how complementarity becomes unequal when age is treated as a shortcut for competence.

Such arrangements may shift the partnership from co-founding towards a mentor–mentee relationship. Younger partners may feel obliged to follow senior decisions because of experience, hierarchy, or respect, while older partners may become concentrated points of authority. Over time, this could create frustration, weaken open disagreement, or make decision-making dependent on one side's judgement. Ramoglou, Zyglidopoulos and Papadopoulou (2023) argue that stakeholders are not passive recipients of entrepreneurial action. Therefore, both partners can be understood as internal stakeholders whose cooperation, interests and expectations shape whether the venture functions effectively. Therefore, stable young–senior co-entrepreneurship may require roles to be based on expertise, flexibility to revise responsibilities, and checks and balances that preserve the partnership.

Trust as a Central Coordinating Mechanism

Trust determines whether stereotypes in young–senior co-entrepreneurship remain barriers or become manageable differences. Both groups expressed assumptions about one another, with younger partners associated with speed and impatience, and older partners with caution or rigidity. However, these perceptions did not necessarily prevent collaboration. Where trust existed, participants were more willing to interpret differences as learning rather than as a reason to avoid partnership.

Co-founding requires delegation of both work and authority. Without trust, partners may hesitate to rely on each other's judgment, leading to repeated checking, overlapping effort, slower coordination and wasted resources. Trust therefore functions as relational governance, helping partners cooperate when every decision, responsibility or conflict cannot be controlled through formal rules alone (Mayer, Davis and Schoorman, 1995; Poppo and Zenger, 2002). From a stakeholder theory perspective, this reflects the need to manage relationships for shared value rather than allowing one group's assumptions to dominate the other (Freeman et al., 2010).

Prior familiarity may reduce stereotypes because some trust may already exist before the venture begins. Where partners are unfamiliar, clear agreements, defined responsibilities, and transparent decision processes can reduce uncertainty without replacing interpersonal trust. Therefore, trust becomes more than a personal preference; it is a practical condition for delegation, coordination and long-term partnership stability.

Temporal Differences as a Structural Constraint

Differences in decision-making pace show that young–senior co-founding depends not only on shared goals, but also on compatible expectations about how quickly business decisions should be made and acted upon. Younger participants often associated venture work with speed and momentum, while older participants emphasised deliberation and risk checking. This should not be treated as a fixed age rule, but as a practical difference shaped by life-stage experiences and working habits.

Pace matters because it shapes venture functioning. A slower decision process may seem responsible to one partner but frustrating to another, while quick action may appear entrepreneurial to one and careless to the other. Such differences can create

coordination friction, with younger partners seeing opportunities as lost through over-caution and older partners attributing mistakes to haste.

This extends stakeholder theory by showing that internal stakeholders bring expectations about the rhythm of action, not only resources, interests and values. Building on Ramoglou, Zyglidopoulos and Papadopoulou's (2023) non-economic considerations, working pace can be one condition that shapes opportunity actualisation. Young–senior ventures may therefore need early agreement on decision timelines, risk thresholds and independent authority, so speed and caution balance rather than become sources of blame.

Capacity Constraints and the Limits of Complementarity

Capacity is not an individual limitation in young–seniorco-entrepreneurship; it becomes a shared condition shaping what the partnership can sustain. Complementarity suggests that one partner's strength can offset the other's weakness, but this only works when differences in availability, energy, digital responsiveness and commitment are openly managed. Otherwise, the same differences that make the partnership attractive can produce misalignment in workload, communication and expectations.

This develops stakeholder theory by showing that internal stakeholders influence opportunity not only through their resources, but also through the constraints they bring into the venture. A partner's limited availability, slower responsiveness or different boundaries between work and personal life may quietly define the pace and scope of the business. Equally, the intergenerational learning exchange may create value, but only if the benefits outweigh the coordination effort required to maintain it. Building on Ramoglou, Zyglidopoulos and Papadopoulou's (2023) view that opportunity actualisation depends on stakeholder conditions, young–senior ventures therefore require early clarity on capacity, commitment and fallback arrangements. Complementarity is most likely to become productive only when differences can be absorbed without weakening the venture.

Implications

The central theoretical insight is that co-founders shape venture feasibility not only through the resources they bring in, but also through the constraints they introduce into the partnership. This extends stakeholder theory by showing that young-senior co-entrepreneurship depends on managing internal stakeholder conditions and relations, not complementarity alone.

Practically, this suggests that young and senior co-founders should assess the partnership, not only the business idea, before committing to a venture. This includes clarifying autonomy, decision rights, responsibilities and handling of disagreements. Furthermore, support organisations should move beyond age-basis matchmaking by helping potential partners test their fit through guided conversations, scenario-based exercises or short collaboration trials. Incubators and accelerators can support this through mixed-age founder programmes that could combine mentoring with shared decision-making practice. For policymakers, the findings suggest that targeted initiatives can make intergenerational entrepreneurship operationally feasible, rather than treating it only as a response to youth employment or ageing workforce challenges.

Limitations

As this study focuses on perceptions of young–senior co-entrepreneurship, its findings should be understood as identifying perceived conditions for feasibility rather than measuring venture outcomes. This does not reduce the value of the study, instead, it highlights the early-stage expectations, concerns and relational conditions that may shape whether such partnerships are attempted and sustained. Future research could build on this by examining active young-senior ventures over time.

Conclusion

This study examined how young adults and retired professionals perceive the feasibility of intergenerational co-entrepreneurship. Drawing on stakeholder theory and qualitative evidence from 20 interviews, the findings show that such collaboration offers clear potential but cannot be assumed to work simply because each generation brings different strengths. Thus, its viability depends on how co-founders organise their relationship.

The study further demonstrates that value in intergenerational ventures extends beyond financial outcomes to include social, emotional, and developmental benefits. By positioning co-founders as internal stakeholders, it reframes intergenerational entrepreneurship as a relational process in which entrepreneurial potential becomes viable only when expectations, contributions and responsibilities are made workable in practice.

These insights highlight practical implications for entrepreneurs, support organisations, and policymakers. Intergenerational ventures require structured support mechanisms, such as matchmaking platforms, training, and institutional guidance, to translate potential complementarities into sustainable collaboration and to address broader challenges related to ageing and youth employment.

However, the study has limitations. As this study focuses on perceptions rather than active ventures, its findings should be interpreted as perceived conditions for feasibility rather than measured venture outcomes. The small, purposively selected sample and uneven gender representation also limit generalisability and the diversity of perspectives captured.

Future research could build on this work through larger, more gender-balanced and culturally diverse samples, as well as longitudinal or observational studies examining how young–senior partnerships develop over time. Further studies could also explore how targeted support mechanisms enable sustainable intergenerational ventures.

References

AXA UK (2024) *AXA UK research shows 60% of young people want to be their own bossbytheageof30* . Available at: https://www.axa.co.uk/newsroom/media-releases/2024/axa-uk-research-shows-60-of-young-people-want-to-be-their-own-boss-by-the-age-of-30/ (Accessed: 20 July 2025).

Bell, E., Harley, B. and Bryman, A. (2022) *Business research methods*. Oxford university press.

Berger, S. (2018) *'How this 22-year-old with Down syndrome built a multimillion dollarbusinessoffhis love ofcrazysocks'* , CNBC, 22 October. Available at: https://www.cnbc.com/amp/2018/10/22/22-year-old-with-down-syndrome-founded-johns-crazy-socks.html (Accessed: 23 August 2025).

Boerner, L., Fritz, T. and Frick, B. (2025) 'Investing in people, not in products: how age, gender, ethnicity, and attractiveness of entrepreneurial teams influence the decision- making of angel investors in Germany', *Journal of Business Economics*, 95(1), pp. 37-74.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative research in psychology,* 3(2), pp. 77-101.

Cauffman, E. and Steinberg, L. (2000) '(Im) maturity of judgment in adolescence: Why adolescents may be less culpable than adults', *Behavioralsciences&thelaw*, 18(6), pp. 741-760.

Conley, C. (2018) *Wisdom at work: The making of a modern elder*. Crown Currency.

Cronin, M. and Cronin, J. (2021) *'Mission-driven podcast: Mark Cronin '80'*, Holy Cross Alumni Relations Podcast, 22 June. Available at: https://hcalumni.libsyn.com/21-mark-cronin-80 (Accessed: 23 August 2025).

Daraban, M. (2023) *'Why economic value creation is crucial for your organization'*, TopCFO, 5 June. Available at: https://www.topcfo.ro/why-economic-value-creation-is-crucial-for-your-organization/ (Accessed: 29 August 2025).

Dent, H. L. (2020) *Boomer Entrepreneurship: The New Retirement. D.B.A*., Jacksonville University, United States -- Florida [Online] Available at: https://bris.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/boomer-entrepreneurship-new-retirement/docview/2458705362/se-2?accountid=9730

Elder, G.H. Jr. (1998) 'The life course as developmental theory', *Child Development*, 69(1), pp. 1–12. doi: 10.2307/1132065.

Elkayaly, D. and Fahim, I. S. (2021) *'Challenges and Motivations for Youth EntrepreneurshipStart-Ups:Empiricalstudyfrom Egypt'* . Third Internation al Sustainability and Resilience Conference: Climate Change: IEEE, 339-341.

Ensari, M. S. (2017) 'A study on the differences of entrepreneurship potential among generations', *Research Journal of Business and Management*, 4(1), pp. 52-62.

Eurostat (2025) *'Work after retirement more common among self-employed'*, Eurostat News, 6 June. Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20250606-1 (Accessed: 30 July 2025).

Fauchart, E. and Gruber, M. (2011) 'Darwinians, communitarians, and missionaries: The role of founder identity in entrepreneurship', *Academy of Management Journal*, 54(5), pp. 935–957. doi: 10.5465/amj.2009.0211.

Forster-Holt, N. and Clark, P. G. (2024) 'I'm Glad I Met You: Ageism Interventions in an Entrepreneurship Course', *Journal of Management Education*, 48(6), pp. 1148-1184.

Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L. and De Colle, S. (2010) *'Stakeholder theory: The state of the art'*.

Hannon, K. (2019) *‘Thefamilybusinessmatchingolderandyoungerentrepreneurs’*, Forbes, 28 March. Available at: https://www.forbes.com/sites/nextavenue/2019/03/28/the-family-business-matching-older-and-younger-entrepreneur/ (Accessed: 23 August 2025).

Hilgers, L. (2016) *Chip Conley's influence at Airbnb.* Spaces Magazine*.* Available at: https://spacesmag.com/stories/people/chip-conleys-influenceatairbnb/#:~:text=When%20the%20company%20hired%20him,Neither%20felt%20right%20to%20Conley (Accessed: 7 July 2025).

Kanama, D., Ito, S., Muranaka, S. and Watanabe, T. (2025) 'Role of the grey entrepreneur in startups: an empirical study of the impact of age diversity on innovation performance', *Journal of Economic Structures*, 14(1), pp. 1-17.

Kautonen, T. (2013) *Senior entrepreneurship*: OECD publishing.

Kibler, E., Wainwright, T., Kautonen, T. and Blackburn, R. (2015) 'Can social exclusion against “older entrepreneurs” be managed?', *JournalofSmallBusiness Management*, 53, pp. 193-208.

Laplume, A. O., Sonpar, K. and Litz, R. A. (2008) 'Stakeholder Theory: Reviewing a Theory That Moves Us', *Journal of Management*, 34(6), pp. 1152-1189.

Li, C., Murad, M., Shahzad, F., Khan, M. A. S., Ashraf, S. F. and Dogbe, C. S. K. (2020) 'Entrepreneurial passion to entrepreneurial behavior: role of entrepreneurial alertness, entrepreneurial self-efficacy and proactive personality'*, Frontiers in psychology*, 11, pp.1611.

Lyons, S. and Kuron, L. (2014) 'Generational differences in the workplace: A review of the evidence and directions for future research', *Journal of organizational behavior*, 35(S1), pp.S139-S157.

Management Editorial Team (2023) 'Longer lifespans mean the need to rethink work and retirement', *NZ Business + Management*, 37(3), pp. Sp(7)-Sp(7).

Manoj, M. (2025) ‘Generational differences in communication preferences and their impact on employee engagement’, *InternationalJournalofResearchPublicationand Reviews*, 6(3), pp. 8605–8616. Available at: https://ijrpr.com/uploads/V6ISSUE3/IJRPR40909.pdf (Accessed: 20 August 2025).

Martin, D. (2024) ‘Self-employed people over 60 hits record high’, *EnterpriseNation*, 29 August. Available at: https://www.enterprisenation.com/learn-something/self-employed-people-over-60-hits-record-high/ (Accessed: 20 July 2025).

Martin, W. F. and Welsch, H. (2019) 'Wasted talent: Battling exclusionary forces among senior entrepreneurs', *The International Journal of Entrepreneurship and Innovation*, 20(2), pp. 130-143.

Mouraviev, N. and Avramenko, A. (2020) 'Senior Entrepreneurship: Ageing, Entrepreneurial Identity, Behaviour and Challenges', *EntrepreneurshipforDeprived Communities*: Emerald Publishing Limited, pp. 67-93.

Murnieks, C. Y. and Cardon, M. S. (2023). *Entrepreneurial Passion*. Oxford University Press.

Mustafa, N. (2024) 'STUDENTS' PERCEPTIONS ON CHALLENGES AND OPPORTUNITIES FOR STARTUP INNOVATION AND ENTREPRENEURSHIP IN AZERBAIJAN', *German International Journal of Modern Science / Deutsche Internationale Zeitschrift für Zeitgenössische Wissenschaft*, (93), pp. 17-25.

Perez-Encinas, A., Bueno, Y., Santos, B. and Nieto-Mejia, C. (2021a) 'Are There Differences and Complementarities between Senior and Young Entrepreneurs? An Intergenerational Perspective', *Sustainability*, 13(9), pp. 5202.

Perez-Encinas, A., de Pablo, I., Bueno, Y. and Santos, B. (2021b) 'Intergenerational Entrepreneurship to Foster Sustainable Development: A Methodological Training Proposal', *Sustainability*, 13(17), pp. 9654.

Polat, S. and Yılmaz, Y. (2020) 'Barriers to intergenerational learning: a case of a workplace in Turkey', *Leadership & Organization Development Journal*, 41(3), pp. 431-447.

Poppo, L. and Zenger, T. (2002) 'Do formal contracts and relational governance function as substitutes or complements?', *Strategic Management Journal*, 23(8), pp. 707–725.

Ramoglou, S., Zyglidopoulos, S. and Papadopoulou, F. (2023) 'Is there opportunity without stakeholders? A stakeholder theory critique and development of opportunity-actualisation', *Entrepreneurship Theory and Practice*, 47(1), pp. 113-141.

Saunders, M. N. K., Lewis, P. and Thornhill, A. (2023) *Research methods for business students.* Ninth edition edn. Harlow, England: Pearson, p. xxvii, 860 pages : illustrations (chiefly colour); 27 cm.

Schøtt, T., Kew, P. and Cheraghi, M. (2015) 'Future Potential. A GEM perspective on youth entrepreneurship 2015', *Global Entrepreneurship Monitor* (GEM).

Scully, C. (2024) '8 ways Gen Z will change the workforce', *Stanford Report*, 14 February. Available at: https://news.stanford.edu/stories/2024/02/8-things-expect-gen-z-coworker (Accessed: 25 August 2025).

Shane, S., Locke, E.A. and Collins, C.J. (2003) 'Entrepreneurial motivation', *Human Resource Management Review*, 13(2), pp. 257–279. doi: 10.1016/S1053-4822(03)00017-2.

Sitoula, T. (2015) *'Challenges and prospects of youth entrepreneurship in Kathmandu'.*

Statista (2023) 'Current and effective retirement age worldwide in 2020', *Statista*, October. Available at: https://www statistacom.bris.idm.oclc.org/statistics/ 268824/retirement-age-in-international-comparison/ (Accessed: 20 July 2025).

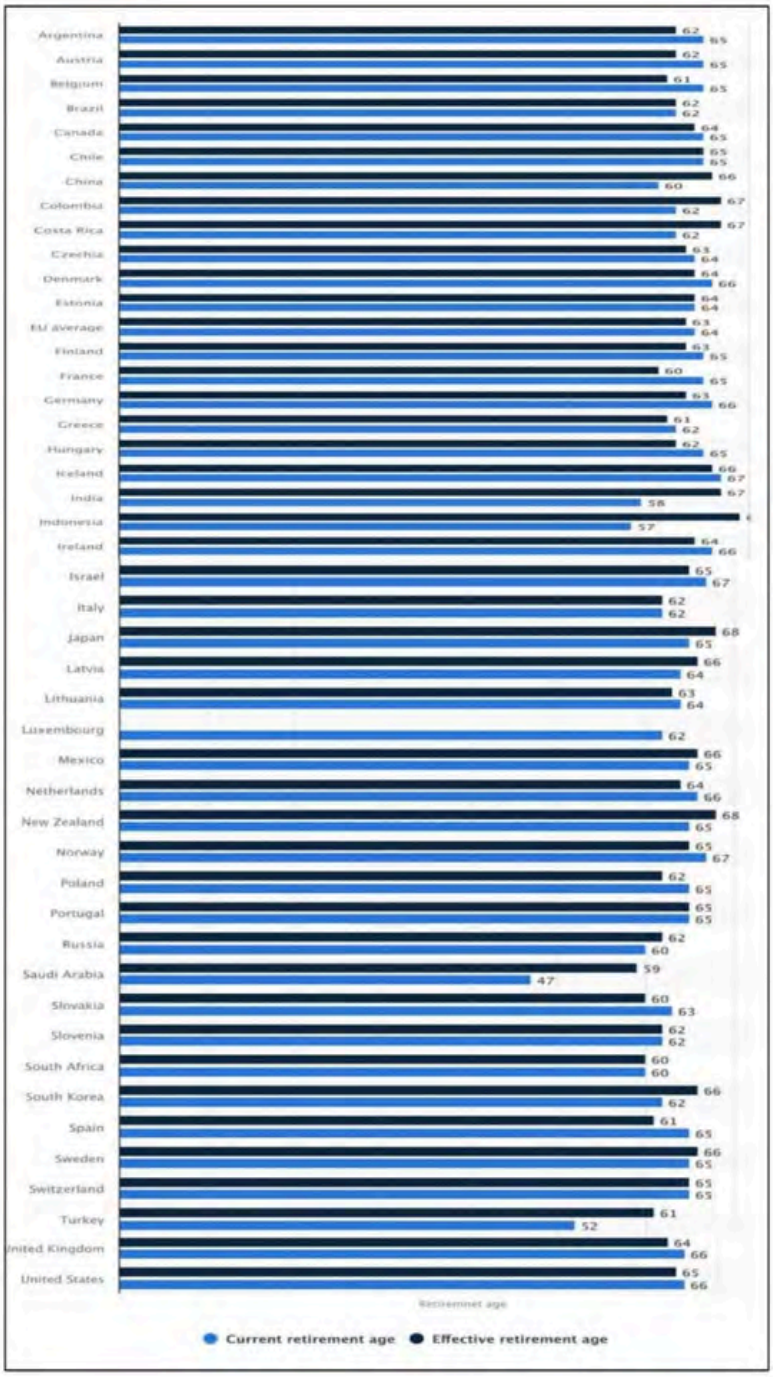
Unger, J.M., Rauch, A., Frese, M. and Rosenbusch, N. (2011) ‘Human capital and entrepreneurial success: A meta-analytical review’, *Journal of Business Venturing*, 26(3), pp. 341–358. doi: 10.1016/j.jbusvent.2009.09.004.

Watkins, H. J. and Omilion-Hodges, L. M. (2025) 'Zooming in on Generational Differences: Exploring Intergenerational Employee Relationships and the Rise of Generation Z in the Workplace', *InternationalJournalofBusinessCommunication*, pp. 1.

Williamson, O. E. (1985) *The Economic Institutions of Capitalism*. New York: Free Press.

Appendix

APPENDIX 1: Current and effective retirement age worldwide in 2020 (Statista, 2023)



APPENDIX 2: Data Analysis – Coded Themes

THEMES	SUB-THEMES	CODES
Understanding and Self-Perception of Being an entrepreneur	Understandings of Entrepreneurship	Survival
		Responsibility
		Problem-Solving
	Self-Identified Role and Capacity	Health problems in seniors
		Family duties for seniors
		Higher stamina in youth
	Motivations to Engage or Refrain from Entrepreneurship	Independence for both groups
		Passion for both groups
		Extra source of income for seniors
		Higher risk as a deterrent for seniors
	Prerequisites for starting a Business	Prior experience
		Business acumen
Motivators for Intergenerational Collaboration	Complementary Skills and Knowledge	Energy and ideas of youth
		Vision and expertise of older
	Mutual Learning and Mentorship	Digital skills of youth
		Guidance from seniors
	Shared Values and Goals	Alignment in vision
	Societal and Community Contribution	Better products
		Longer and healthy life for seniors
		Job opportunities
Barriers and Tensions in Collaboration Collaboration Dynamics	Value and Working style Clashes	Youth portrayed as impatient
		Rigid seniors
	Risk Tolerance gaps	Difference in risk tolerance capacity between both groups
	Stereotypes and Misconceptions	Impatient or lacking discipline (youth)
		Resistant to change (seniors)
	Practical Constraints	Lack of time and energy (Seniors)
	Preferred Roles and Responsibilities	Creative or operational tasks (Youth)
		Strategy, finance, mentoring or planning roles (Seniors)
	Decision-Making Process	Clarity in rules and roles at the start (Mentioned by seniors)
		Efficiency, resolution, and third-party help (Mentioned by youth)
	Trust and communication Patterns	Trust above age in partner selection
		Age gap – a cause of conflict
Enablers and Conditions for successful Ventures	Skill Development and Training Needs	Communication, patience, and empathy
		Technical knowledge and networking
		Honesty, and clarity
	Cultural and Organisational Environment	Inclusivity, and Diversity of thoughts

APPENDIX 3: Interview Question Guide

Interview Guide

Before we dive into the main questions, I'd just like to ask a few quick background details— like your age and where you're from. These are simply to help me better understand and interpret your responses during the analysis.
Your age group or region might be mentioned in the findings, but never alongside your name or anything that could personally identify you. Everything will remain fully confidential and anonymised.

So just to start:

1. Could you please tell me your **age**?
2. And where are you **originally from** or currently living?
3. What kind of work do you **currently do or used to do before retiring**?
4. And if you're retired, may I ask **when you retired**?

Questions:

Theme 1: Personal Situation of Interviewee

1. Can you tell me a little about what you're currently doing in life?
   (e.g., studying, retired, freelancing, exploring new opportunities)
2. Do you see yourself starting a business at any point in the future? Why or why not?
3. Why haven't you started your business yet- and what would need to happen to make you feel ready or confident enough to begin?

Theme 2: Their Idea/Understanding of Business and Entrepreneurship

4. How do you personally define entrepreneurship or running a business?
   (What does being an entrepreneur mean to you?)

Theme 3: Professional Network

5. Do you have friends, mentors, or colleagues from a senior age group/younger age group? If yes, how do you usually connect with them?
6. Have you ever worked with someone from a senior age group/younger age group? If yes, what was that experience like?

Theme 4: Perception of the Other Age Group

7. How would you explain or define the idea of younger and older individuals starting a business together?
8. If you think about people from another generation, what do you believe they can bring to a business partnership?
   (e.g., knowledge, calmness, energy, fresh ideas, experiences, wisdom, tech knowledge)
9. What kind of values or working style differences do you notice in people of senior age group/younger age group?

Theme 5: Past Exposure to Young–Senior Co-Entrepreneurship

9. Have you ever co-created or been part of a project, startup, or group work with someone much older/younger than you? How did it go?

Theme 6: Perception of Young–Senior Co-Entrepreneurship

10. What do you think are the main benefits and difficulties of such cross-generational partnerships?
11. Would you personally be open to such collaboration in business setting? Why or why not?

Theme 7: Complementarities Between Age Groups

13. What are the key strengths you see in older/younger entrepreneurs?
14. Do you think these strengths could complement your own? In what ways?
15. Where do you think the two age groups might struggle or need to adjust when working together?

Theme 8: Hard and Soft Skills Required

16. What technical or professional skills does your age group usually bring to a business setup?
17. What soft skills or personality traits do you think are most useful for working well in a mixed-age team?

Theme 9: Perception of Roles in Collaboration

18. If you were in a team with someone from senior age group/younger age group?, what kind of role would you naturally take on?

19. Are there tasks you'd rather handle yourself—or pass on to your partner? Why?

Theme 10: Common and Specific Skill Needs

20. What are some skills you think both younger and older people should learn or improve to work well together?
21. What kind of training, guidance, or mentorship would help make a mixed-age business team more effective?

Is there anything else you'd like to share about working with other age groups or your thoughts on entrepreneurship?
Thankyou.

APPENDIX 4: Sample of Interview Transcript

Interviewer

Before diving into the main questions, I just wanna ask like again, what is your age currently?

Y10

I'm turning 33 this November.

Interviewer

OK. And you're where you are from exactly?

Y10

From India, I come from Kerala. I work in Bangalore so I have a company. We make films, and videos. We make advertisements, brand films.

Interviewer

OK, you already answered my next question, that was about your occupation. Yeah. So, Y10, can you tell me a little bit about your role, like what you're currently doing in life? And can you also tell me since when you started your company, and what your plans are, a bit about all these things.

Y10

So we established this company just before the COVID outbreak. It was in 2019. We have been freelancing a lot. I was doing a different jobs before that. So the love for passion for cinema and content for that, because of that, me and my friends decide my best friend decided to start a company with another friend and we've been freelancing and then we thought no, let's establish this as a proper protection House 2019. Then COVID came. So just after the establishment COVID came.
And COVID happened. And then five years been working. Works are going on. We work with a lot of clients. So the future plans, yeah. What are from your future plan?
So yeah, you know, securing my job revolves mostly around sales. And I'm a producer too, so I produce films, I onboard clients. So the future plan is going to be the 11 major thing happened was we established a recent office in Dubai. So that is taken care of my partner and he'll be migrating to Dubai. And we will be, you know, onboarding a lot of clients. That's the immediate plan this year. There are a lot of discussions going on.
So then in the next three years, you know we want to and we will be working towards pushing this company to the E Class, that the sought after production house, you know, working with a lot of clients or maybe working with in a couple of like minimal but minimal clients, but good clients, you know. That is that is a future plan.

Interviewer

OK. Just to summarise, basically you have started your company in 2019 with your best friend, and now you guys are running your company. And for your future plans, you plan to, you know, continue with the company and expand it?

Y10

Yes, exactly. Exactly, yeah.

Interviewer
Right, OK. So, Y10, it's just like you mentioned that you started a company with your best friend. May I ask, like, your best friend, must I assume your best friend must be of your age, if I'm not wrong?

Y10

Yes, yes, we are childhood friends.

Interviewer

So, oh, that's amazing. So may I ask, like, why did you plan to start your company with your best friend only? Or were there, like, any other options in your mind while starting your company?

Y10
No, no, no, it it was like he was already in this business. He was already doing this. He was a cinematographer and we discussed it, but I was not in this industry. I was doing some other odd jobs. I was doing different jobs and we decided that, you know, you you do the creative part, I'll do the business part, and that's how it start. So it was not like there were many options and selected this. That was the option in front of me. And our other partner is senior to me. Actually, he's working in an IT sector and he, you know, he's somebody, you know, who loves cinema and want to make films. That's why we joined hands with them. And yeah, that's how we started.
Yes. So there was not many much options out there to choose.

Interviewer

OK, so before like following up with your answer, I'm just gonna ask you that, how do you personally define entrepreneurship or running a business? Like, your personal definition of these terms?

Y10
Entrepreneurship and running a business, right? It's not a classy glamorous thing that people can say, entrepreneurship, and it's responsibilities. You know, it's prioritising a lot of things over your needs. It's about, you know, quick decision making or maybe the best decision making. Sometimes you have to select, you have to go with, you know, what do you actually feel the right one. Then you don't know that, you know, you have to go. Maybe you'll do a lot of trial and error methods. You have to learn from the mistakes. Adapt. So there's a lot of life lessons that you get. For me, entrepreneurship was that.

You know, that that journey of, you know, you you trying and experimenting a lot of things. And sometimes it's not working. Sometimes you feel like, you know, you're gonna quit and that's the end of it. But, you know, that urge, that hunger for, you know, that success, the hunger that, you know, I want to achieve and why I started in the first place, you know, that that that sort of keeps you going. So it's all about the perseverance. It's all about, you know, rising and rising from the ashes. It's all about learning from the mistakes, you know, controlling your ego. I know there's a lot of things. So yes, it's more about life itself and the business. You solving something for your customer in running in businesses. See, that's in most of the times when people, you know, people you've heard about people saying that sales is finding a solution to a problem, right?

It is true, and most of the times before I I really didn't understand that because during your age, maybe, maybe your age, and I was I was doing a lot of job, I used to think why why people are saying that. I mean, it's it's a solution? Sales is just buying and selling, right? Then on the course of the time, when you speak, my job revolves around speaking with a lot of decision makers, a lot of business owners, marketing heads, speaking with them, thinking from their shoes, understanding their pain points. And then you give you a solution, and you're actually agreeing to the final "OK, this works for me." And someone else out there who is a ball-breaking CEO or a top company executive understanding that "OK, this solution works," and we're joining hands and we're seeing the results.
That feeling itself is something that, you know, it's it's pretty hard to explain, because we know that every time you do this, you know that OK, you're on the right and you have this feeling about "OK, OK, we are winning."

Authors Bio

Khushi Agarwal is a graduate of the University of Bristol's School of Management, holding an MSc in Management (Entrepreneurship & Innovation). Her work centres on entrepreneurship, innovation, and intergenerational collaboration, with a particular interest in ventures that create meaningful social value alongside commercial success. Drawn to questions of how ideas emerge and how businesses can address societal challenges, she is especially interested in topics related to well-being, ageing, and quality of life. Outside academia, she enjoys listening to music while taking long walks and observing the world around her, often finding inspiration in everyday conversations, experiences, and human behaviour.

Does the Prosus–Just Eat Takeaway Acquisition Create Value?

A Scenario-Based Valuation of Synergy Uncertainty and Post-Acquisition Financial Outcomes.

Author: Huaijin Xing

ORCID: 0009-0007-5747-007X

Programme: BScEconomics and Finance

Affiliation: School of Economics

Email: do23500@bristol.ac.uk

Abstract

In February 2025, Prosus declared an all-cash proposal of €20.30 per share to take over Just Eat Takeaway.com (JET) and valued the Company at roughly 4.1 billion euro, a 63 per cent premium to its pre-announcement share price. This case study explores the issue of whether the transaction results in shareholder value based on the uncertainty characteristics in the realisation and timing of synergies. This analysis is conducted as a hypothetical, scenario-based valuation exercise rather than an empirical ex-post test of realized value creation. It evaluates the transaction under a set of explicitly stated assumption derived from publicly available information and industry benchmarks. The paper assesses the financial logic of the deal using a mix of valuation benchmarking, earnings-per-share (EPS) accretion analysis and a scenario-based synergy net present value (NPV) model.

These results indicate that although the acquisition itself is a strategic timing, as it is being made after a long period of JET losing its market capitalisation and all being made in cash and without burdening Prosus with any additional leverage, the anticipated synergies only partially justify the acquisition premium when optimistic assumptions are made. Consequently, value development is quite dependent on quality of execution, regulatory restrictions and success of integration as opposed to the deal structure.

Keywords: Scenario-based valuation, Synergy uncertainty, Acquisition premium, EPS accretion analysis, Regulatory constraints

Introduction

With the slowdown of the growth of revenues and the pressure on profitability turning into a real issue after the Covid-19 pandemic, a sudden surge of global food delivery industry has been observed, as the largest corporations in the sector, including Uber Eats, DoorDash, Meituan and Delivery Hero, compete fiercely over scale. It is on this basis that Prosus, the Dutch-listed international consumer internet conglomerate and with significant minority holdings in Platforms such as iFood, Swiggy, Delivery Hero and Meituan, declared a recommended all-cash offer of €20.30 per share to buy Just Eat Takeaway.com (JET) with the target valued at approximately 4.1 billion. This bid was a 63 per cent. premium to the undisturbed JET share price and it came after a time when the market capitalisation of JET had fallen by approximately 15 billion Euros to approximately 2 billion in 2020 (EU-Startups, 2025; Financial Times, 2025a).

The central question that is posed by this transaction with respect to the corporate finance theory and investment practice is: does the Prosus-JET acquisition add shareholder value when the uncertainty regarding post-acquisition synergies takes place? Importantly, the paper evaluates the ex-ate value creation potential of the Prosus-JET transaction under alternative synergy and execution scenarios, rather than measuring the realized, ex-post value creation after completion of the deal. Empirical data indicates that a high percentage of mergers do not achieve the envisioned synergies, especially in technological and platform-based industries where integration is complicated as well as speed of response to competition is high. Though the notion of acquisitions being made at valuation troughs is being promoted as opportunistic and value-accruing, high acquisition premiums may shift a large portion of the anticipated value to target shareholders in the initial phases.

The paper will answer this question by discussing Prosus-JET deal as a specialised case study. Instead of trying to identify the precise firm value in the long term, the analysis is evaluating the transaction using three traditional M&A assessment tools: valuation benchmarking, based on EV/ EBITDA multiples, EPS accretion /dilution analysis of Prosus shareholders and a scenario-based comparison of the synergy NPV and the acquisition premium. The overall verdict is that despite the strategic consistency of the deal and its conservative funding, creating value of the deal is

conditional on the presence of optimistic synergy assumptions and disciplined implementation.

Conceptual Framework

One of the main tenets of mergers and acquisitions is that a transaction should result into value creation in the case where the resulting entity is valued at more than the aggregate of the constituent parts. Practically, this incremental value is usually assigned to synergies which may be divided into operational, revenue-based and financial synergies.

Synergies that are usually related to operations are those that are based on the optimisation of logistics, procurement efficiencies, removal of duplication overheads and are as a rule considered to be highly predictable and immediate. The revenue synergies, such as cross-selling, better customer conversion, and penetrating new market segments are less certain and highly reliant on the effective integration and competitive forces. Value may also be further increased by financial synergies such as tax optimisation, improved access to capital markets or reduced cost of capital, which are usually limited by regulation and taxation, especially in transactions involving cross-border operations. In this paper, operational and revenue synergies are operationalized as incremental cash flow streams that enter a discounted cash flow framework to derive the net present value of synergies under alternative scenarios.

Acquisition premiums are extremely important in the likelihood of synergies creating net value to an acquirer. The high premium is effectively paying anticipated benefits of synergy to the target shareholders at the time of closing, the realised synergies below the premium benefit the deal kills value as viewed by the shareholders of the acquirer. The acquisition premium is operationalised as the difference between the €20.30 offer price and the undisturbed pre-announcement share price multiplied by the number of shares outstanding, which serves as the benchmark against which the synergy NPV is compared. Marketwise, the short-term market consequences are generally evaluated using the accretion/dilution EPS analysis: accretive acquisitions are normally perceived in a positive way, and thus dilutive acquisitions are required to have plausible long-term narratives on value creation. The impact on EPS is

incorporated through a simplified pro-forma accretion/dilution framework that contrasts Prosus's standalone earnings with illustrative post-acquisition EPS under conservative assumption.

These ideas, along with the realisation of synergy, acquisition premium, and the impact on EPS, comprise an analytical basis of determining the acquisition Prosus-JET as the one that may be regarded as value-creating. These conceptual elements including synergies, acquisition premium and EPS, directly underpin the empirical part of the case study by structuring the three analytical tools employed-EV/EBITDA benchmarking, pro-forma EPS accretion/dilution, and a scenario-based comparison of synergy NPV versus acquisition premium.

Case Description: Prosus and Just Eat Takeaway

Prosus is a Dutch-traded multinational consumer internet conglomerate with a portfolio of classifieds, payments and fintech, and food delivery with major minority investments in iFood, Swiggy, Delivery Hero and Meituan. These holdings allowed Prosus to have limited operational control and strategic freedom in the region before the transaction as it was mostly indirectly exposed to European food delivery.

In comparison, Just Eat Takeaway.com was one of the most popular online food delivery Platforms in Europe with solid presence in the United Kingdom, Germany and the Netherlands. Nevertheless, with the pandemic demand rush, JET experienced increased competition, integration, and continued profitability problems, and its net loss in 2024 was also significant, as its reported market capitalisation dropped by about two-thirds to around 2 billion in 2024(EU-Startups, 2025; Financial Times, 2025a).

The deal was organized as a recommended all-cash offer, through the debent of Prosus through its subsidiary MIH Bidco Holdings B. V., at the valuation of €20.30/share, equivalent to an equity worth of approximately 4.1 billion. Prosus said that the considerations would be made through the use of cash resources available and would not involve the introduction of new debt or equity, thus maintaining its investment-grade credit profile and eliminating dilution of shareholders. The acquisition was unanimously approved by the management team and the

supervisory board of JET and was to take place fully through a statutory squeeze out, after the successful completion of the public offer.

The deal gave Prosus direct control of the most strategic markets in Europe including the United Kingdom, Germany and the Netherlands, which complements its positioning in other regions. The communications of the management focused on the implementation of the artificial intelligence of the Prosus, optimisation of logistics, and payment expertise to enhance the operational efficiency of JET and fasten the journey to sustainable profitability, and positioning the united group as a “European food delivery champion”.

Methodology and Assumptions

The analysis is fundamentally assumption-driven-it relies on transparent, scenario-based financial modelling using publicly available data, rather than on proprietary internal information or observed post-deal outcomes. To determine whether the Prosus-JET acquisition ensures value creationtothe Prosus shareholders, this paper implements three complementary testsofanalysis that are usually employed in investment banking and corporate financepractice:

1. **Valuation benchmarking (EV/EBITDA):**
Implied EV/EBITDA multiples are obtainedbycomparing the price of the acquisition with the EBITDA of JET in various historicalandforecasted situations. Since the EBITDA of JET has been negative in someyearsbefore the deal, they are approached with a grain of salt and are usedasapurely contextual tool and not as an absolute valuation tool.

2. **EPS accretion/dilution analysis:**

The accretion/dilution analysis of EPS is done in scenarios of conservative assumptions with no synergy and potential synergies are examined separately in terms of a scenario-based NPV framework. In the analysis, the distinction was made between GAAP EPS and adjusted EPS to consider non-recurring in accordance with the self-reporting practices of Prosus.

3. **Synergy NPV vs acquisition premium:**

The synergy-based valuation is the net present value of anticipated synergies

divided by the acquisition premium contained in the offer of 20.30 per share. The premium is whereby the difference between the offer price and the undisturbed price of shares before the announcement multiplied by the number of shares is obtained. Annual synergies in the range of €0.3–0.5 billion are assumed as an illustrative benchmark. This range is anchored in general M&A integration and synergy studies by global consulting firms, which typically report cost synergies in the mid-single-digit percentage range of the acquired cost base for large-scale consumer and technology transactions, as well as in selected analyst commentary on operational efficiency ambitions in platform businesses. In particular, EY's discussion of M&A integration planning and L.E.K. Consulting's work on synergy realisation support the use of externally benchmarked ranges rather than single-point forecasts in situations where company-specific internal synergy guidance is unavailable. These figures should therefore be interpreted as externally benchmarked, illustrative assumptions rather than management guidance or precise forecasts.

This range is anchored in disclosed cost-synergy outcomes from comparable large-scale food-delivery and consumer-internet transactions, selected analyst and industry reports on typical cost-saving percentages, and management commentary on operational efficiency ambitions, rather than on any formal synergy guidance provided by Prosus or JET. The synergies are modelled at 5 years realisation and discounted at 8% weighted average cost of capital (WACC) in line with the estimates of large and diversified consumer internet groups that have exposure to emerging and developed markets.

To reflect uncertainty, the analysis employs three scenarios:

- **Neutral scenario:** hypothetical operating performance of JET based on its latest actual operating performance before the announcement in the absence of incremental synergies.
- **Historical-average scenario:** smoothing the short-term volatility in margins and cash flows using the arithmetic mean of the 2022-2024 financial indicators of JET.
- **Synergy-enhanced scenario:** reflecting presumed operational efficiency gains and increment of revenues, such as an increase in EBITDA margins and a small increase in revenues through optimisation of logistics and cross-selling.

Under each scenario, EV/EBITDA multiples, EPS outcomes, and synergy NPVs are calculated and compared to the acquisition premium. These synergy figures should therefore be viewed as externally benchmarked, illustrative assumptions used for scenario analysis, not as precise forecasts or management targets.

Results and Discussion

Valuation Benchmarking

The EV/EBITDA multiple suggests that the implied multiple at the point at which the acquisition was made was below the historical high but still high compared to the weaker or more troubled companies and hence the acquisition could not be described as being evidently cheap using only standalone basic. In those years when JET recorded negative EBITDA, the implication of these multiple is mechanically invalid; however, these years also demonstrate the volatility and risk on the operating profile of the target.

The analysis isolates the purchase price relative to the earnings capacity of JET by proxying enterprise value with the equity value paid by Prosus of the purchase, which is the value of the equity of the company in the purchase at 4.1 billion Euros. Although this is a simplification when compared to the standard definition of enterprise value, it ignores differences in net debt and capital structure across firms and thus can distort valuation comparisons. Consequently, the EV/EBITDA benchmarking in this paper should be interpreted as a contextual indicator of pricing rather than as a precise, like-for-like valuation metric.

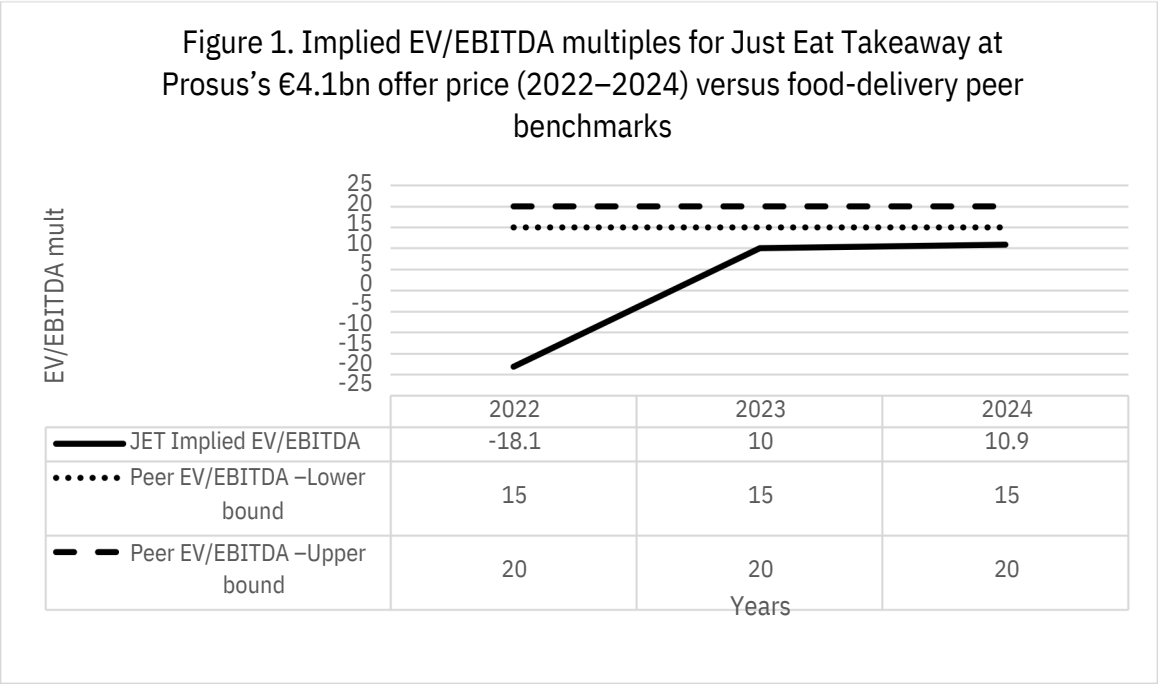


Figure 1. ***Implied EV/EBITDA multiples for Just Eat Takeaway at Prosus's €4.1bn offer price (2022–2024) versus food-delivery peer benchmarks. (Negative EV/EBITDA in 2022 reflects negative EBITDA during post-pandemic operating losses and does not imply negative enterprise value.)***

Accretion / dilution Analysis (Pro-forma EPS)

This section evaluates the significance of acquisition of Just Eat Takeaway (JET) by Prosus on implication of earnings per share (EPS) under a simplified, assumption-based, pro-forma framework. It is not aimed at building a fully consolidated post-transaction income statement, but to assess the directional effect on the EPS under conservative and transparent assumptions, in line with the normal accretion/dilution analysis when discussing M&A cases. The EPS accretion/dilution analysis is deliberately based on a simplified, assumption-driven pro-forma framework. It is intended to illustrate the directional impact of the acquisition on Prosus's EPS under conservative assumptions, rather than to provide a fully consolidated or precise forecast of post-transaction earnings.

The independent (before) EPS values of Prosus are based on Bloomberg-reported GAAP and adjusted Prosus FY2023-2025 earnings. Post-acquisition (after) EPS represents the calculation of the expected contribution of earnings to Prosus that is made by having an estimated contribution of the JET, which is done by keeping the

basic weighted average shares of Prosus constant, implying that the transaction is cash-based and there is no additional equity issue.

The net income contribution is approximated by JET based on the reported disclosed adjusted EBITDA of the company of 460 million and using an effective corporate tax rate of 26 percent and an estimated net income of about 340 million annually (Just Eat Takeaway.com, 2025f; PwC, 2024). There is no assumed operation of synergies, financing effects, or restructuring benefits so that the analysis is conservative. This addition is made straight to the standalone earnings of Prosus in order to come up with adjusted post-acquisition EPS.

A simplified purchase price allocation (PPA) assumption is used to differentiate between the outcomes of the GAAP and adjusted ones. The amortisable intangible assets, take up thirty per cent of the consideration in place of the payment of the consideration namely the Euro 4.1 billion, and amortised over ten years implying that it will be amortised at an annual rate of about 123 million. In the GAAP calculation of post-acquisition EPS, this charge is deducted but not included in adjusted EPS, which is similar to the definition of non-IFRS performance measures used by Prosus itself.

The resultant pro-forma EPS tracts are illustrated in Figure 2. With the addition of the JET earnings, adjusted EPS is modestly accretive over all years of operation considering it has a positive operating contribution. GAAP wise, the transaction would be neutral or slightly accretive since amortisation of acquisition related will offset some of the earnings of JET in the near term. Significantly, within this simplified framework, the illustrate results are obtained even without reference to synergy assumptions, suggesting that the transaction need not be immediately value-destructive on an earning basis. However, given the modelling simplifications, these EPS outcomes should be interpreted as indicative only and not as definitive evidence of value creation.

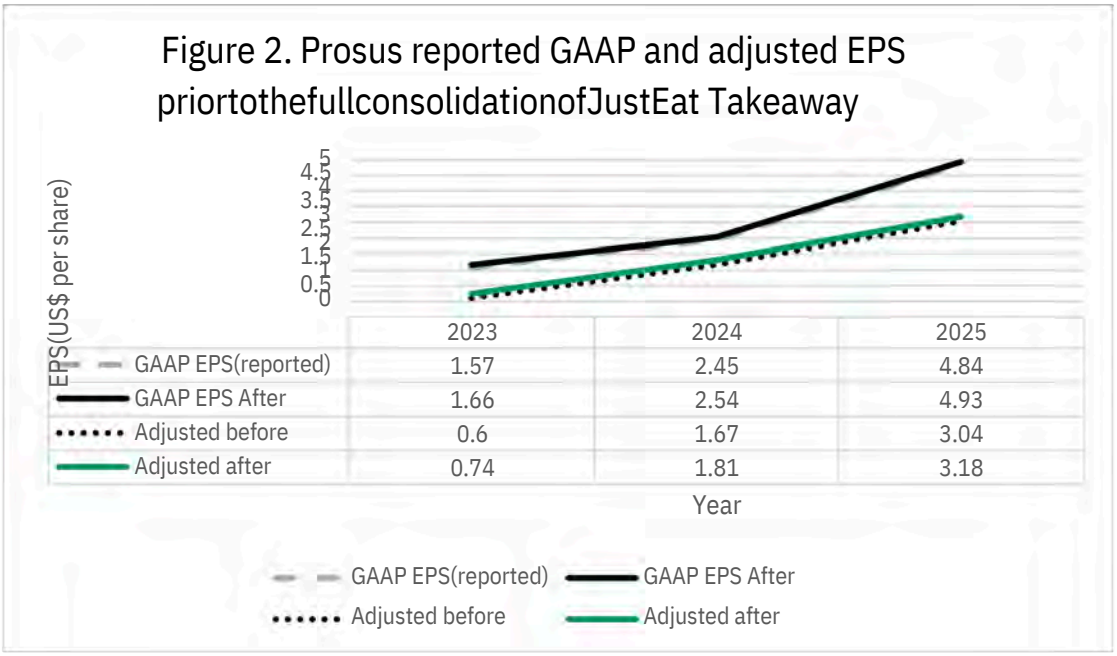


Figure 2. ***Pro-forma comparison of Prosus GAAP and adjusted EPS before and after the acquisition of Just Eat Takeaway. Post-acquisition EPS figures are illustrative and based on simplified assumptions regarding JET earnings contribution and acquisition-related amortisation, rather than on fully consolidated reported results.***

Synergy NPV vs Acquisition Premium

The synergy NPV analysis gives a conservative view regarding the creation of value of the transaction by specifically assessing the approximated present value of the possible synergies against the acquisition premium reflected in the offer price. Since there is no clear quantitative guidance on the guidance of the synergy in the offer documentation at Prosus, the annual synergies of 0.3-0.5 bn are modeled as industry-cored assumptions, not as management predictions. This spectrum is supposed to run across a pessimistic to optimistic spectrum of results, inclusive of the size of the Just Eat Takeaway cost base as well as typical efficiency advantages realized in large-scale food delivery and consumer internet dealings.

In the upper-bound synergy scenario (realised by the 8% WACC discounting of 5 years, the synergies are expected to be of a net value of about 2.1 billion), the net present value of the projected synergies is estimated at 2.1 billion, in comparison to an acquisition premium of 2.6 billion. Even in this favourable case, desired synergies are not enough to cover the premium paid to full extent meaning there is a small

margin of safety to the Prosus shareholders assuming it is on a purely financial basis. Reduced synergy assumptions in the modelled range will lead to value creation of somewhat smaller NPVs, which supports the sensitivity of value creation to realised operational improvements.

In addition to the scenario analysis, a break-even calculation is used to identify the minimum annual synergy required for the NPV of synergies to exactly match the acquisition premium. Holding the five-year realisation horizon and an 8% WACC constant, the break-even annual synergy is approximately €0.65 billion. This implies that Prosus would need to realise synergies close to the upper end of the assumed range on a sustained basis for the transaction to be financially neutral on a premium-versus-synergy basis. Below this threshold, the deal would destroy value for Prosus shareholders on a purely financial basis.

The sensitivity analysis also points out the weaknesses of the case of investment. A small adjustment in assumed annual synergies or a rise in the discount rate causes a strong compression in NPV and more aggressive synergy assumptions cannot be easily justified due to the high levels of inter-industry competition, regulatory pressures and mixed history of successful large scale platform integration in the food delivery business. Combined, the evidence indicates that the value proposition of the transaction is extremely conditional to the capacity of Prosus to perform well and achieve synergies on the high end of the perceived range, not only to deal structure. A regulation-constrained scenario is also considered to reflect the European Commission’s conditional clearance, which required Prosus to reduce its stake in Delivery Hero and relinquish board representation. In this scenario, reduced strategic flexibility and coordination constraints are proxied by a haircut to the achievable synergies (for example, a 20–30% reduction in the annual synergy level and/or a slower ramp-up over the five-year horizon). Under such regulation-constrained assumptions, the resulting synergy NPV falls materially short of the acquisition premium, highlighting regulation as an additional brake on value creation beyond pure financial considerations. These regulatory remedies are therefore explicitly incorporated into the adverse synergy scenario, reinforcing that regulatory constraints can significantly erode the synergy headroom available to Prosus.

Overall, the conclusion that modelled synergies only marginally justify the acquisition premium is highly sensitive to the assumed synergy level. Small changes in the annual synergy assumption within a plausible range are sufficient to move the transaction from marginal value creation into value destruction, underscoring that the results are driven by the chosen inputs rather than by independently verified synergy outcomes.

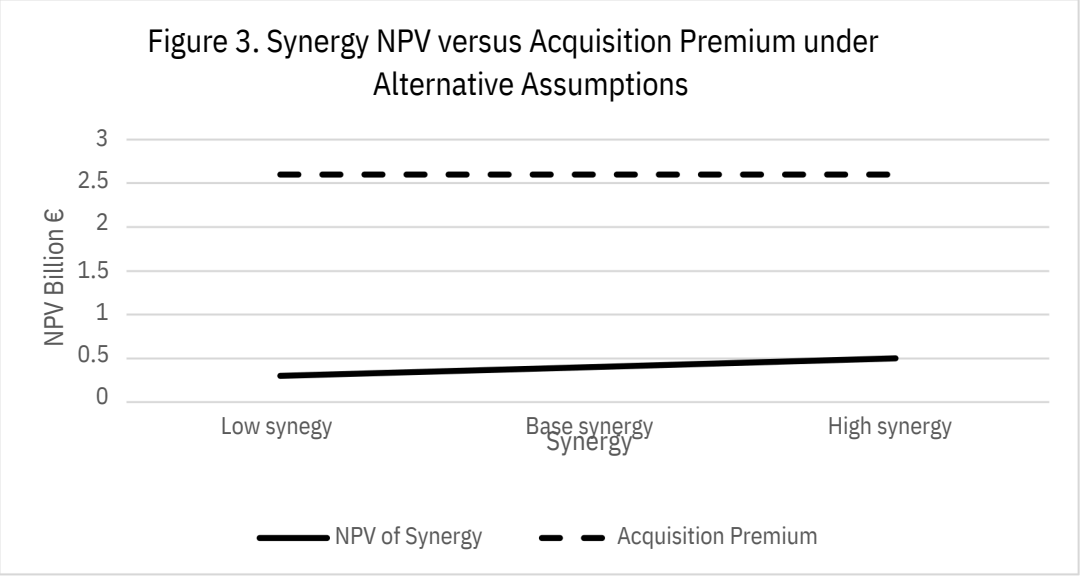


Figure 3*. Synergy NPV versus Acquisition Premium under Alternative Assumptions. The acquisition premium is defined as the difference between the €20.30 offer price and the undisturbed pre-announcement share price multiplied by shares outstanding.*

Sensitivity Analysis

To systematically assess parameter sensitivity, a simple sensitivity grid is constructed (Tables 1 and 2). First, holding the discount rate at 8%, annual synergies are varied by ±10% and ±20% around a €0.4 billion base case. Under these variations, the synergy NPV ranges from approximately €1.28 billion at -20% to €1.92 billion at +20%, remaining below the approximately €2.6 billion acquisition premium throughout. Second, fixing annual synergies at €0.4 billion and varying the discount rate between 7%, 8% and 9% yields a synergy NPV range of roughly €1.64 billion, €1.60 billion and €1.56 billion respectively. These sensitivity checks show that modest changes in assumptions materially affect the valuation outcome and reinforce that the transaction does not comfortably cover the premium under central assumptions.

Table 1

Assumption change	Annual synergy(€bn)	NPV of synergies (€bn)
-20%	0.32	1.28
-10%	0.36	1.44
Base case	0.40	1.60
+10%	0.44	1.76
+20%	0.48	1.92

Table 2

WACC	NPV of synergies at €0.4bn base case
7%	1.64
8%	1.60
9%	1.56

Regulatory and Antitrust Context

The regulatory consent was a key factor in the viability of the deals. All required regulatory approvals were received by August 2025 to allow the offer to be completed and the squeeze-out. To resolve issues with cross-shareholdings and coordination in the European food delivery markets, the European Commission issued conditional clearance on 11 August 2025, on the condition that Prosus would reduce its stake in Delivery Hero to less than 10 percent and give up its seats in the board, (Reuters, 2025a; Reuters, 2025b; Silicon Republic, 2025).

These solutions point to the fact that the deal was not considered in vacuo but rather as a part of a more extended ecosystem where Prosus already had influential roles. Although the circumstances did not undermine the deal, they limited the flexibility of Prosus to use its portfolio of food delivery assets to further strategically integrate and increased the execution hurdle to deliver synergies that were primarily in the standalone operations of JET. Regulatory constraints are thus another source of

execution risk on top of the strictly financial measures that are reflected in the synergy NPV and EPS models.

Conclusion and Limitations

Prosus-JET acquisition is reviewed in this case study in terms of the synergy uncertainty, regulatory restriction as well as the financial performance following the acquisition. On the basis of the scenario-based assumptions adopted, the analysis suggests that the transaction appears strategically well-timed, after JET had significantly declined in its valuation, and conservatively financed without the need to raise leverage or dilute shareholders. The quantitative findings however show that the anticipated synergies are only marginally sufficient to explain the acquisition premium, even with good assumptions and there is not much space to make an error in execution.

There are a number of limitations that need to be recognized. First, the analysis is based on estimations of modelled synergy, not quantitative instructions given by the management, which supports the absence of comprehensively reported synergy in statutory filings and offer documents. Second, the framework is more concerned with financial measures, and it does not explicitly simulate the issue of cultural integration, labour market trends, and possible strategic repositioning to deal with competitive actions by competitors. Third, the analysis does not reflect post closing performance and realized synergies despite the fact that there will be a conservative pro-forma evaluation of near-term benefit impacts, but such future value creation will hinge upon the quality of execution and the degree to which anticipated operational enhancement will be achieved. Fourth, for simplicity, the analysis proxy's enterprise value by the equity value implied by the offer price and does not fully adjust for differences in net debt, which limits the precision of EV/EBITDA-based comparisons. All conclusions are therefore conditional on the assumed inputs and should not be interpreted as definitive, empirical evidence of actual value creation.

References

EY (2024) 'Four tips for estimating mergers & acquisitions integration costs', *EY Insights*. Available at: https://www.ey.com/en_cn/insights/strategy-transactions/four-current-trends-estimating-mergers-acquisitions-integration-costs (Accessed 3 January 2026).

EU-Startups (2025) *Just Eat Takeaway acquired by Prosus in €4.1 billion deal amid €1.6 billion net loss*. EU-Startups, 23 February. Available at: https://www.EU-Startups.com/2025/02/just-eat-takeaway-acquired-by-Prosus-in-e4-1-billion-deal-amid-e1-6-billion-net-loss/ [Accessed 3 January 2026].

Financial Times (2025a) *Acquisition – Prosus offer for Just Eat Takeaway.com N.V.* FT.com Markets data, 23 February. Available at: https://markets.ft.com/data/announce/detail?dockey=1323-16910941-019SR55LQP91FBFO2V2JOMHST0 [Accessed 3 January 2026].

L.E.K. Consulting (2024) 'Keys to unlocking merger value: cost synergies', *L.E.K. Insights*. Available at: https://www.lek.com/insights/post-merger-integration/keys-unlocking-merger-value-cost-synergies (Accessed 3 January 2026).

Just Eat Takeaway.com (2025f) *Full Year 2024 Results*, 24 February. Available at: https://newsroom.justeattakeaway.com/en-WW/247233-full-year-2024-results/ (Accessed: 3 January 2026).

PwC (2024) *Taxation in the Netherlands*. Available at: https://www.pwc.nl/en/insights-and-publications/services-and-industries/tax/doing-business-in-the-netherlands/taxation-in-the-netherlands.html (Accessed: 3 January 2026).

Reuters (2025a) *Prosus offers to slash Delivery Hero stake to address EU concerns about Just Eat deal*. Reuters, 18 July. Available at: https://www.Reuters.com/sustainability/boards-policy-regulation/Prosus-offers-slash-delivery-hero-stake-address-eu-concerns-about-just-eat-deal-2025-07-18/ [Accessed 3 January 2026].

Reuters (2025b) *Prosus wins conditional EU antitrust nod for Just Eat Takeaway*

deal. Reuters, 11 August. Available at: https://www.Reuters.com/sustainability/boards-policy-regulation/Prosus-wins-conditional-eu-antitrust-nod-just-eat-takeaway-deal-2025-08-11/ [Accessed 3 January 2026].

Silicon Republic (2025) *EU greenlights Prosus acquisition of Just Eat Takeaway*. Silicon Republic, 11 August. Available at: https://www.siliconrepublic.com/business/Prosus-just-eat-takeaway-delivery-hero-acquisition-European-commission [Accessed 3 January 2026].

Authors Bio:

Huaijin Xing is a 2026 graduate of the University of Bristol's School of Economics, holding a Bachelor's degree in Economics and Finance. Her research interests focus on Corporate Finance and Financial Market, particularly in the area of M&A. She enjoys reading and further research, applying these skills both academically and in Career Development.

Pink PP as Cultural Branding: How Luxury Brands Extract Community Meaning into Commercial Value

Author: Shenzimo Zhang
ORCID: 0009-0003-1585-3495
Programme: MSc Marketing
Affiliation: School of Management
Email: we25059@bristol.ac.uk

Abstract

This article examines how Valentino's Pink PP campaign transforms cultural significance into commercial value within luxury branding. Pink's symbolic meaning was not invented by the brand, but emerged through history, identity politics, and online community practices. This paper adopts a conceptual case study approach, synthesising a four-stage model from existing literature with Pink PP as an explanatory case. Based on the McCracken (1986) meaning transfer model, this paper proposes a four-stage conceptual framework: communities generate meaning; the brand extracts emotional value and reconstructs it through aesthetic language; digital communities amplify and disseminate this meaning; and the brand ultimately captures commercial value. This paper's original contribution lies in two added dimensions: digital community amplification and asymmetric value distribution. Taking Pink PP as an explanatory example, this study explains how the model operates in practice: Pink PP keeps the emotional core of identity politics but weakens its political context. Collective struggle becomes personal expression and visual participation. Digital communication helps the campaign travel further, but the communities involved do not receive equal returns. This paper reveals the internal tension between symbolic inclusion and commercial extraction.

Keywords: Cultural branding, meaning transfer, luxury marketing, brand community, symbolic value

Introduction

Over the past decade, luxury brands have shifted their focus from product quality to cultural significance. Brands used to rely on scarcity, craftsmanship and tradition to convey value, but now many consumers, especially the younger generation, are also expressing identity when buying luxury goods. Colour, style, and visual function as cultural symbols, enabling brands to connect with identity, emotions, and social values (Holt, 2004).

The available literature shows that cultural significance is often formed from within the community. Consumer culture theory holds that brands do not create these meanings independently. ‘meaning’ comes from social behaviour, common history and group identity. However, there is less discussion on how luxury brands can extract, repackage, and convert their emotional value into commercial value when luxury brands enter these pre-existing cultural fields.

This study addresses the following question: How can luxury brands extract cultural significance from the communities and translate it into commercial value? Valentino Pink PP was selected as an illustrative rather than a representative case: it is not used to represent all luxury brand campaigns, but to explain how a colour with a cultural load is transformed into an identifiable, spreadable and sellable brand symbol.
Launched in Paris in March 2022 as part of the Valentino autumn and winter series, the campaign was led by creative director Pierpaolo Piccioli, applying a single PP Pink hue across multiple brand touchpoints, including shows, clothing, advertising, and packaging, transforming pink from decorative elements into a unified brand language.

This article integrates the theories of brand meaning, cultural meaning transfers and community value creation, proposing a four-stage model: communities generate meaning; the brand extracts emotional value and reconstructs it; the community expands meaning through digital communication; and the brand captures the final

value. The model adds two dimensions based on McCracken (1986): digital community amplification and value asymmetry distribution. Pink PP is used to illustrate how this conceptual model operates in specific brand campaigns.

Methodology

This article adopts a conceptual case study path approach. The four-stage model integrates existing literature on meaning transfer, brand community and cultural brand, extending McCracken's (1986) framework by adding two dimensions: digital community amplification and asymmetric value distribution.

Valentino Pink PP activity (2022) was selected as an illustrative case because it systematically applies cultural symbols across many brand touchpoints, and the influence of digital communication is outstanding, making the process of commercialising cultural meaning especially visible. Evidence draws from published industry reports, fashion media sources, and public social media data from TikTok, Weibo and Google Trends (accessed 23 May 2026).

Literature Review and Theoretical Framework

Meaning comes from the community, not brand invention

Brand meaning is not created unilaterally by the brand but gradually formed in community practice. Consumers express their identity through brands, so brand meaning is shaped by both social concepts and marketing messages.

Batra (2019, p. 535) argues that brand meaning is a collection of associations in consumer memory, which determines how consumers understand their relationship with the brand. Belk (1988, p. 139) further put forward the concept of ‘extended self’, pointing out that individuals incorporate brands into their own identity narratives. Escalas and Bettman (2005, p. 378) indicate that brands enable consumers to tell personal stories within a shared cultural framework. Muñiz and O’Guinn (2001, p. 412) add that brand

meaning is socially co-created through shared consciousness, rituals and a sense of responsibility within brand communities.

Therefore, brand significance is first and foremost a cultural consensus reached by the community. Belk (1988) focuses on individual dimensions, and Muñiz and O'Guinn (2001) extended it to the level of collective co-creation, which together constitute the theoretical basis of the first stage of the four-stage model in this article. The brand then entered this meaning space and participated in the subsequent cultural interpretation.

How brands use community meaning: Emotional extraction and reframing

Brands tend to select emotionally appealing parts of the symbolic meaning generated by communities around a given symbol and use them for commercial purposes. According to McCracken's (1986, p. 71) meaning transfer model, cultural meanings flow from the cultural world to consumer goods, and then to individual consumers. Brands do not produce meaning but mediate and reframe it.

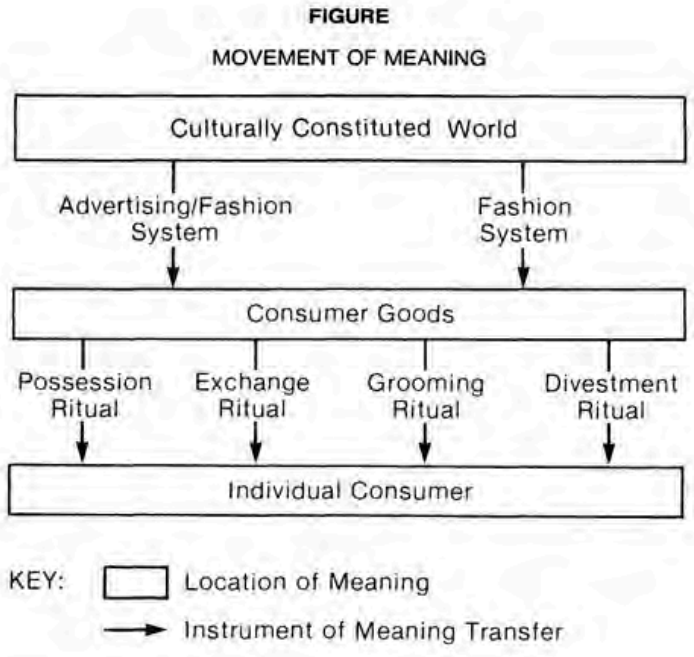


Figure 1. ***McCracken's Meaning Transfer Model***

This transfer is selective in luxury brands: brands tend to retain the emotional core of a symbol (such as freedom, self-expression or resistance), while stripping out specific political content to avoid risks and expand universality. The final result is not a complete historical account, but a simplified message that can be circulated globally.

However, McCracken's model does not address how digital communities spread meaning, nor how value is unevenly distributed among participants. In this study, the four-stage model extends these two gaps, which constitute a conceptual extension rather than an empirical discovery.

Community communication expands brand meaning

After the brand completes the initial coding, meaning enters social circulation. Muñiz and O'Guinn (2001, p. 412) show that meaning develops through interaction rather than one-way communication. Escalas and Bettman (2005, p. 378) explained that this process contributes to building an individual identity. Therefore, the brand provides symbols; the community makes it come alive. The cycle of encoding, decoding and re-encoding together shapes the final meaning of cultural symbols.

Meaning circulates through community interaction, but brands often gain more economic value. Batra (2019, p. 535) finds that when brand meaning relates to personal identity, consumers are more loyal and more willing to purchase, translating cultural recognition into economic benefits. However, the community perspective of Muñiz and O'Guinn (2001) does not fully explain how the economic value jointly created by the community is distributed. Communities receive symbolic returns such as visibility and emotional recognition, while brands gain sales and cultural authority. This gap between creators and beneficiaries is central to Stage 4 of the four-stage model in this paper.

Case Context

Historical accumulation of pink meaning

The cultural meaning of pink has constantly evolved throughout history, indicating that colour meaning is a product of social construction rather than natural attributes. In European court culture, pink silk and embroidery were once symbols of male aristocratic identity and elegance (Paoletti, 2012). Later, through mass marketing and gendered consumer culture, pink was redefined as a symbol of ‘sweetness and femininity’ and became a visual expression of femininity in gender norms (Paoletti, 2012). The pink triangle was originally imposed on persecuted gay men as a symbol of Nazi persecution and was later reappropriated by the LGBTQ community as a symbol of resistance and liberation (United States Holocaust Memorial Museum, n.d.; Holocaust Centre North, 2021). Therefore, the meaning of pink is rooted in the emotional expressions of different historical periods and different communities, rather than being the creation of any single brand.

Pink in digital culture

Pink in digital culture has gradually become an emotional code on social platforms. TikTok and Instagram users integrate pink into styling and visual content to express identity and personal values, rather than participating in political debates. Therefore, when brands use pink, they are entering a symbolic field that already exists in digital culture, rather than constructing meaning from scratch.

This field was formed before the emergence of Pink PP. Around 2016, Millennial Pink became popular on Tumblr and Instagram, evolving into a digital aesthetic associated with youthful identity and emotional expression. Pink PP entered this already activated cultural context in March 2022 and systematised pink into a luxury brand language. Lyst's annual report (Lyst, 2022c) shows that the number of searches increased by 152% in the week after the Valentino show, and the collection subsequently received more than 76.9 million views on TikTok. This suggests that the scale of dissemination is difficult to explain through brand advertising alone.

At the same time, after the release of Margot Robbie's on-set photos in June 2022, ‘Barbie core’ quickly became a significant fashion trend, and searches for related ‘pink items’ increased by 416% (Lyst, 2022c). Lyst listed ‘Barbie core’ as the annual trend of 2022, while also pointing out that ‘the defining pink of the year came from Valentino’. By the official release of the film in 2023, pink had surpassed its status as a brand colour or passing trend to become a cultural symbol connecting identity, nostalgia and self-expression. Pink PP and ‘Barbie core’ together suggest that the cultural influence of pink is shaped by brands, media, platform dynamics, and communities; Valentino did not create this meaning but converted the already circulating cultural energy into luxury brand value.

How Pink PP enters the cultural narrative and creates marketing value

Valentino did not create a pink meaning from scratch, but systematised a symbol already activated by history and digital culture into a complete brand visual language. A single shade ‘PP Pink’, covers runways, shows, clothing, advertising and packaging, forming a highly recognisable brand code. The campaign conveys the concept of colour symbolising freedom and personal expression through aesthetic language (Vogue Runway, 2022), rather than drawing on the political history behind pink, thus extracting emotional meaning from symbols and converting it into a carrier of identity expression.

This highly recognisable visual code lowered the barrier for users to imitate, recreate and share content across platforms. The dissemination of Pink PP therefore did not only rely on brand advertising but drew on the emotional energy that pink already carried within communities and platform communication mechanisms, expanding a brand campaign into a broader cultural phenomenon. This is the core of Pink PP's marketing value: the brand selected an already activated symbol and built it into a transmissible and participatory global luxury visual strategy.

Results: The Four-Stage Model in Practice

Emotional extraction: entering an existing meaning field

This paper proposes that luxury brands do not create cultural meaning from scratch but selectively enter the symbolic field where communities have accumulated emotional value.

Piccioli described PP Pink as the colour of ‘love, community, energy and freedom’ and a powerful symbol of ‘equality and love’ (Anne of Carversville, 2022). Valentino officially stated that Pink PP ‘celebrates a beauty of tolerance beyond gender’ (Valentino, 2022), and the show presented both men's and women's clothing (Vogue Runway, 2022).

However, these official expressions remained at the emotional level, without engaging with specific political contexts such as queer history or feminist activism. This can be read as a selective extraction: the brand retains the emotional core and abstracts the political background in order to reach a diverse audience and reduce political risks, converting pink’s emotional resonance into brand identity resources (Kapferer and Bastien, 2012).

Aesthetic reframing: systematising cultural symbols into brand language

By converting cultural symbols into a unified visual system, the brand achieves the key transformation from emotional meaning to commercial code. A single shade, ‘PP Pink’, covers runway shows, clothing, advertising and packaging, forming a highly recognisable brand code. Piccioli described this strategy as ‘maximising the possibility of expression in the apparent lack of possibility’ (WWD, 2022a). In other words, a single colour eliminates visual noise so that the tailoring, material and individual characteristics of the clothing can be more fully presented. This suggests that monochromatic branding is not merely an aesthetic preference but a deliberate visual tool for amplifying individual expression.

Zhou et al. (2026) show that innovative luxury brands use high-saturation colours to enhance perceived brand status. Valentino's collaboration with Pantone to develop the exclusive shade PP Pink can therefore be understood as an embodiment of this visual code strategy. This aligns with an extension of McCracken (1986) meaning transfer framework: not only the ‘transfer’ of meaning, but also an active ‘reframing’. Specific political and historical content is replaced by universal aesthetic language, and the cultural meaning is reorganised into a brand language that is easier to identify, replicate and consume.

Digital circulation: collaborative amplification between platforms and communities

The imitation and re-creation of brand visual codes in the digital community constitute the third stage not covered by McCracken's original model. Lyst's annual report (Lyst, 2022c) shows that Pink PP campaign received more than 76.9 million views on TikTok, and the number of searches increased by 152% in the week after the show; Weibo-related t hashtags accumulated more than 140 million reads (Sina Finance, 2022). This scale is difficult to explain through brand advertising alone and is better understood as the result of the combined effects of community participation and platform dynamics.

Platform dynamics also played a role. Baumann et al. (2026) indicate that TikTok systematically amplifies content that matches users' interests. The highly recognisable visual qualities of PP Pink lower the threshold for UGC creation and dissemination, allowing brand codes to spread at the community level.

Value capture: emotional capital converted into brand assets

Value asymmetry is the core extension of the four-stage model and also a structural gap not addressed in McCracken's original model. During the Pink PP campaign, Valentino achieved brand asset accumulation at multiple levels. At the financial level, brand sales reached 1.419 billion euros in 2022, a year-on-year increase of 15%, with direct channels growing by 21% (FashionNetwork, 2023). It should be noted that this growth

was driven by multiple factors, including direct channel expansion and channel restructuring, and cannot be attributed to Pink PP alone.

At the consumer perception level, the Lyst brand popularity index (which aggregates search volume, views and sales data across platforms) shows that Valentino ranked fourth globally in Q2 2022, surpassing Louis Vuitton (Lyst, 2022a). Industry media reporting on the Lyst Q3 Index noted that consumers' positive response to Pink PP was an important factor in the brand's sustained rise in popularity (WWD, 2022b). This growth in consumer attention shows that Pink PP expands brand reach from a traditional luxury audience to a wider range of digital cultural participants.

At the cultural authority level, the Lyst annual report regards Valentino as a brand that defines pink vision in 2022 (Lyst, 2022c), which can be understood as strengthening the brand's cultural leader in colour and identity expression (Business of Fashion, 2022; Business of Fashion, 2024).

The emotional labour contributed by the community through the creation of UGC, the spread of visual symbols and the construction of cultural connotations are eventually transformed into the brand's private assets in the form of sales growth, brand popularity and cultural authority (Arvidsson, 2005), and the rewards obtained by the community mainly stay at the level of visibility and emotional satisfaction. Co-creation does not guarantee value sharing, even if co-creation theory usually frames it as mutually beneficial value production (Prahalad and Ramaswamy, 2004). In the Pink PP case, there is a clear structural gap between cultural production and economic returns. This insight is positioned as a conceptual proposition of this case, rather than a general law of all luxury brand cultural marketing.

Critical Discussion

Identity politics as a symbolic resource

Pink PP deserves critical examination at the ethicallevel.Pinkisnotaneutralaesthetic element. It has been linked to gender norms, queerhistoryandresistancepolitics,soit has a strong identity resonance with the LGBTQcommunityandthegender-fluid audience. Valentino's strategy is not to directlystatepoliticalpositions,buttousethese emotional associations to build a ‘pioneer, freeandinclusive’brandimage.

This ambiguity is the key: brands use colourswithahistoryofstruggletosendsolidarity signals but avoid explicitly acknowledgingtherelevantpoliticalcontextinofficial communication. In other words,PinkPPusestheemotionalmobilisationofidentity politics but translates it into asaferandmorewidelycirculatingbrandlanguage.This shows that borrowing symbolsfromcultureisnotaneutralact;culturalmeaningcanbe transformed into brand value,whiletheoriginalresistancecontextisweakenedinthe process of commercialisation(Arvidsson,2005).

Limitations of visual diversity

The visual presentation ofPinkPPcanbeunderstoodasasymbolicinclusion.It presents multiple identitiesthroughcasting,stylingandcolourlanguage,butthis diversity still exists withintheluxuryaestheticsystem.Inotherwords,thecampaign allows different identitiestoenterthepicture,butitdoesnotfundamentallychangethe luxury brand's definitionofbeauty,body,anddesire.

Therefore, the inclusivenessofPinkPPcanstillbeunderstoodwithinmainstream aesthetic norms (Gurrieri andCherrier,2013).Identityexpressionispackagedasa luxury experience, and equalityandfreedomcanalsobetransformedintoconsumable visual symbols. This reveals aninherentcontradiction:brandswanttoestablish emotional connections with diversegroups,butstillrelyontraditionalluxurybrand systems to produce desirability andscarcity.Diversityismorelikeavisualsignalhere, not necessarily a fundamental changeinthevaluesystem.

Value asymmetry: who benefits from cultural meaning

Pink PP also exposes value asymmetry in cultural co-creation.Themeaningofpink comes from long-accumulated history, community practice anddigitalcirculation,butit is the brand that ultimately systematises, amplifies and monetisesit.Community participation allows symbols to gain visibility and emotional life, whilebrandsgainsales, cultural authority and symbolic capital.

This does not mean that communities have not received any rewards; they have indeed gained visibility, identity affirmation and emotional satisfaction. But these returns are mainly at the symbolic level, and the brand obtains more durable economic and brand assets. Therefore, co-creation does not necessarily mean value sharing, even if co-creation theory usually understands it as a mutually beneficial value production process (Prahalad and Ramaswamy, 2004). In this case, activism is repositioned as personal expression, and cultural history becomes brand equity, which raises a core question: when cultural meaning is commercialized, who really benefits from it?

Conclusion

Taking Valentino Pink PP as an explanatory case, this paper applies and elaborates a four-stage conceptual framework extended from McCracken's (1986) meaning transfer model.

Analysis shows that the symbolic meaning of pink is not created by the brand from scratch, but gradually formed through historical accumulation, identity politics and digital community practice. Pink PP selectively extracts the dimensions of freedom, equality, and inclusivity from this affective field, while weakening the specific political context and transforming it into an aesthetic language that can be circulated worldwide. The digital platform mechanism, together with the community UGC, amplifies this visual symbol, allowing the campaign to go beyond a single seasonal project and enter wider cultural circulation. At the same time, value distribution is not symmetrical: the community contributes emotional labour and communication momentum, while brands gain sales growth, brand prominence and cultural authority.

The original contribution of this paper is to add two dimensions of digital community amplification and value asymmetry capture to the McCracken framework and elaborate them as a conceptual extension. This analysis reveals the core tension in luxury cultural branding: the tension between symbolic inclusion and commercial extraction: brands can build a pioneering image with the help of the emotional power of community history, but do not necessarily bear the corresponding political responsibility.

This paper is limited to a single case and public data, does not draw on brand internal decision-making materials, and does not conduct systematic consumer research. Future research can apply this framework to more luxury cultural marketing cases or test the perception and response of different audiences to the commercialisation of cultural symbols through primary research.

Reference List

Anne of Carversville (2022) 'Valentino's Pink PP Fall 2022 Campaign'. Available at: https://anneofcarversville.com/fashion/2022/8/3/valentino-fall-2022-campaign (Accessed: 23 May 2026).

Arvidsson, A. (2005) 'Brands: A critical perspective', *Journal of Consumer Culture*, 5(2), pp. 235–258.

Batra, R. (2019) 'Creating brand meaning: A review and research agenda', *Journal of Consumer Psychology*, 29(3), pp. 535–546.

Baumann, F., Arora, N., Rahwan, I. and Czaplicka, A. (2026) 'Dynamics of algorithmic content amplification on TikTok', *EPJ Data Science*, 15, article 18. Available at: https://doi.org/10.1140/epjds/s13688-026-00629-2 (Accessed: 23 May 2026).

Belk, R.W. (1988) 'Possessions and the extended self', *Journal of Consumer Research*, 15(2), pp. 139–168.

Business of Fashion (2022) 'Is hot pink here to stay?' Available at: https://www.businessoffashion.com/articles/marketing-pr/hot-pink-trend-valentino-trendy-millenial-pink-barbiecore-barbie-margot-robbie/ (Accessed: 23 May 2026).

Business of Fashion (2024) 'How Valentino's Pierpaolo Piccioli harnesses the power of colour'. Available at: https://www.businessoffashion.com/articles/luxury/how-valentinos-pierpaolo-piccioli-harnesses-the-power-of-colour/ (Accessed: 23 May 2026).

Escalas, J.E. and Bettman, J.R. (2005) 'Self-construal, reference groups, and brand meaning', *Journal of Consumer Research*, 32(3), pp. 378–389.

FashionNetwork (2023) 'Valentino: sales up 15% in 2022'. Available at: https://us.fashionnetwork.com/news/Valentino-sales-up-15-in-2022,1509678.html (Accessed: 23 May 2026).

Gurrieri, L. and Cherrier, H. (2013) 'Queering beauty: Fat, femme and resisting normative consumer culture', *Journal of Macromarketing*, 33(3), pp. 128–140.

Holocaust Centre North (2021) 'Reclaiming the pink triangle'. Available at: https://holocaustcentrenorth.org.uk/blog/reclaiming-the-pink-triangle/ (Accessed: 23 May 2026).

Holt, D.B. (2004) *How Brands Become Icons: The Principles of Cultural Branding*. Boston: Harvard Business School Press.

Kapferer, J.-N. and Bastien, V. (2012) *The Luxury Strategy: Break the Rules of Marketing to Build Luxury Brands*. London: Kogan Page.

Lyst (2022a) 'The Lyst Index: Fashion's hottest brands and products Q2 2022'. Available at: https://www.lyst.com/data/the-lyst-index/q222/ (Accessed: 23 May 2026).

Lyst (2022b) 'The Lyst Index: Fashion's hottest brands and products Q3 2022'. Available at: https://www.lyst.com/data/the-lyst-index/q322/ (Accessed: 23 May 2026).

Lyst (2022c) *The Year in Fashion 2022*. Available at: https://www.lyst.com/data/year-in-fashion-2022/ (Accessed: 23 May 2026).

McCracken, G. (1986) 'Culture and consumption: A theoretical account of the structure and movement of the cultural meaning of consumer goods', *Journal of Consumer Research*, 13(1), pp. 71–84.

Muñiz, A.M. and O'Guinn, T.C. (2001) 'Brand community', *Journal of Consumer Research*, 27(4), pp. 412–432.

Paoletti, J.B. (2012) *Pink and Blue: Telling the Boys from the Girls in America*. Bloomington: Indiana University Press.

Prahalad, C.K. and Ramaswamy, V. (2004) 'Co-creation experiences: The next practice in value creation', *Journal of Interactive Marketing*, 18(3), pp. 5–14.

Sina Finance (2022) 'Valentino "粉色风暴"背后的战略与思考'. Available at: https://cj.sina.com.cn/articles/view/6192937794/17120bb4202001xikx (Accessed: 23 May 2026).

United States Holocaust Memorial Museum (n.d.) 'Gay men under the Nazi regime', *Holocaust Encyclopedia*. Available at: https://encyclopedia.ushmm.org/content/en/article/gay-men-under-the-nazi-regime (Accessed: 23 May 2026).

Valentino (2022) 'Valentino Pink PP Collection'. Available at: https://www.valentino.com/en-gb/fashionshow/valentino-pink-pp-collection (Accessed: 23 May 2026).

Vogue Runway (2022) 'Valentino Fall 2022 Ready-to-Wear Collection'. Available at: https://www.vogue.com/fashion-shows/fall-2022-ready-to-wear/valentino (Accessed: 23 May 2026).

WWD (2022a) 'Pantone and Valentino debut Valentino Pink PP at Paris Fashion Week'. Available at: https://wwd.com/sourcing-journal/industry-news/pantone-valentino-pink-pp-exclusive-color-paris-fashion-week-pierpaolo-piccioli-1238791954/ (Accessed: 23 May 2026).

WWD (2022b) 'Lyst: Italian brands and comfort shoes are piping hot'. Available at: https://wwd.com/sourcing-journal/sj-denim/lyst-index-q3-2022-gucci-miu-miu-birkenstock-ugg-genderless-1238805610/ (Accessed: 23 May 2026).

Zhou, X., Xiao, C., Yoon, S. and Zhu, H. (2026) 'The colour of status: Color saturation, brand heritage, and perceived status of luxury brands', *Journal of Consumer Research*, 52(6), pp. 1232–1252. Available at: https://doi.org/10.1093/jcr/ucaf029 (Accessed: 23 May 2026).

Authors Bio:

Shenzimo Zhang is an MSc Marketing student in the School of Management at the University of Bristol. Her research interests focus on cultural branding, brand communities, and the symbolic value of brands in contemporary consumer culture. She has gained practical experience in advertising, esports business development, and brand communications, including working on communications and public relations projects for luxury beauty brands such as Valentino Beauty at L'Oréal. Her research explores how brands move beyond products to create cultural meaning, identity, and emotional connections with consumers.

Research on Green Transformation, Financing Constraints, and Corporate Performance

Author: Wenxu Yu
ORCID: 0009-0004-5067-3884
Programme: BScFinance and Accounting
Affiliation: Business School
Email: zg25304@bristol.ac.uk

Abstract

As globalconcern for environmental protection and sustainable development continuesto grow, green transformation has become an inevitable choice for enterprisesseeking sustainable development and has also attracted considerable attention from academia, industry, and policymakers. In view of this, this study takes 37,114 listed companies on the Shanghai and Shenzhen stock exchanges as the research sample, and theoretically analyses and empirically examines the impact of green transformation on corporate performance and its mechanism. The empirical results show that green transformation is positively associated with corporate performance, and financing constraints play a mediating role in this relationship. After a series of robustness tests, the above conclusions remain valid. This study aims to explore ways to resolve the tension between enterprise development and environmental protection, help enterprises shift away from the traditional development model characterised by high energy consumption and high pollution, thereby improving corporate performance and enabling enterprises to identify the endogenous motivation for fulfilling environmental responsibility. Meanwhile, this study provides a new perspective for understanding the economic consequences of corporate green transformation and offers useful references for policymakers and corporate managers.

Keywords:green transformation; financing constraints; corporate performance

Introduction

Growing attention to environmental protection and sustainable development has made green transformation an increasingly important issue for enterprises. In China, a series of policies, such as the green credit policy and the 2024 Opinions on Accelerating the Comprehensive Green Transformation of Economic and Social Development, have further encouraged firms to pursue greener development paths. Existing research generally suggests that green transformation may help enterprises improve resource allocation efficiency, reduce environmental-related costs and strengthen long-term market competitiveness, which could ultimately contribute to better corporate performance.

At the same time, scholars have paid increasing attention to factors affecting firms' ability to carry out green transformation. In particular, studies on ESG performance, environmental regulation, corporate social responsibility and information disclosure suggest that financing conditions may play an important role in supporting sustainable investment activities. Although previous literature has discussed the relationship between green transformation and firm performance, less attention has been given to how financing constraints may influence this relationship, especially in the context of Chinese listed companies.

Based on this research gap, this study focuses on companies listed on the Shanghai and Shenzhen stock exchanges to examine the relationships among green transformation, financing constraints and corporate performance. The empirical analysis uses data collected from the China City Statistical Yearbook, the China Environmental Statistical Yearbook, annual reports, corporate social responsibility reports and the Wind database. Corporate performance is measured by return on equity (ROE), green transformation is proxied by green total factor productivity (GTFP), and financing constraints are measured using the SA index.

To analyse these relationships, the study employs descriptive statistics, correlation analysis, a two-way fixed effects model and a mediating effects model. It further examines whether financing constraints may serve as a possible mechanism through which green transformation affects corporate performance. Robustness tests are also conducted to improve the reliability of the empirical results. Overall, the study aims to provide additional evidence on the economic implications of corporate green

transformation and offer practical insights for enterprises seeking sustainable development.

Literature Review and Research Hypotheses

Green Transformation and Corporate Performance

In recent years, as global concern for environmental protection and sustainable development continues to increase, corporate green transformation has become a research focus in both academia and practice. Porter and Van der Linde (1995) propose that a green strategy can bring competitive advantages to enterprises and break the deadlock between environmental regulation and economic development, arguing that environmental innovation can reduce corporate costs and create new market opportunities. Yang et al. (2013) are among the earlier scholars to examine the impact of green strategy on corporate performance, pointing out that in the context of a transition economy, green strategy may positively influence corporate performance by improving resource utilisation efficiency and market competitiveness. Li and Ramaswami (2018) focus on the relationship between different types of environmental regulation and corporate environmental performance. Taking China as an example, they find that stringent environmental regulation helps enterprises improve environmental performance and thus promotes sustainable development. Kalyar et al. (2020) emphasise that through green supply chain management practices and institutional pressure, enterprises can improve both environmental performance and financial performance, indicating that green transformation may contribute not only to environmental protection but also to corporate economic performance. Shah and Soomro (2021) examine the relationship between internal green integration and environmental performance, and point out that proactive environmental strategy, supplier greening, and environmental cooperation with suppliers have significant predictive power for improving corporate environmental performance.

However, some studies suggest that green transformation may impose substantial short-term adjustment costs and place pressure on corporate profitability during the transition period (Li and Ramaswami, 2018).

In summary, the impact of green transformation on corporate performance is

multifaceted. It involves both short-term cost input and long-term gains in benefits, while enterprises in different industries and of different sizes face different opportunities and challenges in the process of green transformation.

The Mediating Effect of Financing Constraints

In the field of research on the mediating effect of financing constraints, scholars have in recent years conducted in-depth discussions from different perspectives. Yi and Sun (2023) find that ESG performance can effectively alleviate corporate financing constraints by expanding financing channels. Good ESG performance helps enterprises obtain more opportunities for both equity financing and debt financing, thereby reducing the degree of financing constraints (Xiong and Liu, 2023). Xie et al. (2025) focus on the mechanism through which ESG information disclosure affects financing constraints.The study finds that ESG information disclosure may reduce the cost of debt financing and thereby exert a positive effect on the improvement of corporate performance, with debt financing cost playing a mediating role in this process . Based on empirical evidence from privately owned listed manufacturing companies in China, they find that greater fulfilment of corporate social responsibility and increased R&D investment can reduce the cost of debt financing and thereby alleviate financing constraints, indicating that investment in social responsibility and innovation has a significant mediating effect . These studies reveal the mediating mechanism of financing constraints from different dimensions and provide rich theoretical foundations and empirical support for understanding the paths through which corporate financing constraints can be alleviated.

Theoretical Analysis

The internal relationship among corporate green transformation, financing constraints and corporate performance is supported by rich theories. The impact of green transformation on corporate performance is based on the resource-based view and Porter hypothesis: the resource-based view holds that green transformation can improve resource utilisation efficiency and form unique resource advantages for enterprises; the Porter hypothesis points out that green transformation as an innovative response to environmental regulation can help enterprises develop green

products and open up new market space.

The mediating role of financing constraints can be explained by information asymmetry theory and pecking order theory: from the perspective of information asymmetry theory, green transformation enterprises pay more attention to information disclosure, reduce the information gap with investors, and make investors more willing to provide financial support, thus alleviating financing constraints; according to pecking order theory, although green transformation requires initial capital investment, it can reduce production costs and improve internal financing capacity in the long run, and green projects are supported by national policies, which can broaden external financing channels and optimise the financing structure of enterprises.

Research Hypotheses

Green transformation can improve enterprises' resource utilisation efficiency, reduce production costs, and meet the market demand for green products, expand market share and establish a good social image, thus improving corporate performance. Based on this, Hypothesis 1 is proposed:

H1: Green transformation is positively associated with corporate performance.

Green transformation enables enterprises to align with national policy orientation, improve operational stability and information transparency, send positive signals to external stakeholders, gain the favour of financial institutions and investors, and thus alleviate financing constraints. Based on this, Hypothesis 2 is proposed:

H2: Green transformation may help alleviate corporate financing constraints. Financing constraints affect the smooth implementation of green transformation: severe financing constraints will hinder green transformation and thus affect corporate performance, while alleviated financing constraints can promote green transformation and exert a positive effect on corporate performance. Based on this, Hypothesis 3 is proposed:

H3: Financing constraints may represent a potential pathway linking green transformation and corporate performance.

Research Design

Sample Selection and Data Sources

This study takes 37,114 listed companies on the Shanghai and Shenzhen stock exchanges as research samples, and the research data are obtained from China City Statistical Yearbook, China Environmental Statistical Yearbook, listed company annual reports, corporate social responsibility reports and Wind database, ensuring the consistency and reliability of data sources.

Variable Definitions

The explained variable is corporate performance (ROE), measured by return on equity; the explanatory variable is green transformation (GTFP), measured by green total factor productivity; the mediating variable is financing constraints (SA), calculated by the formula $SA = -0.737Size + 0.043Size^2 - 0.04Age$. The control variables include asset-liability ratio (Lev), firm size (Size), total asset turnover (ATO), cash flow ratio (Cashflow), inventory ratio (INV) and fixed asset ratio (FIXED), all of which are obtained from the Wind database except for the calculated variables.

Table 1. Variable Definitions and Data Sources

Category	Variable	Measure	Data Source
Explained variable	ROE	Return on equity	Wind database
Explanatory variable	GTFP	Green total factor productivity	Calculated
Mediating variable	SA	Financing constraints	Calculated
Control variable	Lev	Asset–liability ratio	Wind database
	Size	Firm size	Wind database
	ATO	Total asset turnover	Wind database
	Cashflow	Cash flow ratio	Wind database
	INV	Inventory ratio	Wind database
	FIXED	Fixed asset ratio	Wind database

Model Specification

Baseline Regression Model

The main focus of this study is the impact of green transformation on corporate performance. Accordingly, the first part examines the direct effect of green transformation on corporate performance. This study employs a two-way fixed effects model for the empirical analysis, and the model is specified as follows:

$$ROEit=\alpha0+\alpha1GTFPit+\alpha iControlit+\varepsilon i+yt+oit \quad (1)$$

Where ROEit denotes the explained variable, namely corporate performance; GTFPit denotes the explanatory variable, namely green transformation; Controlit denotes the control variables; εi denotes the individual dummy variable; yt denotes the time dummy variable; oit denotes the random disturbance term; α0 denotes the constant term of the explanatory variable; α1 denotes the coefficient of the explanatory variable; and αi denotes the coefficients of the control variables.

The two-way fixed effects model controls for differences across firms that do not change over time, as well as year-specific shocks that may affect all firms simultaneously. By including firm fixed effects and year fixed effects, the model aims to reduce estimation bias arising from unobserved firm characteristics and broader macroeconomic changes. However, the model cannot fully address time-varying omitted variables, reverse causality, or simultaneity between green transformation and corporate performance. Therefore, the findings should be interpreted as conditional associations rather than strict causal estimates.

Mediating Effect Model

According to the explanation of mediating effects provided by Wen Zhonglin and Ye Baojuan (2014), the stepwise regression method is the most stringent approach for testing mediating effects, and its results are also more accurate. The first step of the stepwise regression method is the same as that of the baseline model. In the second step, the mediating variable SA is regressed as the explained variable. In the third step, the mediating variable SA is incorporated into the baseline model for regression analysis. The model is specified as follows:

$$SA_{it}=\beta_0+\beta_1 GTFP_{it}+\beta_i Control_{it}+\tau_i+g_t+\vartheta_{it} \quad (2)$$

$$ROE_{it}=\rho_0+\rho_1 SA_{it}+\rho_2 GTFP_{it}+\rho_i Control_{it}+n_i+f_t+\pi_{it} \quad (3)$$

where SA_{it} denotes the mediating variable; τ_i and n_i denote the individual dummy variables; g_t and f_t denote the time dummy variables; and ϑ_{it} and π_{it} denote the random disturbance terms.

Empirical Results and Analysis

Descriptive Statistics

Table 2 presents the descriptive statistics of the key variables in this study, including the number of observations, mean, standard deviation, minimum value, and maximum value. Among the 37,114 observations, the mean value of corporate performance (ROE) is 0.065, with a standard deviation of 0.113, indicating relatively small variation in corporate profitability across the sample. The mean value of green total factor productivity (GTFP) is relatively high at 1.003, with a standard deviation of 0.123, suggesting that the achievements of enterprises in technological progress and resource allocation optimisation are relatively concentrated. The mean value of financing constraints (SA) is -3.838, with a standard deviation of 0.258, indicating considerable differences in financing capacity among enterprises. The mean asset–liability ratio (Lev) is 0.431, suggesting that the average level of corporate indebtedness is moderate, while the negative minimum value of the cash flow ratio (Cashflow) reflects the profitability challenges faced by some enterprises.

Table 2. Descriptive Statistics

Variable	N	mean	min	max	Sd
ROE	37114	0.065	-0.962	0.414	0.113
GTFP	37114	1.003	0.720	1.255	0.123
SA	37114	-3.838	-5.728	-2.300	0.258
Lev	37114	0.431	0.027	0.925	0.199
Size	37114	22.207	19.406	26.440	1.241
ATO	37114	0.651	0.055	3.106	0.442

Cashflow	37114	0.049	-0.224	0.283	0.069
INV	37114	0.145	0.000	0.778	0.132
FIXED	37114	0.219	0.002	0.765	0.161

Correlation Analysis

The correlation analysis results in Table 3 reveal the strength and direction of the linear relationships among the various financial indicators involved in this study. Green transformation (GTFP) is significantly positively correlated with corporate performance (ROE), with a correlation coefficient of 0.057***. This indicates that as the degree of corporate green transformation deepens, corporate performance also tends to improve. This preliminary relationship is consistent with the possibility that green transformation may be associated with improved competitiveness and profitability.

Asset–liability ratio
(Lev) are significantly negatively correlated with green transformation (GTFP), with a coefficient of -0.082***, and are also negatively correlated with corporate performance (ROE), with a coefficient of -0.082***. This indicates that firms with higher leveragetend to exhibit lower levels of green transformation and corporate performance.

Table 3. Correlation Analysis

	ROE	GTFP	Lev	Size	ATO	Cashflow	INV	FIXED
ROE	1							
GTFP	0.057***	1						
Lev	-0.082***	-0.082***	1					
Size	0.145***	0.192***	0.453***	1				
ATO	0.307***	-0.059***	0.132***	0.025***	1			
Cashflow	0.383***	0.051***	-0.152***	0.063***	0.190***	1		
INV	0.003	-0.099***	0.217**	-0.006	0.215***	-0.193***	1	

FIXED	-0.089***	-0.110***	0.048***	0.015***	0.112***	0.244***	-0.199***	1

Regression Analysis

Analysis of the Baseline Regression Results

Based on the baseline model constructed in Model (1), this section examines the relationship between corporate green transformation and corporate performance. The estimation results are reported in Table 4. Before conducting the regression analysis, this study first performs a Hausman test. The test results show that the p- value is less than 0.05, and the fixed effects model is therefore selected for the regression analysis. At the same time, firm fixed effects and year fixed effects are also controlled for in the fixed effects model . The regression results in Table 4 show that green transformation is positively associated with corporate performance to a certain extent. In column (1), the coefficient of corporate green transformation is significantly positive at the 5 per cent level. After additional control variables are included, the coefficient of corporate green transformation in column (2) remains significantly positive at the 5 per cent level, although the regression coefficient becomes smaller. This indicates that, at the enterprise level, there are other factors affecting corporate performance. It is therefore necessary to include additional control variables on the basis of the original model, so as to better estimate the conditional relationship between green transformation and corporate performance. Taking column (2) as an example, for every one-unit increase in corporate green transformation, corporate performance increases by 0.059 units, indicating that green transformation is significantly positively associated with corporate performance. While the estimated relationship is statistically significant, the magnitude of the coefficient remains relatively modest, suggesting that green transformation is only one of multiple factors associated with corporate performance.

Table 4. Regression Results of the Baseline Model

Variable	ROE	ROE
GTFP	0.079**	0.059**

	(0.032)	(0.029)
Lev		-0.227***
		(0.005)
Size		0.034***
		(0.001)
ATO		0.121***
		(0.002)
Cashflow		0.287***
		(0.009)
INV		0.066***
		(0.008)
FIXED		-0.132***
		(0.007)
Constant	0.048	-0.650***
	(0.034)	(0.040)
Year	Yes	Yes
Industry	Yes	Yes
N	37114	37114
R2	0.031	0.204

Note: Standard errors in parentheses， *** p<0.01, ** p<0.05, * p<0.1

Analysis of the Regression Results of the Mediating Effect Model

Based on Models (2) and (3), a mediating effect model is constructed to test the mediating role of financing constraints in the relationship between corporate green transformation and corporate performance. As shown by the regression results in Table 5, column (1) reports the effect of corporate green transformation on financing constraints. The regression coefficient of corporate green transformation is 0.033 and is significantly negative at the 10 per cent significance level. Column (2) reports the potential role of financing constraints in the relationship between green transformation and corporate performance. The regression coefficient of corporate green transformation is 0.058 and is significantly positive at the 5 per cent significance level, while the regression coefficient of financing constraints is 0.028

and is significantly negative at the 1 per cent significance level, indicating that the results provide suggestive evidence of a mediating relationship involving financing constraints and that this effect passes the Sobel test. These findings provide suggestive evidence that financing constraints may represent a potential pathway linking green transformation and corporate performance.

Table 5. Regression Results of the Mediating Effect Model

Variable	SA	ROE
GTFP	-0.033*	0.058**
	(0.018)	(0.029)
SA		-0.028***
		(0.009)
Lev	-0.029***	-0.228***
	(0.003)	(0.005)
Size	0.000	0.034***
	(0.001)	(0.001)
ATO	-0.014***	0.121***
	(0.001)	(0.002)
Cashflow	0.013**	0.288***
	(0.005)	(0.009)
INV	-0.008	0.066***
	(0.005)	(0.008)
FIXED	0.033***	-0.131***
	(0.004)	(0.007)
Constant	-3.464***	-0.746***
	(0.025)	(0.050)
Year	Yes	Yes
Industry	Yes	Yes
N	37114	37114
R2	0.893	0.204

Note: Standard errors in parentheses， *** p<0.01, ** p<0.05, * p<0.1

Robustness Tests

This study conducts robustness tests by replacing the measurement of the explanatory variable. With reference to the studies by Zhou et al. (2022), 113 keywords related to corporate green transformation are selected from multiple dimensions. The frequency of each keyword appearing in the annual report texts of listed enterprises is then counted to form the frequency measure of green transformation terms. The natural logarithm of this frequency plus one is used to characterise corporate green transformation. As shown in Table 6, the signs and significance levels of the coefficients of the explanatory variable and the mediating variable are consistent with those obtained in the baseline regression results and the mediating effect tests, providing additional support for the stability of the observed relationships.

Table 6. Robustness Test Results

Variable	ROE	SA	ROE
GTRAN	0.004***	-0.001**	0.004***
	(0.001)	(0.001)	(0.001)
SA			-0.028***
			(0.009)
Lev	-0.227***	-0.028***	-0.228***
	(0.005)	(0.003)	(0.005)
Size	0.034***	0.001	0.034***
	(0.001)	(0.001)	(0.001)
ATO	0.121***	-0.014***	0.120***
	(0.002)	(0.001)	(0.002)
Cashflow	0.287***	0.013**	0.287***
	(0.009)	(0.005)	(0.009)
INV	0.066***	-0.008	0.066***
	(0.008)	(0.005)	(0.008)
FIXED	-0.133***	0.033***	-0.132***
	(0.007)	(0.004)	(0.007)

Constant	-0.605***	-3.489***	-0.702***
	(0.033)	(0.021)	(0.045)
Year	Yes	Yes	Yes
Industry	Yes	Yes	Yes
N	37114	37114	37114
R2	0.204	0.893	0.204

Discussion

Limitations in Data Coverage and Measurement Bias

First, the research sample is only limited to listed companies on the Shanghai and Shenzhen stock exchanges, excluding small and medium-sized enterprises that account for more than 99% of China's total enterprises. Small and medium-sized enterprises face more severe financing constraints and have fewer green transformation resources, and their unique transformation mechanisms are not included in the study, resulting in limited explanatory power of the conclusions for small and medium-sized enterprises. Second, the measurement of core variables is simplified: green transformation is measured only by GTFP, which cannot distinguish between passive compliance-based transformation and proactive innovation-driven transformation with very different performance effects; financing constraints are measured only by the SA index, ignoring carbon account assets, green patent pledge capacity and other key factors, leading to estimation bias of the mediating effect. In addition, the lag and incompleteness of environmental data disclosure, such as the low proportion of listed companies disclosing carbon emissions and the two-year lag of environmental statistical data, make the research conclusions unable to capture the latest policy effects and may underestimate the recent effect of green transformation in alleviating financing constraints.

Endogeneity Risks in the Research Design and Ambiguity of the Mechanism

First, the research model cannot fully solve the endogeneity problem, and there is a two-way causal relationship between green transformation and financing constraints: the improvement of corporate performance will also increase the enterprise's

investment capacity in green transformation, and the current two-way fixed effects model cannot eliminate this reverse causality, which may overestimate the positive effect of green transformation on performance. Second, the study treats financing constraints as a single dimension and does not distinguish the different mediating mechanisms of different financing channels such as green credit, green bonds and supply chain finance, resulting in the lack of specificity of policy recommendations. In addition, the study does not conduct industry heterogeneity analysis, and high-pollution and low-pollution industries have significant differences in green transformation costs, performance improvement paths and financing methods. The lack of grouped tests makes the cross-industry policy recommendations too uniform and weakens the practical value of the conclusions.

Conclusion

The Direct Effect of Green Transformation on Corporate Performance: Green transformation may require substantial capital investment in the short term and may therefore place certain pressure on corporate profitability. In the long run, however, green transformation may contribute to improved corporate profitability by increasing resource utilisation efficiency and reducing operating costs (Zhu and Chen, 2024). The Mediating Role of Financing Constraints: Green transformation may help alleviate corporate financing constraints, and reduced financing constraints may be associated with improved corporate profitability. This provides suggestive evidence that financing constraints may represent a partial pathway linking green transformation and corporate performance. In addition, the robustness test results based on the Bootstrap method show that the mediating effect of financing constraints is robust, which further validates the findings of this study.

In summary, the findings suggest that green transformation may contribute to both environmental and financial performance under certain conditions by alleviating financing constraints. When carrying out green transformation, enterprises should therefore fully consider the impact of financing constraints and actively seek diversified financing channels, so as to achieve a win-win outcome between green transformation and corporate performance.

References

Fang, D., Wang, P. & Cao, G., 2024. The collaborative evolution mechanism of green innovation and performance in the power sector under social networks. Journal of Management Sciences in China, 27(10), pp.35-52.

Fang, H., Xie, H. & Zhao, Q., 2024. The establishment of big data comprehensive pilot zones, the empowerment of data factors, and corporate green transformation.World Economy Studies, 11, pp.93-107.

Gong, J., Lan, X., Hu, Z. et al., 2024. The effectiveness, limitations, and policy optimisation of green finance in promoting corporate green transformation: Evidence from the Yangtze River Economic Belt. South China Journal of Economics, 10, pp.28-52.

Hua, Y., Liu, Y. & Lu, J., 2025. Green credit policy and the performance of polluting enterprises: From the perspectives of investment and financing constraints and internal control. Technology Economics, 44(1), pp.140-154.

Jiang, H., Han, Y. & Qin, Y., 2024. Research on the interactive relationship between digital transformation and green transformation in manufacturing. Ecological Economy, 40(11), pp.47-52.

Kalyar, M.N., Shoukat, A. & Shafique, I., 2020. Enhancing firms' environmental and financial performance through green supply chain management practices and institutional pressures. Sustainability Accounting, Management and Policy Journal, 11(2), pp.451-476.

Lei, Z. & Song, J., 2025. How does local fiscal expenditure promote the green transformation of industrial structure? From the perspective of developmental local government. Finance and Accounting Research, 2, pp.3-14.

Li, R. & Ramaswami, R., 2018. Exploiting the relationships between different types of environmental regulations and environmental performance: Evidence from China.

Journal of Cleaner Production, 196, pp.1329-1340.

Liu, G., Zhao, J. & Chen, T., 2024. A brief analysis of the path through which coal-hydrogen industrial co-ordination supports the green transformation of the coal industry under the background of new quality productive forces. China Coal, 50(12), pp.25-32.

Liu, S., Wang, M. & Qi, J., 2024. Green investment, financing constraints, and corporate financial performance: Empirical evidence from listed enterprises in the energy industry. Statistics and Decision, 40(6), pp.172-177.

Liu, W., Li, Y. & Bi, K., 2024. Research on the impact of green technology integration on the asymmetric catch-up of latecomer firms: Based on the perspective of dynamic competition and co-operation. Science Research Management, 45(8), pp.95-105.

Long, X. & Deng, X., 2024. The synergistic effect of corporate social responsibility and competitive strategy. Management Review, 36(10), pp.185-198.

Lu, Q. & Lu, X., 2025. Research on the green and low-carbon transformation path of coal-fired power enterprises in Shaanxi Province. Journal of Safety and Environment, 25(2), pp.779-789.

Mao, Q. & Shi, B., 2024. The road towards green development: Intelligent manufacturing and corporate green transformation. The Journal of World Economy, 47(9), pp.152-182.

Nie, M. & Zhang, X., 2025. The impact of green technological innovation on corporate financial performance under the “dual carbon” goals: Empirical evidence from heavily polluting manufacturing industries. Finance Theory and Practice, 46(2), pp.90-98.

Porter, M. & Van der Linde, C., 1995. Green and competitive: Ending the stalemate. In: The Dynamics of the Eco-efficient Economy: Environmental Regulation and Competitive Advantage, 33, pp.120-134.

Shah, N. & Soomro, B.A., 2021. Internal green integration and environmental performance: The predictive power of proactive environmental strategy, greening the supplier, and environmental collaboration with the supplier. Business Strategy and the Environment, 30(2), pp.1333-1344.

Shang, Y., 2025. Strategic thinking and practical paths for promoting the comprehensive green transformation of economic and social development. Commercial Culture, 6, pp.130-133.

Sun, A. & Mo, F., 2025. Reflections on innovating the value realisation mechanism of ecological products from the perspective of Chinese-style modernisation: A case study of Inner Mongolia's green transformation and development model. Price Monthly, 1, pp.9-16.

Wang, J., Li, M., Jiang, Z. et al., 2025. Research on the supply chain spillover effect of corporate green technological innovation. Journal of Management Sciences in China, 28(1), pp.77-99.

Wang, K., Sun, Z. & Chen, P., 2024. The impact of digital transformation on corporate financial performance: The mediating and moderating roles of dynamic capabilities. Management Review, 36(12), pp.249-262.

Wang, L. & Li, X., 2024. New models and strategies for green development under the background of the digital economy. International Public Relations, 22, pp.28-30.

Wen, Z. and Ye, B., 2014. Analyses of mediating effects: The development of methods and models. Advances in Psychological Science, 22(5), pp.731–745.

Xie, L., Peng, Z., Tong, X. and Zhou, W., 2025. ESG disclosure and financing constraints: Evidence from Chinese listed firms. Economic Analysis and Policy, 89, pp.742–756.

Xiong, Z. & Liu, J., 2023. Research on the impact of fintech on green innovation in

manufacturing enterprises. Chinese Journal of Management, 21(12), pp.1812-1819.
Yang, J., Shi, J., Li, M. and Liu, J., 2013. How does green strategy influence corporate performance? Evidence from the context of transition economies. Science of Science and Management of S.& T., 34(7), pp.141–149.

Yi, X. and Sun, H., 2023. Can ESG performance effectively alleviate corporate financing constraints? Evidence from financing channels. Finance and Economics, 7, pp.65–75. https://doi.org/10.19622/j.cnki.cn36-1005/f.2023.07.006

Zhang, C. & Hu, Z., 2025. Research on the industrial chain linkage effect of enterprises' co-ordinated digital and green transformation: Based on the coupling co-ordination degree model and the perspective of social division of labour and collaboration networks. Journal of Anhui University (Philosophy and Social Sciences Edition), 49(2), pp.154-164.

Zhou, D., Zhai, L., Hu, S. et al., 2024. Digital servitisation in manufacturing enterprises: Connotation analysis, research evolution, and the construction of an integrated framework. Collected Essays on Finance and Economics, 7, pp.91-101.

Zhou, K., Wang, R., Tao, Y. et al., 2022. Corporate green transformation and stock price crash risk. Management Science, 35(6), pp.56–69.

Zhu, S. & Chen, W., 2024. Green product imports and corporate green transformation: Evidence from Chinese listed companies. World Economy Studies, 8, pp.75-89+136-137.

Zou, Y. & Xiao, Z., 2024. Research on the impact of ESG greenwashing on corporate performance. Contemporary Finance & Economics, 11, pp.152-164.

Authors Bio:

My name is Wenxu Yu, and I am currently pursuing an MSc in Management at the University of Bristol Business School. Throughout my studies, I have developed a strong interest in sustainable development, corporate strategy, and environmental governance. As global environmental challenges continue to intensify, I believe that green transformation has become an essential pathway for enterprises to achieve long-term competitiveness while contributing to social and environmental well-being. My research focuses on the relationship between green transformation, financing constraints, and corporate performance, exploring how sustainable business practices can generate both environmental and economic value. I am particularly interested in understanding how firms can balance profitability with environmental responsibility, and how financial mechanisms can support this transition. Studying these issues has strengthened my belief that sustainable business development is not only necessary for addressing environmental challenges but also represents an important source of innovation and growth for modern enterprises.

Education andDigital Society

Meet the Subject Editor

Anna Shahbazyan
Education and Digital Society

Anna Shahbazyan is currently pursuing a Master's degree in Education Policy and International Development at the University of Bristol. With previous editorial experience across several undergraduate literary magazines, she was excited to join the Student Research Journal as an Education, Policy, and Arts Subject Editor. The role has provided her with the opportunity to engage with a wide range of interdisciplinary scholarship and gain insight into the diverse and innovative research being undertaken by students across disciplines. Her academic interests include multilingual education, refugee and migrant education, educational inclusion, and language policy implementation. She hopes to continue working in the field of education, contributing to the development of policies that promote equitable and inclusive learning opportunities for students around the world.

Student Feedback Literacy from a Critical Perspective: A Neurodivergence-Affirming Review

Author: Nathaniel Joyce

ORCID: 0009-0009-3944-6150

Programme: MSci Biology

Affiliation: SchoolofBiological Sciences

Email: fx22495@bristol.ac.uk

Abstract

A feedback literate student possesses the competencies necessary to make effective use of feedback. The framework of student feedback literacy has informed institutional interventions to improve student progress in higher education. However, current conceptualisation of student feedback literacy is yet to consider minority perspectives, including those of neurodivergent people. It is therefore unclear how minoritised students' experiences with developing feedback literacy differ. Synthesising neurodivergent students' perspectives with sociocultural and political understandings of feedback, this critical review analyses implicit power imbalances, discrimination, and normative expectations in the feedback literacy framework. The review focuses on the conception of making judgements and managing affect, critiquing the implicit objective conception of evaluative judgement and individualisation of responsibility for affective response. The review proposes two concepts in the student experience of feedback literacy: Evaluative Masking as a behaviour suppressing authentic academic expression, and Constructive Defensiveness as a misunderstood but valid response to feedback. Future research directions are noted to facilitate reconceptualisation of the framework to create a neurodivergent-affirming understanding of student feedback literacy. This review lays the basis for future minority perspectives to be incorporated into feedback literacy, and the adoption of practices that will make developing student feedback literacy more accessible and equitable.

Key Words Neurodiversity, Feedback Literacy, Evaluative Judgement, Managing Affect, Inclusive Education

Thank you to Sheila-Amici Dargan for supervising me with this review.

Introduction

Undergraduate students remain unsatisfied by the feedback they receive from markers on assessments, and this acts as a barrier to student progress (Boud & Molloy, 2013). Alongside improving the content of feedback, improving student uptake is crucial. One recent conceptual advance considering this is the development of the student feedback literacy framework, which recognises core competencies of a learner who effectively uses feedback (Carless & Boud, 2018). For effective action-taking: a student must appreciate the feedback process, make suitable judgements on quality, and manage affect in the feedback process. However, minority perspectives on feedback are lacking, with homogeneity of cohorts assumed even in studies examining learner sociocultural context (Xu & Carless, 2017). In not recognising these perspectives, the feedback process can be unsupportive to specific needs, reducing psychological safety.

Incorporating neurodiverse perspectives on feedback is critical. Given the breadth of neurodivergence – incorporating ADHD, autism, dyslexia and mental health conditions, it is likely experiences of feedback are heterogeneous. However, across these cognitive profiles neurodivergent students report poor experiences in higher education (McDermott & Mosley, 2023). There is increasing recognition that institutional policy must be built with neurodivergent needs centred (Kennedy et al., 2025), and this should include feedback processes, which can be designed based on neurodivergent-affirming conceptions.

This review critiques feedback literacy from a neurodivergence-affirming perspective. It aims to synthesise literature from pedagogy and neurodiversity studies to critique assumptions and biases underpinning the current framework. This theoretical 2econceptualization will support inclusive principles in feedback research and implementation. The review covers two competencies associated with feedback literacy as described by Carless & Boud (2018) – Evaluative Judgement and Managing Affect. Each is analysed for conflicts with broader concepts in pedagogy and inclusivity studies. The effect of these conflicts on the neurodivergent student's learning is explored, alongside future research and policy directions.

Evaluative Judgement

Improving evaluative judgement facilitates actioning on feedback through refinement of work according to internal judgements (Boud & Carless, 2018). In considering the current barriers, scholars have recommended more opportunities for self-evaluation and deeper engagement with marking criteria. While opportunities for understanding criteria are likely to benefit neurodivergent students significantly (Bhandari & Rainford, 2023), it is important to recognise the unique barriers faced by neurodivergent students (Kennedy et al., 2025). Development of evaluative judgement requires the acquisition of tacit knowledge (Carless & Boud, 2018). This is often conveyed poorly to neurodivergent students, especially in more subjective subjects where markers often disagree (Bloxham et al., 2016). Moreover, neurodivergent students may possess special interests that reshape their judgement of quality (Boven, 2018) Taking neurodivergence as a diverse but valid sociocultural context, this section therefore aims to identify how current understandings of evaluative judgement may take an overly narrow view of its development and presentation.

Implicit Objectivisation of Quality

Despite the conceptualisation of feedback literacy as socioculturally grounded (Gravett, 2022), research often treats evaluative judgement as an objective skill (e.g. Tai et al., 2018). However, this risks sidelining accepting alternative but valid evaluations of work. Evaluative judgement is built on the quality against which it is practiced. However, instead of challenging how quality came to be conceptualised, many trust consensus views (Sadler, 2013). Feedback literacy as conceptualised implicitly trusts the consensus, and focuses on calibration and accuracy over developing an individual academic perspective and expression of quality (Carless & Boud, 2018).

Conceptualisation of quality as objective has shaped empirical methods. For example, Ibarra-Saiz et al. (2020) measured project management students' evaluative judgements relative to the consensus-based QUASS metric, rather than accepting the multiple valid styles of project management. Implicitly, this quantification frames minority perspectives as deficient, even when well-justified.

Although accuracy framing is beginning to be critiqued (Fernandez-Toro & Duensing, 4 2021), this is yet to influence the conception of feedback literacy or be applied to disciplines traditionally viewed as more standards-based. In failing to search for alternative perspectives on quality, research is yet to understand how identity influences university students' work. This may curtail the ideas generation and epistemic insights expressed by many neurodivergent students (White & Shah, 2011; Dror, 2022). We may therefore recognise that without active consideration, neurodivergent perspectives are at risk of being overlooked.

Dialogue and Identity

If we accept evaluativejudgement as identity-based, discussion is critical to develop broader appreciationsofjudgements This aligns with the increased interest in dialogue in feedback(Steen-Utheim et al., 2017). Dialogue can challenge the teacher-student hierarchy, improving trust and encouraging feedback-seeking behaviour (Price et al., 2011). Further, in challenging a consumption-based unidirectional view of feedback, we recognise dialogue as an opportunity for selfreflection and academic transformation. However, for truly open and non-hierarchical dialogue, efforts should be made to encourage constructive disagreement and bidirectional uptake of insights. Often, dialogue spaces can silence full expression of student voice (Hookway, 2012). More open dialogue is an opportunity to develop evaluative judgement in students through encouraging justification of their perspectives. This could provide space for dialectics, creating perspectives of quality reflective of both academic and student perspectives. More developed academic literacies have been more successful in actively challenging top-down standardisation in favour of actively seeking different student manifestations (Lea & Street 1998). If feedback literacy research followed this approach, dialogue becomes a bidirectional space to explore quality, and to genuinely deconstruct the hierarchies that reinforce unjust conceptions of quality (Hookway, 2012). This makes dialogue an incredibly powerful way to recognise the often missed contributions of neurodivergent people to the academic space and exploring expressions that need greater recognition (Bailey, 2023).

Calibration and Conformity

By creating psychologically safe dialogue spaces, we can ensure students feel confident voicing their perspective (Gibbs, 2006). However, communication of standards as objective and fixed may implicitly undermine trust, even when aimed at justifying institutional policies. Such a cognitive environment rewards conformity, not transgression, which while useful in some practical concepts, may inhibit expression of a student's academic voice (Hookway, 2012). Safer environments would more strongly challenge conformity and reward unique academic expressions. Conformity in subjective academic settings rewards safety over risk-taking. An objectivity-based environment allows many students develop confidence in their ability to calibrate to standards (Little et al., 2023). Yet when neurodivergent perspectives are not validated, they may feel less confident than peers (Cai & Richdale, 2015; Buckley et al., 2024), including when they are shown to be no different in ability (Mahak et al., 2025). There may therefore exist an implicit pressure for students to develop internal feedback that suppresses under-recognised cognitive skills. This is particularly relevant for neurodivergent students (White & Shah, 2011; Milton, 2014). In dialogue and peer settings, their strengths may be under-recognised, and therefore unvoiced. This implicitly reinforces the neurodivergent insecurity that their cognitive profile is deficient (Cai & Richdale, 2015), rather than a potential strength to be shared. To open itself to true dialogue, research and practice must therefore actively work to avoid creation of cognitive environments that pressure conformity. One way researchers have started to recognise more nuanced expressions of student judgement, even in corrective settings, is through greater appreciation of the role of metacognition (Kolovelonis, 2023; Clark et al., 2025). Metacognitive students, aware of both the quality of work and thought processes behind it, may more clearly distinguish between when feedback reflects epistemic injustice (Fricker, 2007).

Evaluative Masking

Neurodivergent students will be familiar with having to suppress their behaviours to conform to societal standards, known as masking. There is growing recognition that 6 masking can be applied beyond social interaction to viewpoints and perspectives (Chapman et al., 2022). Perhaps a neurodivergent student's judgement process can

be perceived not as evaluative judgement but as evaluative masking – suppression of academic voice in fear it would be poorly received (Boven, 2018). Evaluative masking may suppress exploration of focused interests, which could critically harm neurodivergent student's motivation (Milton & Sims, 2016). Intense interests and pursuits clash with a system that often considers undergraduate hyperfocus a deficit (Hua et al., 2012). The masking student is pressured to broaden or abandon their interests, harming their wellbeing (Grove et al., 2018). Though not yet applied to evaluative judgement, the significant toll of masking in higher education is increasingly recognised (Syharat, 2023). This calls for the redesigning of standards to accept focused interests as valid, and as an opportunity to motivate neurodivergent students (Boven, 2018).

Managing Affect – A Critical Lens

Carless and Boud (2018) identify management of affect as a core competency of feedback-literate students. However, given the inherent sociocultural placing of affective response and management, considerably little literature focuses on what constitutes constructive affective response. This is particularly noticeable in contrasting the attention given to teacher affect (Carless & Winstone, 2023). Given the increased proximity education researchers have with educators, this likely represents an empathy gap that creates an inherent theory-practice rift in understandings of affect. Recognising the emotional and political factors underpinning feedback, this passage examines how under-exploration of affect in feedback may implicitly reinforce normative expectations distant from neurodivergent experience. It moves past viewing affective response as a study skill as traditionally theorised (Carless & Boud, 2018) and suggests that "poor affect management" (Winstone & Carless, 2019) itself is not the barrier to improvement, but poor institutional support for diverse affective needs.

Management of Affect may be Disproportionally Relevant for Neurodivergent Students

One factor potentially explaining the under-exploration of management of affect is that it is generally not seen as a barrier to feedback literacy (Winstone & Carless, 7 2020). This is justified by student surveys that show typically only 10% of respondents view report challenges in managing affect. These learners are often dismissed as “low achieving”, with the implicit idea that these learners are less capable of developing feedback literacy competencies more generally (Winstone & Carless, 2020). Little attention has been paid as to how non-response bias or feelings of shame in these optional surveys may shape the low proportion of students reporting affect as a barrier. Moreover, from a critical pedagogy angle, it is crucial to remember that those 10% of students represent a systemic failure, not an individual deficit. Minoritised groups often face greater rejection sensitivity to feedback (London et al., 2012), this is particularly relevant for neurodivergent students where rejection sensitivity stems from both trauma-based and neurobiological sources (Müller & Pikó, 2024). When we synthesise these psychological sensitivities with the negative impact on wellbeing of higher education of neurodivergent students (McDermott & Mosley, 2023; Kennedy et al., 2025), we recognise the disproportionate difficulty neurodivergent students face to present idealised affective responses. Negative emotional responses to feedback may thus not be universally seen as opposition to improvement as traditionally viewed (Winstone & Carless, 2019), but as a valid reaction rooted in social and contextual factors. The theory that neurodivergent students may experience stronger affective responses to feedback aligns with neurodivergent students’ voices. Neurodivergent students generally find higher education involves a great deal of emotional labour (Bhandari & Rainford, 2023, Jackson et al., 2017), including to feedback they perceive as insensitive (Syharat et al., 2023). One could therefore argue that affective expectations can be implicitly ableist, expecting neurodivergent students to present emotional responses that misrepresent the pressures they face. A neurodivergent student can face the stress of masking their feelings, or the social and academic rejection from refusing to suppress their emotional response (Chapman et al, 2022).

Identifying Institutional Responsibility

Recognising affective response as sociopolitical, we understand that individuals may have limited autonomy to independently shape healthy affective expression to 8 feedback. Yet this is not fully recognised in current feedback literacy literature, where little attention is paid to how institutions can aid students in managing affect in feedback (Carless and Boud, 2018). Instead, conceptions have implicitly responsibilised the self, reflecting the prevailing neoliberal attitude on emotional regulation common in Western society (Trnka & Trundle, 2014). This is reflective in the exemplification of students independently managing their own affect in response to negative feedback (Winstone & Carless, 2019) through students whose survey responses fit idealised conceptions of reflectiveness and affective management (Shields, 2015). This is an inherent extrapolation: to assume that one student's emotional cognition is reflective of what all should aim towards. From this normative perspective, alternative management of affect may be viewed as a deficit. This strongly contrasts with views on affect from neurodivergent students and the wider student body: as reflective of cognitive differences and not as an individual deficits to address (Molloy et al., 2020; Bhandari & Rainford, 2023; Syharat et al., 2023). This aligns much more closely with the sociocultural situation of feedback literacy that scholars often cite (Nieminen & Carless, 2023). The rift between theorisation and student perspective and experience indicates that individualisation of affect has failed to develop student feedback literacy. This can threaten the creation of psychologically safe spaces (Carless & Young, 2024). Therefore, before we begin to consider how affective scaffolding can be universally designed, we must first explicitly and thoughtfully consider the institutional responsibility of affective support in feedback.

Institutional Responsibility and Identity Although management of affect is still generally viewed as a skill over a sociopolitical construct, some authors do accept that institutions bear some responsibility to support affect. For example, Xu and Carless (2017) praise the consideration of cultural perceptions and teacher-student relationship in mediating anxieties in the feedback process for Chinese students. This recognition of best practice is positive, but generalised the challenges to those faced across the entire cohort, clashing with evidence that neurodivergent students face not only intensified challenges to the neurotypical cohort, but also different

challenges (Kennedy et al., 2025). Authors acknowledge the cultural differences between Chinese and Anglosphere classrooms 9 (Xu & Carless, 2017) but do not extend this to the cognitive and cultural heterogeneity in Western institutions. As secondary education becomes more inclusive, more neurodivergent students attempt academic degrees (ACDET, n.d.), increasing the diversity of lecture theatres. Therefore, we may not only view the institution as responsible for whole-cohort wellbeing, but also for proactively supporting individual needs, aligning with universal design principles.

Without recognising the different affective needs in the classroom, instructors risk implicitly reinforcing deficit framings, a stigmatisation increasingly recognised as prevalent in education (Jense van Rensburg & Liang, 2025). This stigmatisation reduces neurodivergent students' confidence in asking for support (Cai & Richdale, 2015), reducing feedback-seeking and support-seeking behaviours. This is in contrast to an expressed desire of neurodivergent students to feel more confident in these elements (Syharat et al., 2023). Future work must therefore move beyond identity-blind empirical studies and policies (Little et al., 2023), and centre these as contextual factors reshaping a learner's affective needs.

It cannot be fully theorised how neurodivergent students' affective management processes can be fully supported and accommodated in feedback, as empirical research is entirely lacking. However, the first step for theory and practice is a conceptual shift from a deficit-based to an emotionally-affirming understanding of the affective challenges of neurodivergence in education and feedback (Rowney-Smith et al., 2024; Kroll et al., 2025). This extends work that already recognises feedback literacy's manifestation could look different for disabled students (Nieminen et al., 2022) but moves to specifically recognise and root out additional challenges related to identity, discrimination and rejection sensitivity. Framed this way, we can begin open conversations with neurodivergent students to understand their experiences, and adapt feedback policy following inclusive best practice in tertiary education (Bhandari & Rainford, 2023). Relational aspects also likely shape neurodivergent affective management as they do with students more generally (Ellis, 2010; Xu & Carless, 2017), and increased emphasis on relations in feedback, particularly between peers who share identities, may share and reduce emotional labour (Simonsmeier et al., 2020). Together, this recognises that the institution's

responsibility is not only to calm whole-cohort fears, but to proactively prepare feedback spaces to be inclusive of neurodivergent and minoritised student needs.

Discrimination and Affective Response

When we acknowledge both the wider discriminatory nature of society and its manifestation in normative standards of quality, we recognise that educational practices can be reflective of broader discrimination (Tannock, 2007). This leads to the appreciation that higher education can be built for neurotypical academics and students (McDermott & Mosley, 2023). From this, the poor mental health of neurodivergent students begins to make sense (Kuder et al., 2021). Analogising to more researched minoritises, such as LGBTQ+ students, we can predict significant improvements in wellbeing when an inclusive institutional culture is created (Huang & Fan, 2024). As well as viewing feedback as a current contributor to discrimination, we can therefore recognise it as a future mitigator.

If feedback reinforces discrimination, negative emotions to feedback may be seen as not only understandable but justified. Instead of reflecting misunderstandings of feedback (Carless & Boud, 2018), they reflect the empirically-backed and personally frustrating undermining of neurodivergent ability (Bailey, 2023; Marcotte, 2023; Beagan et al., 2025). Given neurodivergent people are often justice sensitive, it may be particularly frustrating that their manifestation of academic ability is so underformulated. Even in critical literature, there is a gap in recognising neurodivergent minority stress compared to the academic stresses recognised in other minorities (Vaccaro et al., 2021). Only from neurodivergent individuals themselves is there any discussion of the distress and discomfort created by the strands of discrimination running through higher education (Botha, 2021; Butcher & Lane, 2024). This can make feedback inherently personal while reducing autonomy in the development process, which together can make developing a growth mindset exceedingly difficult (Westby, 2020). Thus, at the root of what is considered poor affective response may be systemic discrimination.

Reframing Defensiveness

When we recognise affective response as influenced by discrimination, educators can appreciate the importance of critical creating compassionate feedback spaces in mitigating distressing affective responses. However, given that societal

discrimination for neurodivergent students is deeply embedded, reducing institutional guilt may not absolve responsibility for institutions to work with negative affective 11 reactions. We may therefore decide to design systems that can work with emotional reactions typically framed as negative. One particularly negatively-framed response is defensiveness. Defensiveness is often considered entirely maladaptive (Rudolph et al., 2007), or even antithetical to reflectiveness (Schwartzman, 2007). However, this may limit our appreciation of defensiveness as a contextual response. We know from empirical research that defensiveness to critical feedback is elevated in women (London et al., 2012), students of colour (Cohen et al., 1999), and likely in neurodivergent students (Bhandari & Rainford, 2023). Moreover, it is often ignored that when evidence-based, justified defensiveness is built on self-reflection. Viewing defensiveness as a skill, a form of self-reflection, and a natural inclination of rejection-sensitive learners (Babinski et al., 2019), it may therefore be valuable to engage with defensiveness openly.

Reframing defensiveness as something to work with may not only be helpful in reshaping a deficit framing of emotional responses that is damaging for neurodivergent students (Kennedy et al., 2025), but also in extending the dialogic interpretation of feedback (Carless & Chan, 2016) to include engaging with Constructive Defensiveness. Constructive defensiveness to feedback includes a student's meta-cognition of injustice within feedback, and may present a powerful opportunity for self-advocacy that increases workplace readiness in environments where anti-neurodivergent discrimination is still commonplace (McDermott & Mosley, 2023). Constructive engagement with defensiveness could involve listening to student perspectives and providing opportunities for further engagement for those who feel strongly: a potentially critical and underexplored method of driving increased feedback literacy and improving institutional practice. Institutions can reconceptualise processes in which educators and students communicate on feedback from improving student concordance with fixed expectations (Westby, 2020), to a bidirectional and dialogic practice that recognises not only recognises an exchange in ideas of quality, but also of emotional perspective (Lea & Street, 1998).

Conclusion

This review examined an underexplored proposition - that neurodivergence may shape student feedback literacy. Owing to gaps in direct evidence, the review provided conceptual directions through synthesis of other research fields. Through this, it has laid the foundations for empirical research incorporating neurodivergent perspectives on assessment and feedback. A key theme throughout the review was the normative expectations that arise when identity is unconsidered. It is critical that future studies centre identities such as neurodivergence when examining student experiences with feedback. Incorporation of neurodivergent perspectives will begin to address a fundamental problem in feedback literature – the gap between theory and practice. Engaging in more radical co-construction of feedback frameworks and practice will increase empathy and broaden our perspective on the sociocultural construction of feedback literacy. Such engagement recognises that manifestation of feedback literacy is shaped by learner context. This may help influence the development of psychologically safe feedback environments, opening the feedback space for bidirectional dialogue.

The development of inclusive and socioculturally contextual feedback literacy presents an opportunity to close attainment gaps for minoritised students. By moving away from a study skills conception, we recognise feedback literacy as the development of a learner aware of their position, their areas for continuous development, and of justice in the feedback process. Through reimagining feedback literacy through a critical pedagogical lens, this review has recognised that feedback literacy is inherently political, built on power relations, and an opportunity for developing critical consciousness. Whether we engage with further critical analysis and incorporation of traditionally marginalised perspectives may determine if feedback literacy policies unintentionally perpetuate educational inequality or can contribute to equity in higher education.

Bibliography

ADCET. (n.d.). Higher Education Statistics. *AustralianDisabilityClearinghouse on Education and Training*. https://www.adcet.edu.au/inclusive-teaching/understandingdisability/higher-education-statistics#Ref. Accessed 9 December 2025

Babinski, D.E., Kujawa, A., Kessel, E.M., Arfer, K.B. and Klein, D.N., 2019. Sensitivity to peer feedback in young adolescents with symptoms of ADHD: Examination of neurophysiological and self-report measures. *Journalofabnormal child psychology*, 47(4), pp.605-617.

Bailey, C., 2023. 'Neurodivergent literacies': exploring autistic adults' 'ruling passions' and embracing neurodiversity through classroom literacies. *Literacy*, 57(2), pp.120-131.

Beagan, B.L., Gulliver, M., Sibbald, K., Pride, T. and Yee, B., 2025. Mad, mentally ill and neurodivergent professionals: epistemic injustice in practice. *Disability & Society*, pp.1-22.

Bhandari, R. and Rainford, J., 2023. Exploring the transitions of neurodivergent access students to level one study: Narratives of study skills and support. *International Journal of Educational and Life Transitions*, 2(1).

Bloxham, S., Den-Outer, B., Hudson, J. and Price, M., 2016. Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, 41(3), pp.466-481.

Botha, M., 2021. Academic, activist, or advocate? Angry, entangled, and emerging: A critical reflection on autism knowledge production. *Frontiers in psychology,* 12, p.727542.

Boud, D. and Molloy, E., 2013. Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in higher education*, 38(6), pp.698-

712.

Boven, F., 2018. Special interests and inclusive academic learning: an autistic perspective. *Advances in Autism*, 4(4), pp.155-164.

Buckley, E., Sideropoulos, V., Pellicano, E. and Remington, A., 2024. Higher levels of neurodivergent traits associated with lower levels of self-efficacy and wellbeing for performing arts students. *Neurodiversity*, 2, p.27546330241245354.

Butcher, L. and Lane, S., 2024. Neurodivergent (Autism and ADHD) student experiences of access and inclusion in higher education: an ecological systems theory perspective. *Higher Education,* 90(1), pp.243-263.

Cai, R.Y. and Richdale, A.L., 2015. Educational experiences and needs of higher education students with autism spectrum disorder*. Journal of autism and developmental disorders*, 46(1), pp.31-41.

Carless, D. and Boud, D., 2018. The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), pp.1315-1325.

Carless, D. and Winstone, N., 2023. Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in higher education*, 28(1), pp.150-163.

Carless, D. and Young, S., 2024. Feedback seeking and student reflective feedback literacy: a sociocultural discourse analysis. *Higher Education*, 88(3), pp.857-873.

Chapman, L., Rose, K., Hull, L. and Mandy, W., 2022. “I want to fit in… but I don’t want to change myself fundamentally”: A qualitative exploration of the relationship between masking and mental health for autistic teenagers. *Research in Autism Spectrum Disorders*, 99, p.102069.

Clark, S.A., Rivers, M.L. and Overono, A.L., 2025. The Role of Answer Justification in Multiple-Choice Testing: Effects on Performance and Metacognitive

Accuracy.*Behavioral Sciences* , 15(4), p.477.

Cohen, G.L., Steele, C.M. and Ross, L.D., 1999. The mentor's dilemma: Providing critical feedback across the racial divide*. Personality and social psychology bulletin*, 25(10), pp.1302-1318.

Dror, L., 2022. Is there an epistemic advantage to being oppressed?. *Noûs,* 57(3), pp.618-640.

Ellis, R., 2010. Epilogue: A framework for investigating oral and written corrective feedback. *Studies in second language acquisition*, 32(2), pp.335-349.

Fernández-Toro, M. and Duensing, A., 2021. Repositioning peer marking for feedback literacy in higher education. *Assessment & Evaluation in Higher Education,* 46(8), pp.1202-1220.

Fricker, M., 2007. Epistemic injustice: *Power and the ethics of knowing*. Oxford university press.

Gibbs, G., 2006. How assessment frames student learning. In *Innovative assessment in higher education* (pp. 43-56). Routledge.

Gravett, K., 2022. Feedback literacies as sociomaterial practice. *Critical Studies in Education*, 63(2), pp.261-274.

Grove, R., Hoekstra, R.A., Wierda, M. and Begeer, S., 2018. Special interests and subjective wellbeing in autistic adults. *Autism Research*, 11(5), pp.766-775.

Hookway, C., 2012. Some varieties of epistemic injustice: Reflections on Fricker. *Episteme*, 7(2), pp.151-163.

Hua, O., Shore, B.M. and Makarova, E., 2012. Inquiry-based instruction within a community of practice for gifted–ADHD college students. *Gifted Education International*, 30(1), pp.74-86.

Huang, X. and Fan, B., 2024. The association between campus climate and the mental health of LGBTQ+ college students: A systematic review and metaanalysis. *Sexuality & Culture*, 28(4), pp.1904-1959.

Ibarra-Sáiz, M.S., Rodríguez-Gómez, G. and Boud, D., 2020. Developing student competence through peer assessment: the role of feedback, self-regulation and evaluative judgement. Higher Education, 80(1), pp.137-156.

Jackson, S.L., Hart, L., Brown, J.T. and Volkmar, F.R., 2017. Brief report: Selfreported academic, social, and mental health experiences of post-secondary students with autism spectrum disorder. *Journal of autism and developmental disorders*, 48(3), pp.643-650.

Janse van Rensburg, M. and Liang, B., 2025. Improving autistic students' experiences in higher education: Developing a community framework for individual autistic student and autistic community flourishing. *Autism in Adulthood*, 7(2), pp.141-154.

Kennedy, L.J., Richdale, A.L. and Lawson, L.P., 2025. Comparing disclosure and supports used by higher-education students with neurodivergent or mental health conditions. *Autism in Adulthood*.

Kolovelonis, A., 2023. Exploring university students' justifications for making metacognitive judgments of learning. *Trends in Higher Education*, 2(3), pp.421-433.

Kroll, E., Lederman, M., Kohlmeier, J., Ballard, J., Zant, I. and Fenkel, C., 2025. Examining the mental health symptoms of neurodivergent individuals across demographic and identity factors: a quantitative analysis. *Frontiers in Psychology*, 16, p.1499390.

Kuder, S.J., Accardo, A.L. and Bomgardner, E.M., 2021. Mental health and university students on the autism spectrum: A literature review. *Review Journal of Autism and Developmental Disorders*, 8(4), pp.421-435.

Lea, M.R. and Street, B.V., 1998. Student writing in higher education: An academic literacies approach. *Studies in higher education*, 23(2), pp.157-172.

Little, T., Dawson, P., Boud, D. and Tai, J., 2023. Can students' feedback literacy be improved? A scoping review of interventions. *Assessment & Evaluation in Higher Education,* 49(1), pp.39-52.

London, B., Downey, G., Romero-Canyas, R., Rattan, A. and Tyson, D., 2012. Gender-based rejection sensitivity and academic self-silencing in women. *Journal of personality and social psychology*, 102(5), p.961.

Mahak, S., Malone, S., Elsherif, M., Hand, C.J. and Morsanyi, K., 2025. Academic anxiety and cognitive reflection in neurodivergence based on evidence from a large international sample. *Scientific Reports*, 15(1), p.37522.

Marcotte, M., 2023. *Using formal epistemology to model epistemic injustice against neurodivergent people* (Doctoral dissertation, The University of Western Ontario (Canada)).

McDermott, L.G. and Mosley, N.A., 2023. " Academia, as a whole, is structured entirely without any consideration for neurodivergency," and other things neurodivergent students want you to know. In *Physics Education Research Conference.*

Milton, D.E., 2014. Autistic expertise: A critical reflection on the production of knowledge in autism studies. *Autism*, 18(7), pp.794-802.

Milton, D. and Sims, T., 2016. How is a sense of well-being and belonging constructed in the accounts of autistic adults?. *Disability & Society*, 31(4), pp.520-534.

Molloy, E., Boud, D. and Henderson, M., 2020. Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*, 45(4), pp.527-540.

Müller, V. and Pikó, B., 2024. From ADHD to well-being: The Role of Rejection Sensitivity in college life. *European Psychiatry*, 67(S1), pp.S284-S285.

Nieminen, J.H. and Carless, D., 2023. Feedback literacy: A critical review of an emerging concept. *Higher Education*, 85(6), pp.1381-1400.

Nieminen, J.H., Tai, J., Boud, D. and Henderson, M., 2022. Student agency in feedback: beyond the individual. *Assessment & Evaluation in Higher Education*, 47(1), pp.95-108.

Price, M., Handley, K. and Millar, J., 2011. Feedback: Focusing attention on engagement. *Studies in higher education*, 36(8), pp.879-896.

Rowney-Smith, A., Sutton, B., Quadt, L. and Eccles, J.A., 2024. The lived experience of rejection sensitivity in ADHD-a qualitative exploration. *medRxiv,* pp.2024-11.

Rudolph, Almut & Wiesner, Anja & Schuetz, Astrid. (2007). Self-Esteem Discrepancies and Defensive Reactions to Social Feedback. International Journal of Psychology - INT J PSYCHOL. 42. 174-183. 10.1080/00207590601068134.

Sadler, D.R., 2013. Ah!… so that's 'quality'. In *Assessment* (pp. 130-136). Routledge.

Schwartzman, L., 2007. Student defensiveness as a threshold to reflective learning in software design*. Informatics in Education-An International Journal*, 6(1), pp.197-214.

Shields, S., 2015. 'My work is bleeding': exploring students' emotional responses to first-year assignment feedback. *Teaching in Higher Education*, 20(6), pp.614-624.

Simonsmeier, B.A., Peiffer, H., Flaig, M. and Schneider, M., 2020. Peer feedback improves students' academic self-concept in higher education. *ResearchinHigher Education*, 61(6), pp.706-724.

Steen-Utheim, A. and Wittek, A.L., 2017. Dialogic feedback and potentialities for

student learning.*Learning, Culture and Social Interaction* , 15, pp.18-30.

Syharat, C.M., 2023, June. Burnout: The Cost of Masking Neurodiversity in Graduate STEM Programs". In *2023 ASEE Annual Conference & Exposition*. ASEE.

Syharat, C.M., Hain, A., Zaghi, A.E., Gabriel, R. and Berdanier, C.G., 2023. Experiences of neurodivergent students in graduate STEM programs. *Frontiers in Psychology*, 14, p.1149068.

Tai, J., Ajjawi, R., Boud, D., Dawson, P. and Panadero, E., 2018. Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher education*, 76(3), pp.467-481.

Tannock, S., 2007. The problem of education-based discrimination. *British Journal of Sociology of Education*, 29(5), pp.439-449.

Trnka, S. and Trundle, C., 2014, April. Competing responsibilities: Moving beyond neoliberal responsibilisation. In *Anthropological forum* (Vol. 24, No. 2, pp. 136-153). Routledge.

Vaccaro, A., Miller, R.A., Kimball, E.W., Forester, R. and Friedensen, R., 2021. Historicizing minoritized identities of sexuality and gender in STEM fields: A grounded theory model. *Journal of College Student Development*, 62(3), pp.293-309.

Westby, C., 2020. Growth mindsets: ideas from Carol Dweck. *Word of Mouth*, 31(5), pp.1-3.

White, H.A. and Shah, P., 2011. Creative style and achievement in adults with attention-deficit/hyperactivity disorder. *Personality and individual differences*, 50(5), pp.673-677.

Winstone, N. and Carless, D., 2019. *Designing effective feedback processes in higher education: A learning-focused approach*. Routledge.

Xu, Y. and Carless, D., 2017. ‘Only true friends could be cruelly honest’: cognitive scaffolding and social-affective support in teacher feedback literacy. *Assessment& Evaluation in Higher Education*, 42(7), pp.1082-1094.

Authours Bio:

Nathaniel is a final year Biology MSci Student at the University at the University of Bristol. In his final year, he has focused on improving the educational experiences of neurodivergent people. Drawing empathy from his own experience and a passion for social justice, he has worked to identify the implicit discrimination in education systems. He believes in the universal design of education, working to proactively remove biases and develop a learning environment that works for everyone. Nathaniel now works as England and Wales Coordinator of the Plant-Based Universities campaign, working to make university catering affordable, inclusive, and sustainable.

Digital Care and the Shift of Responsibility in AI Intimacy: Evidence from Character.AI and Love and DeepSpace

Authors: Zixiao Fu

ORCID: 0009-0008-6277-0867

Programme: Sociology MSc

Affiliation: FacultyofArts,Law and Social Sciences

Email: dw25623@bristol.ac.uk

Abstract

This study explores how digital care, particularly through chatbots and dating simulation games, reshapes the experience and responsibility of emotional support. Drawing on qualitative analysis of user comments from platforms such as Character.AI and Xiaohongshu, alongside existing literature, the research examines how individuals engage with digital intimacy in everyday life. The findings suggest that while digital care provides accessible and personalised emotional support, it also shifts responsibility for wellbeing from social institutions to individuals. Users actively manage their emotional experiences, investing time and affective labour, which may lead to stress or dependence. By linking these observations to theories of platform capitalism and social reproduction, the study argues that digital care reflects a broader shift towards individualised responsibility. This raises important ethical and policy questions about the role of technology in care provision and the need to maintain collective support structures.

Keywords Digital care, AI intimacy, Dating sim games, Chatbot, Social Reproduction

Introduction

In recent years, digital technologies such as chatbots have not only been used for task assistance, but also increasingly for emotional interaction. A notable example emerged in China in 2024 when a woman shared that she had been in a relationship with a chatbot built on ChatGPT' s DAN mode and introducing the AI partner to her family as her boyfriend (Jiang, 2024). This case illustrates how users actively customise AI systems to fulfil emotional needs, blurring the boundary between functional tools and relational companions. Similarly, Character.AI had approximately 20 million monthly active users by early 2025, with young adults (18–24) comprising over half of the user base, indicating high engagement and widespread adoption among younger generations (Reigns, 2025). These cases illustrate that individuals are increasingly turning to digital systems for emotional companionship, raising questions about the distribution of caregiving responsibilities in contemporary society.

This study examines whether digital care shifts the responsibility for emotional care from society onto individuals, and how this transformation reshapes the experience and organisation of caregiving. In this study, "digital care" refers to forms of emotional support, companionship, and affective interaction that are mediated through digital technologies (non-human), including chatbots, AI companions, and dating sim games. Unlike traditional forms of care, which are typically embedded in social relationships and institutional structures, digital care is often individualised and mediated through personalised technological interfaces. Digital care does not only involve the provision of support by technology, but also requires active engagement from users, including emotional investment, time, and attention. As such, it can be understood as a form of care practice that is co-produced through human–technology interaction. This dual nature raises important questions about the distribution of responsibility: while digital systems appear to offer emotional support, they may also shift the burden of managing emotional wellbeing onto individuals and mask the responsibility of society and institution.

There are two distinct but related forms of digital intimacy in this study, namely AI-powered chatbots (Character.AI) and romance simulation games (Love and Deepspace). While both involve emotional engagement with non-human entities, they differ in structure, interaction, and the user communities they generate. Chatbots use

large language models to generate open-ended, conversational responses, allowing for flexible and unpredictable interaction. However, Love and Deepspace is a narrative-driven game with pre-scripted dialogue, voice acting, and animated scenes. Its 3D character design and immersive storytelling are key differentiating features. Unlike the free-form conversation of chatbots, players in Love and Deepspace develop relationships through structured choices and gameplay. Moreover, the game extends beyond the screen, it has cultivated a vibrant community of fans who engage in fan- made creations, follow voice actors across social media, and organise both online events and offline meetups. These fan activities and community interactions are part of the game's appeal and help sustain emotional engagement beyond the game itself. Despite these differences, both are included in this study because they share a common function, which is providing emotional companionship through digital mediation. Both require users to invest time, attention, and affective labor, and both have been described by users as sources of comfort for real-life relationships. By examining them together, this study aims to identify broader patterns in how digital care operates across different technological forms, rather than claiming they are identical.

Existing research highlights a growing expectation that individuals manage their own wellbeing through digital tools. Lupton (2016) argues that self-tracking technologies frame health and emotional wellbeing as matters of personal responsibility, encouraging users to monitor and optimise themselves. Srnicek (2017) extends this perspective, suggesting that platform capitalism transforms social services into on-demand, individualised interactions, effectively relocating responsibilities from collective institutions to users. Applied to digital care, these insights imply that emotional support may increasingly be accessed through personalised technologies rather than public or social provision. While this may enhance the user experience, it may also introduce certain drawbacks.

This shift has important implications when considered through the lens of depletion. Feminist scholars argue that when care responsibilities are concentrated at the individual level, they lead to the exhaustion of emotional, temporal, and cognitive resources, disproportionately affecting women (Federici, 2019, p. 180; Power, 2020, pp.67-73). Digital care may therefore obscure structural inequalities by presenting care

as something that can be individually managed. Nancy Fraser's (2022, pp. 56-58) theory of social reproduction provides a critical framework for understanding these dynamics. Fraser conceptualises care as a collective responsibility. From this perspective, the rise of digital care raises concerns about a potential reconfiguration of care responsibilities, in which technological solutions may substitute for, rather than complement, social provision. Building on these perspectives, Dong & Ding (2025) examine AI intimacy and their ethical implications. She argues that virtual intimacy can deepen subjectivity crises, exacerbate the emotional impoverishment of the public sphere, and affect social trust, while also potentially reinforcing existing gender hierarchies. Dong & Ding's (2025) analysis highlights how individualised digital care may mask structural inequalities, showing that users' emotional labour is not only a personal burden, but also reflects broader social and ethical tensions. This study complements existing literature by connecting the micro-level experience of digital intimacy with macro-level concerns about social structures and ethical responsibilities. A competing perspective within the literature questions whether AI intimacy represents a genuinely new phenomenon or a continuation of existing media practices. Scholars such as Turkle (2011) have long argued that people form emotional attachments to computational objects, suggesting that contemporary AI companions may be a technologically advanced but conceptually familiar extension of earlier phenomena like virtual pets or relationship simulators. Others caution against technological determinism — the assumption that technology directly causes social change — arguing that how users integrate AI into their lives is mediated by existing social norms, gender roles, and economic structures (Couldry & Mejias, 2019). This study does not assume that digital care determines new forms of responsibility; rather, it examines how platform design and user practice co-produce a shift that may already be underway in neoliberal societies.

While existing literature has examined the individualisation of care and the rise of digital intimacy, less attention has been paid to how these shifts are experienced and normalised in everyday digital practices. This study therefore asks: does digital care shift the responsibility for care from society or the state onto individuals? Further, by investigating the intersection of technology, intimacy and care, this study aims to explore how digital care transforms both the experience of caregiving and the social norms that shape it.

Methodology

Guided by the research question: how digital care, particularly through chatbots and dating simulation games, reshapes the experience and responsibility of emotional support, this study uses a qualitative research design that combines literature-based analysis with platform-based content analysis to examine how digital care is experienced and how responsibility for care is articulated in digital contexts. Rather than aiming for statistical generalisation, the study seeks to provide a theoretically informed interpretation of emerging forms of digital intimacy and care.

First, a literature-based approach is employed to establish the theoretical framework. Key concepts such as the individualisation of care, platform capitalism, depletion, and social reproduction are drawn from existing scholarship (Lupton, 2016; Srnicek, 2017; Rai, 2024; Fraser, 2022), alongside recent studies on AI-mediated intimacy (Dong& Ding, 2025). This provides a conceptual basis for analysing how digital care may reshape the distribution of responsibility. Second, the study draws on qualitative analysis of user-generated content from digital platforms. This includes posts and comments from Xiaohongshu discussing relationships with AI systems, as well as publicly available user discussions related to Character.AI. These materials are used to explore how users describe emotional support, companionship, and their own role in managing wellbeing through digital interactions. A purposive sampling strategy is used to select relevant content that explicitly engages with themes of loneliness, intimacy, and AI-mediated relationships. Third, illustrative case analysis is used to extend the discussion to digital intimacy in platform-based environments such as romance simulation games (e.g. Love and Deepspace). These cases are not treated as primary empirical data but as examples that highlight how emotional engagement is structured, monetised, and embedded within digital platforms. This allows the study to situate digital care within broader dynamics of platform capitalism and the commodification of intimacy. The data is analysed using thematic analysis, focusing on how users construct meanings around care, responsibility, and emotional support. Attention is paid to instances where care appears to be individualised, as well as to indications of emotional labour and potential depletion.

Ethical considerations are addressed by using only publicly available data and anonymising any identifiable information. Following the Association of Internet

Researchers (AoIR) guidelines, I recognise that "publicly available" does not automatically mean the data are free for all research uses without further ethical consideration (Franzke et al., 2020). The ethical justification for using these data rests on three considerations: the data are limited to content users have deliberately made public; all identifiable information has been anonymised; and the analysis focuses on thematic content rather than individual users. To mitigate the ethical ambiguity of using public social media data without explicit user consent, the study prioritises anonymisation, limits data use to the stated purposes, and does not share raw data. While this approach provides rich qualitative insights, it is limited by the non-representative nature of platform data and the influence of platform-specific cultures and algorithms on user expression.

Several methodological limitations should be acknowledged. First, users' comments on Reddit and Xiaohongshu are subjective and produced by a self- selected population: those who are willing to publicly discuss their experiences. This may overrepresent highly engaged users or those with strong positive or negative views, while silent users with ambivalent or mundane experiences are absent from the sample. Second, platform algorithms shape which comments become visible or recommended, potentially amplifying certain voices over others. The comments analysed here are therefore not necessarily representative of all Character.AI or Love and Deepspace users. Third, the absence of demographic information (age, gender, socioeconomic background) limits the ability to examine how digital care experiences vary across social groups. This study does not claim generalisability. Instead, these findings are treated as illustrative of themes that emerge within these online communities, offering qualitative insights that can inform further research.

Results

The analysis of user-generated content from Character.AI and secondary reports on Xiaohongshu reveals several recurring themes regarding digital care experiences: Emotional support and companionship. Many users reported that AI interactions provided a sense of emotional comfort. For instance, one Character.AI user said, "it gave me some comfort and also helped me to avoid spiralling in my thoughts … after I was done venting with the conversation I left it feeling positive and better about myself" (Reddit user, 2025). Another likened the AI to a responsive diary: "I guess I

treat it more like a diary, but one that responds?" (Reddit user, 2025). Players of Love and Deepspace similarly described immersive emotional experiences. One player wrote, "like he accompanies me every day to work and coffee," while another stated, "that moment I really felt moved to tears" (Xiaohongshu reports, 2024). Beyond the game itself, players' emotional engagement extends into fan communities. They produce and share fan art and fiction based on the context of the game, follow voice actors on social media, and attend offline fan events (Li, 2026; GameSpot, 2025). As one player put it, "Dive into Love and Deepspace, embrace your happy place" (玩《恋与深空》，做幸福女人) (Xiaohongshu reports, 2024). These examples suggest that player investment involves not only the narrative in game, but also participation in a broader fan culture that reinforces emotional bonds and community belonging.

Risks of dependence and emotional exhaustion. Some users reported over-reliance on digital companions or negative impacts on their emotional wellbeing. Comments such as "Felt I was living in my head too much. So, I had to block the app everywhere" (Reddit user, 2025) or "…from my experience, I can totally agree with the lawsuit, the app is destroying your brain. I saw it on myself, my mental health went down the hill with Character.AI…" (Reddit user, 2025) indicate that intensive engagement may generate emotional strain, highlighting the potential for depletion of cognitive and affective resources. Although such accounts may not be universally representative of all users, they serve to highlight a recurring concern within these communities

regarding the potential adverse consequences associated with excessive engagement.

Several users reflected on the limits of digital companionship, such as "do you still feel the need to talk to real humans … or does AI companionship eliminate that need?" (Reddit user, 2025). This suggests that while AI can provide emotional support, it does not fully replace human interaction, and users must navigate the boundaries between digital and social responsibilities. Players of dating sim games similarly expressed deep emotional engagement, noting immersive experiences, highlighting that players participate in the game, and invest money and time to derive feelings of happiness and emotional satisfaction. Together, both of cases indicate that digital care in such games is co-produced by users and technology, characterised by self-managed and personalised, rather than being provided as a social or institutional service.

Overall, these results suggest that digital care platforms offer both affective benefits and potential emotional burdens. Users experience companionship and support but also face risks of cognitive and emotional depletion, indicating a shift in care responsibility from collective structures to individuals.

Discussion

These results show that digital care change where emotional support comes from. AI companions and dating sim games give comfort and enjoyment. At the same time, they shift the responsibility of caring from structure and organisation such as community networks, family support, and publicly available mental health services to individual users, which means users need to solve their own emotional problems and manage their own wellbeing. Care is no longer just social. This matches Srnicek's (2017) idea of platform capitalism and Lupton's (2016) view of self-tracking. The duty of care moves from institutions to the private part. In addition, user comments also show a double effect. While many users report feeling emotionally supported, a subset describes experiences of stress or emerging dependence. These accounts indicate that engagement with digital care can generate cognitive and affective strain, consistent with Rai's (2024) concept of depletion. Technology gives short-term benefits, but users take on new responsibilities. Unlike traditional care, where effort is shared, digital care makes each person manage interactions and feelings by themselves.

Digital care technologies appear to shift the source of emotional support from social institutions to individuals. This change affects how people experience responsibility for care. According to Fraser's (2022) framework, care is ideally shared across households, markets, and the state. When AI companions and romantic simulation games provide emotional support, this collective responsibility is partially transferred to users. Rai's (2024, pp. 23-27) concept of depletion helps explain the personal cost of this shift, as users must invest time, attention, and emotional energy, which can lead to stress or mental fatigue.

Platform capitalism further shapes this process. Digital care platforms present emotional support as personalised and flexible, offering users seemingly free choices. However, this model requires individuals to manage their own wellbeing and engage in continuous emotional labour. In this context, digital care can be understood as

aligning with a neoliberal logic, where individual freedom is emphasised while collective responsibility is reduced. This logic manifests in two key ways. Srnicek (2017) notes that social services are increasingly delivered as on-demand, individualised interactions, while Lupton (2016) shows how digital tools encourage people to monitor and optimise themselves. Applied to digital care, these insights suggest that emotional support is framed as a self-managed task rather than a shared social responsibility. Technology expands access to care, but it also embeds care within user labour.

It is important to consider whether digital care necessarily transfers responsibility or whether it may supplement existing forms of support. Critics might argue that technologies do not replace social care but rather fill gaps where traditional support is unavailable, such as during night loneliness or social withdrawal (Turkle, 2011). From this perspective, digital care could be seen as an addition to the care landscape rather than a substitution. Furthermore, users are not merely passive recipients of platform logic. They exercise agency in choosing when and how to engage with digital companions. A user who treats an AI as a responsive diary (Reddit user, 2025) is actively shaping the terms of that relationship, not simply being exploited. This study acknowledges these possibilities. However, even when digital care supplements rather than replaces, the underlying structure remains individualised. Users must still manage their own emotional wellbeing, and platforms are not accountable for care outcomes. Therefore, although supplementation and user agency are genuine, they do not negate the shift in responsibility. In fact, they may obscure it.

However, it is worth noting that the gaming context complicates the picture of individualised digital care. Unlike chatbots, where interaction remains primarily private, games like Love and Deepspace generate collective experiences through player communities. While the core emotional attachment is between the player and a virtual character, the social practices (such as creating fan works, participating in online communities, and attending offline meetups) surrounding it, which introduce a communal dimension. This suggests that not all forms of digital care are equally individualising. In some cases, the platform may provide a focus for social connection, even as the care itself is mediated through technology.

Ethical problems become clearer when we look at the wider social context. Dong & Ding (2025) argue that AI intimacy can change how people relate to others. When users get used to interacting with AI, they may prefer these virtual interactions, because they are easier to control and do not require effort from another person. Over time, this may reduce their willingness to deal with real relationships, which are more complex and require mutual care. This can weaken social trust, because trust depends on interacting with real people. At the same time, users still have to manage their own emotions. Digital care can offer comfort, but it does not truly share responsibility for care. This means that individuals continue to carry the burden of emotional wellbeing on their own. In this way, digital care may increase individual responsibility rather than reduce it. This also shows a tension: technology makes support easier to access, but it does not replace the need for social and collective care. Policies should therefore not only focus on the utility of technology but also make sure that real social support systems remain strong.

Conclusion

In a world where AI companions and romantic simulation games offer comfort and connection, emotional support is no longer simply given, instead, it is co-created. Users take on an active role, shaping their own sources of care, navigating feelings of loneliness, and managing the boundaries between virtual intimacy and real-life relationships. These findings suggest that this shift subtly transfers the weight of emotional wellbeing from society onto individuals, revealing both the promise and the pressure of digital care. In forming attachments to these virtual partners, we become Pygmalion, sculpting our own sources of comfort while bearing the responsibility they entail. Technology may ease solitude, but it cannot replace collective care, the task of supporting one another still belongs to families, communities, and social institutions. Recognising this, policymakers might offer guidance and safeguards, ensuring that digital intimacy enriches human experience without becoming a burden. Ultimately, these platforms illuminate a tension at the heart of modern life: the desire for connection mediated by technology, and the enduring need for shared, social responsibility.

References

Couldry, N. and Mejias, U.A. (2019) *The Costs of Connection: How Data is ColonizingHumanLifeandAppropriating it forCapitalism* . Stanford: Stanford University Press.

Dong, K. and Ding, L. (2025) ''AI Lovers' and emotional traps: Virtual intimate human-machine relationships and their ethical dilemmas in the age of AI' ["AI 恋人"与情感陷阱：人工智能时代的虚拟亲密人机关系及其伦理困境], Journal of Northwestern Polytechnical University (Social Science Edition) [西北工业大学学报（社会科学版)], 2025(4), pp. 52-59. (in Chinese)

Federici, S. (2019) *Re-enchanting the World: Feminism and the Politics of the Commons*. Oakland, CA: PM Press.

Franzke, A.S., Bechmann, A., Zimmer, M. and Ess, C. (2020) Internet Research: Ethical Guidelines 3.0. Association of Internet Researchers. Available at: https://aoir.org/reports/ethics3.pdf (Accessed: 8 July 2026).

Fraser, N. (2022) *Cannibal Capitalism: How Our System Is Devouring Democracy, Care, and the Planet – and What We Can Do About It*. London: Verso.

GameSpot (2025) *'How Love And Deepspace Became A Viral Phenomenon In 2025',* 28 December. Available at: https://www.gamespot.com/articles/how-love-and-deepspace-became-a-viral-phenomenon-in-2025/1100-6537076/ (Accessed: 7 July 2026).

Jiang, I. (2024) 'Chinese woman goes viral for falling in love with chatbot, has romantic talks, goes on 'beach date', introduces 'him' to mother', *South China Morning Post*, 25 June. Available at: https://www.scmp.com/news/people-culture/trending-china/article/3263280/chinese-woman-goes-viral-falling-love-chatbot-has-romantic-talks-goes-beach-date-introduces-him (Accessed: 15 March 2026).

Li, Y. (2026) *'Beyond parasocial relationships: fictoromance in the post-subculture of Chinese women'*, Continuum: Journal of Media & Cultural Studies, pp. 1-13. doi:10.1080/10304312.2026.2648707.

Lupton, D. (2016) *The Quantified Self: A Sociology of Self-Tracking*. Cambridge: Polity Press.

Power, K. (2020) The COVID-19 pandemic has increased the care burden of women and families. *Sustainability: Science, Practice and Policy*, 16(1), pp.67–73. doi:10.1080/15487733.2020.1776561.

Rai, S. (2024) *Depletion: The Human Costs of Caring*. Oxford: Oxford University Press.

Reddit User (2025) User comments on Character.AI. [Online] Available at: https://www.reddit.com/r/CharacterAI/comments/1kxboua/user_comments_on_characterai/ (Accessed 14 March 2026).

Reigns, D (2025) Character AI statistics and user trends 2026. AboutChromebooks.com. [Online] 18 November (updated 23 February 2026). Available at: https://www.aboutchromebooks.com/character-ai-statistics/?utm_source (Accessed 14 March 2026).

Srnicek, N. (2017) *Platform Capitalism.* Cambridge, UK: Polity.

Turkle, S. (2011) *Alone Together : why we expect more from technology and less from each other*. New York: Basic Books.

Xiaohongshu Reports (2024) Player experiences in Love and DeepSpace. [Online] Available at: https://www.xiaohongshu.com/search_result?keyword=%25E6%2581%258B%25E4%25B8%258E%25E6%25B7%25B1%25E7%25A9%25BA%25E9%2599%25AA%25E4%25BC%25B4%25E6%2584%259F&source=web_search_result_notes (Accessed 14 March 2026).

Authors Bio:

Zixiao is a MSc student at the University of Bristol. Her research interests focus on care, particularly in the area of care on digital platforms among young adults. She enjoys reading, writing, and is interested in exploring everyday experiences through a sociological lens. Beyond her academic work, she has a strong interest in travel and enjoys visiting places connected to the books and ideas she has encountered in her reading, using these experiences as a way to connect intellectual interests with everyday life.

“Modelling optimal networks: Countering the rational terrorist”

Author: Henry Arbiter

ORCID: 0009-0009-0696-8278

Programme: BScEconomics and Management

Affiliation: SchoolofLaw and Social Sciences

Email: gf23427@bristol.ac.uk

Abstract

This paper presents a novel approach in thinking about covert network structure, employing emerging thought on the rationality of terrorist actors. By using a specific function to model environmental constraints on terrorist groups, and by then searching for optimal solutions using a genetic algorithm, this paper moves to present derived findings that support empirical observations. It is found that the modelled results present in the paper are broadly consistent with what is seen in terror organisations around the world. Additionally, policy implications arising from such a novel understanding are laid out, and future applications of the modelling approach are suggested.

Keywords: Covert Network Optimization, Terrorist Organisational Structure, Counterterrorism policy

Acknowledgements: I would like to thank my personal tutor Professor Alberto Franco for his guidance and advice throughout this project

** All code used in this project is available upon request

Introduction

As both an academic, and practical pursuit, the understanding of the structure and hierarchies of covert criminal organisations has become immensely pertinent in recent times. As the philosophy of terror transfers from contained, national, secessionist movements towards larger, more globally integrated networks, the execution of efficient, reactionary counterterrorism measures have become increasingly difficult to implement (Enders and Sandler, 2012).

Understanding the internal structures of covert networks is inherently difficult; however, certain mathematical models are becoming central in the conversation over both their persistence and also vulnerabilities. Lindelauf, Borm and Hamers (2009) use algorithmic construction to identify the effects of secrecy on network construction, with Enders and Jindapon's (2009) seminal work on network externalities illustrating, theoretically, the trade-offs present in covert network construction. These tradeoffs between secrecy and efficiency can then be observed empirically in the diverse structural patterns and hierarchies observed across terrorist networks worldwide and have serious consequences on network design.

Building on these theoretical foundations, this study employs a genetic algorithm to explore the abstract network solutions that arise under the constraints of communication, secrecy, and resilience to infiltration. By identifying recurring trade-offs, this study aims to provide actionable insights for policymakers; highlighting conditions that favour certain structures, and quantifying points of fragility where counterterrorism efforts are best focused. This study, to the best of my knowledge, is novel in its combination of the network externality framework, and its algorithmic exploration of network space.

Literature review

To understand terror organisations in the context of mathematical equilibria can be met, initially, with some sort of resistance. It is common, even compulsory, in economic analysis to represent individuals as rational actors; those that optimise behaviour towards some goal, subject to some constraint. If such parameters change, then a rational actor is expected to respond in a predictable fashion.

Rationality is a concept not exclusive to this quantitative definition, and it is the other connotations that has meant economic analyses of terror actors has remained on the fringe of the surrounding academia. The idea that the goals and methods of terror, past their abhorrentness, can be reconciled as “rational” is a concept foreign to most. To represent terror as rational is, simply, to be able to understand, and attempt to predict, manifestations of ideology, in order to minimise the harm that comes from it.

Itmay be reductionist to label every individual terrorist actor as rational. In fact, there isalarge catalogue of literature that asserts the numerous reasons why a conclusion ofrationality is misguided. Horgan, in his 2005 book “The Psychology of Terrorism”, concludes that there seems to be a lack of any identifiable psychopathology that is common amongst terrorists, but demonstrates how “unremarkable”, in psychological terms, those who engage in terror activity actually are. This actually suggests some lack of a clear rational motive that is common across actors, reducing the viability of any attempt to create some universal model for the terrorist individual. Crenshaw (2008) also presents some evidence towards this idea. She argues that organisational dynamics, internal politics, and the need to satisfy competing factions often drive leadership to choice, more than the ideas of external optimisation.

This paper does not reject that, on an individual level, terror actors may not always behave rationally. In fact, there is the innate understanding from wider economic modelling that this is simply not possible. Instead, I must posit that the rationalisation of terror networks is possible through the averaging of individual rationale across an entire organisation. It is axiomatic that there must be some collective, organising, logic of terror groups, and so it is the rational adaptation of this that can be modelled. In this assumption, the difference between individual actors becomes increasingly negligible.

This idea of modelling the rational behaviour of organisations over the individual was first explored by Landes in his 1978 study of US plane hijacking. He derived an “offense function”, a set of equations that related the number of hijackings to policy variables, such as the probability of apprehension, present defence mechanisms and length of sentence.

Using empirical data, this function was able to ascertain the impact of each policy variable on the number of attacks; the first instance of mathematical modelling of rational terrorist actors to carry direct influence over policy decisions. Over the next decades, further economic forays expand on this methodology, building theoretical models and verifying such through empirical data (most notably Brophy-Baermann, and Conybeare 1994, and Li 2005). This paper aims to reflect this methodology, building a quantitative model with qualitative grounding, and verifying its outputs through empirically observed structures.

The study of covert network structure and hierarchy is already present in political violence scholarship. Crenshaw (1981) and Rapoport (2013) emphasised hierarchies, cell structures, and different operational modes in their respective research, with models initially distinguishing between centralised and decentralised networks, and attempting to understand the trade-offs leading to their emergence. However, their frameworks remain largely descriptive, with examples of how groups are organised, but no suggestions of why these configurations emerge under certain conditions. Kydd and Walter (2006) move closer to an explanation, showing how,

by linking structural choices to differences in the goals of organisations, different structural designs emerge. However, their model remains in the more qualitative realm; it employs strategic logic over formal optimisation. The literature on covert network theory can therefore be seen to be lacking in “mechanism”, that is, the ability to derive network design endogenously, through theoretical constraints, and not from pre-existing empirical cases.

Network theory offers an attempt at a quantitative understanding of such dynamics. Network properties such as clustering, degree distribution, and path length have been shown to influence both efficiency and resilience (Borgatti & Everett, 2006). This work suggests that centralised systems are fragile, while fully distributed networks are resilient but operationally slower; valuable in establishing the measurable trade-offs that emerge from structural choices, but failing to present reasoning as to which trade-offs rational organisations would accept. It is a mere description of network topologies, without an understanding of the organising problem faced by those within them.

The closest attempt to the optimal network generation objective in this paper, was performed by Lindelauf, Borm and Hamers (2009). Here, they use multi-objective optimisation and bargaining game theory in order to analyse the trade-offs between secrecy and operational efficiency. This study produces optimal network topologies borne from individual exposure probability and provides a strong foundational influence on how a terrorist network is mathematically defined in this study. However, despite similar initial premises, this paper aims to use a different type of algorithmic approach in order to attempt a more reasonable "evolution" process of organisations. The 2009 study employs two, more trivial, algorithms in order to search the solution space: the first being exhaustive, random generation. The second starts with a circular graph, and adds edges and nodes until there becomes zero benefit from continuing. While illustrative, these processes unfortunately require much more computing power in order to reach an optimal solution, with no guarantee that such a solution is the most optimal. Finally, there is no consideration of the cost of communication in any terms other than its relation to discovery. The cost of additional edges therefore arises from the game theoretic function used to understand the influence of secrecy on the networks. Here is the clear gap in the literature: there must be a provision of a more concrete, optimised, algorithmic approach to the constrained optimisation problem of terror network structure.

This paper aims to incorporate the contributions of Enders and Jindapon (2009) in their employment of network externalities. By formalising the costs and benefits associated with information exchange in a covert network, they highlight how changing links affects both operational capacity, and exposure to surveillance. This provides a foundation for the calculus used in the following examination of network formation; however, it stops short of identifying what whole networks look like under given constraints. Computational approaches have also explored network adaptation in response to counterterrorism interventions (Clauset and Gleditsch, 2012), and yet, these studies only focus on network classification and disruption, rather than the construction of an optimised organisational structure.

The most fundamental challenge to this paper’s premise is that of Abrahms (2008), who presented empirical evidence that terrorist groups systematically fail to achieve amajority of their stated political goals. This sits in direct tension with the concept of

the rational optimiser assumption. Without results, it is hard to see how terrorist groups can be understood to be effectively pursuing their goals in a rational manner. This would undermine the foundations of the derived model. In response, a distinction is drawn between structural rationality, and strategic effectiveness. This paper does not claim that rationally optimised internal structures always lead to operational efficiency, but merely that, under a certain set of goals and constraints, they arise endogenously.

There are three gaps in the current literature that this study therefore hopes to address: (1) the lack of models that integrate both resilience and efficiency considerations within a unified framework, (2) the absence of algorithmic optimisation in rational network construction, (3) the need to connect abstract computational results to policy relevant insights. This study offers a novel lens to understand how covert organisations, when treated as a network of rational actors, design strategic networks under operational constraints.

Preliminaries

The derivations of the fitness function found in the following sections, in addition to the proper functioning of the genetic algorithm rely on the ability of the terrorist network to be represented in formal mathematical terms. This section provides a short mathematical, preliminary definition.

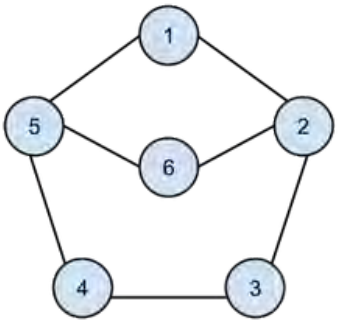
A network can be understood as a collection of individuals, known formally as vertices (V), and the links between such individuals, known formally as edges (E). This gives a definition of a graph as G (V, E). n defines the number of total vertices in the graph. d (i, j) defines the direction and the placement of an edge, that is between vertices i and j. For this study, it is assumed that all networks studied are connected, in the fact that there will never be an individual vertex with E = 0. Finally, the degree of a vertex, the number of vertices to which it is connected, is represented by b, where i is the vertex in question.

$$i$$

As in most economic analysis, these mathematical conditions can be understood in context. When a node or vertex is referred to, it can be interpreted as an individual terrorist. An edge denotes the existence of any kind of communication channel between two terrorists, whether that be open, or just general knowledge of the other individual.

Other graphing assumptions remain incredibly similar to those found in Lindelauf, Borm and Hamers (2009). All calculations on graphs used by the algorithm are done on a matrix, such that each 1 represents a connection between individuals:

Figure 1.



0	1	0	0	1	0
1	0	1	0	0	1
0	1	0	1	0	0
0	0	1	0	1	0
1	0	0	1	0	1
0	1	0	0	1	0

Example Network (n=6), with accompanying matrix representation

By understanding the preliminary mathematical setup, the understanding of the qualitative to quantitative derivations of the fitness function becomes easier.

Fitness function

In order to create an optimal network solution, there must first be an understanding of the factors that covert organisations prioritise. A fitness function is constructed that prioritises three factors: efficiency, robustness, and secrecy. It must be noted that the fitness function expressed in this paper could be easily changed in reaction to any new academic advances in the field, or a specific use case context. It does, also, remain a rather simplified effort; the proposed fitness function is presented as a mere proof of concept to understand the efficacy and future applications of this algorithmic method. In reality, many other constraints may face terror network design, including

recruitment capacity, legitimacy within communities, and ideological coherence; these constraints remain unexplored but are a promising avenue for further extension of such a methodology.

Efficiency is a well-documented conversation in the field of network theory. The efficiency measure employed in the fitness function, mirroring that laid out by Latora and Marchiori (2001), ranks networks based on “global efficiency”.

This measure defines efficiency as the average amount of connections that each individual in the network has to any other (Ek, VerSchneider and Narayan, 2015): ideological coherence; these constraints remain unexplored, but are a promising avenue for further extension of such a methodology.

Efficiency is a well-documented conversation in the field of network theory. The efficiency measure employed in the fitness function, mirroring that laid out by Latora and Marchiori (2001), ranks networks based on “global efficiency”. This measure defines efficiency as the average amount of connections that each individual in the network has to any other (Ek, VerSchneider and Narayan, 2015):

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d(i,j)}$$

$$E(G) = 0, when\ n \leq 1$$

The higher this average number of “edges”, the quicker information is able to be transmitted across the entire system; a reflection of overall integration and connectivity. Fundamentally, efficiency is important as it directly impacts a network’s operational ability. The need for transmission of information and instructions to all areas of the organisation is imperative in maintaining ideological and operational cohesion. Efficiency also allows for networks to adapt their strategies and targets more easily, maintaining their edge in operational manoeuvrability over classical intelligence services.

The fitness function employs two methods of quantifying network robustness. Cascade resilience is the ability of a network to withstand, and recover from, failures that trigger a chain reaction of subsequent adverse effects (Song et al., 2024).

Cascade resilience is scored as how evenly distributed influence is across the network, with cascading failures becoming more severe when a small number of nodes carry disproportionately large importance:

$$R(G) = 1 - \sum_{i \in V} \left(\frac{b_i}{\sum_{j \in V} b_j}\right)^2$$

$$where: R(G) = 1 \; if \sum_{i \in V} b_i = 0$$

In application, this reflects the ability of a network to withstand the capture and interrogation of any number of members in the network: will the network fail if a certain person reveals all of the information they know about said organisation? In a similar measure, algebraic connectivity scores networks highly if they can still maintain efficient communication despite the removal of an arbitrary amount of individual nodes (Maragatham, 2024). Identical to the conventional approach to calculation, algebraic connectivity is represented by the second largest eigenvalue of the solution's matrix:

$$\frac{\lambda(L)}{2}$$

Alongside these three main operational conditions, there is also an allowance for the externality term expressed in Enders and Jindapon (2010): The idea is that members efficiency benefits from close proximity to others. Terrorists are assumed to work more efficiently as part of a team, than as a lone individual. A novel approach to this idea has been attempted by including a distinction between the externality of being close to a "leader" in the network, with an individual becoming more efficient the closer they are to a more senior member. This adaptation of Enders and Jindapon's idea comes through a synthesis with the Gutfraind (2009) model of dynamic membership:

$$D(G) = \frac{1}{n} \sum_{i \in V} \left(\frac{1}{(n-1)\gamma_{max}} \sum_{j \in V, i \neq j} \frac{\gamma(j)}{d(i,j)^{\eta}}\right)$$

$$where: \gamma(j) = \{\gamma_{leader} \; if \; j = leader, \gamma_{follower} \; otherwise\}$$

$$and: \gamma_{max} = max(\gamma_{leader}, \gamma_{follower})$$

The optimality condition also introduces a representation of secrecy through an “edge cost” term, where m is the “cost” of an edge, and the penalty represents total edge density:

$$P_{edge}(G) = -\left|\frac{(m)^*(\sum_{d_c} d_i(i,j))}{n(n-1)}\right|$$

Links between individuals in a terrorist network come at a cost, both material and otherwise (FATF, 2018). Covert communication methods, recently, have often relied on a sophisticated online presence, requiring significant technical and financial investment. There is also an immaterial cost of connection that represents the price of maintaining secrecy and trust across that relationship. In this way, the more “edges”, the more susceptible to discovery an organisation is, and therefore the less optimal the network:

$$P_{deg}(G) = -\left|\sum_{i \in V} max(0, deg(i) - d_{max})\right|$$

There also must be a consideration for penalising too many leaders in a network, in the way that numerous leaders, with potentially conflicting operational preferences may reduce the efficiency of operations. This is represented by the first term below, where a penalty is placed on the density of leaders relative to other individuals. However, it is recognised that having a leader in an organisation is still important, and so a solution will be punished for having not enough, represented by the second term:

$$P_{leaders}(G) = -\left|[\frac{\alpha(L=|\iota|)}{n}] + [\delta \sum_{i \in \iota} max(0, deg(i) - d_{max})]\right|$$

By combining the above equations, a general fitness function for an optimal network solution is produced. However, this paper aims to go further by understanding how results vary across different operational constraints. The three “preset” situations: secrecy, neutral, and efficient, reflect three different counter-terror policy priorities. The secrecy preset reflects an environment where authorities monitor groups

closely. In these times, groups are assumed to prioritise secrecy in an attempt at preservation. In this case, the fitness function is weighted towards maximising cascade resilience, and algebraic connectivity. The opposite scenario, efficiency, reflects situations in which there are no strong counter-terror measures. Here, there isminimal need to prioritise secrecy and preservation, and it is therefore assumed

that the organisations maximise relative to efficiency of communication. Neutral represents the middle ground between the two weightings. The preset weights used remain placeholders for this proof of concept and are available in the code upon request to the author. Combining the fitness function with these preset weights, a final optimality condition is created to which each individual network solution can be scored against each other:

$$TotalScore = \sum_{X} w_{s}(G) * (X(G))$$

This fitness function is combined with the genetic algorithm approach in order to determine a “most optimal” terror network structure under the given environmental constraints.

Genetic algorithms

Genetic algorithms (GAs) employ principles found in biological evolution in order to find appropriate solutions to complex optimisation problems (Forrest, 1993). In the present case, the derived fitness function is used to evaluate a large population of randomly generated candidate network solutions. These candidates are then ranked, and a certain number are allowed to “crossover” traits in order to produce a new generation. The algorithm iteratively applies certain, novel operators, introducing small random changes to the network, which mimic mutation. The population quickly evolves towards a final, optimal, solution output.

A genetic algorithm has been employed in this setting primarily for the fact that it does not require gradient information to produce an output (Forrest, 1993). Due to the style of needed result, many commonly employed optimisation methods fall short with a lack of capability in handling a non-linear objective function. However, GAs

excel at this kind of optimization, employing a stochastic search style in order to explore a larger, and more complex search space (Hassanat et al., 2019). This is the advantage they possess over alternative optimisation methods. Simulated annealing similarly handles non-linear search spaces but remains less suited to such an incredibly large search area (Kirkpatrick, Gelatt and Vecchi, 1983). Bayesian optimisation and exhaustive search also both fail to be viable simply due to their computational expense when handling such a large number of possible candidate solutions (Snoek, Larochelle and Adams, 2012). While not guaranteed to find the globally optimal solution, a GA is able to produce incredibly high performing solutions within a feasible computational budget.

Additionally, due to the way that GAs evolve, they are able to handle imprecise and incomplete data, which, in a field where data sets on covert organisations are not readily available, nor always precise, allow for a way to generate solutions with a non-negligible amount of predictive power.

While other algorithmic approaches have been employed to study the structure of terrorist groups, there has typically been a focus on network destruction rather than construction. Callahan et al's (2012) GREEDY_FRAGILE algorithm identifies "individuals" whose removal would increase network centrality, making the organisation more susceptible to collapse.

This is an important theoretical exercise; however, it is both the opinion of this paper, and that of the authors of the algorithm, that this approach remains only useful in a handful of practical cases, due to the need for a complete data set on a network. In contrast, the algorithm used in this essay focuses on construction of rational networks, an ability that can be used not only in the absence of information, but in lieu of it.

The algorithm uses a population size of 30, which evolves over 300 generations. Networks of 10 to 30 nodes were optimised using tournament selection, a crossover rate of 0.9, and mutation rates of 0.05 for both edge and role assignments. These

parameters were calibrated to balance solution quality and computational efficiency, but remain easily changed in accordance with any new literature that may arise in this research area. With the limited population and generation sizes, it must be acknowledged that the findings illustrated below may not be the final, optimal solutions, but represent the extent to which the networks were optimised by the end of the process.

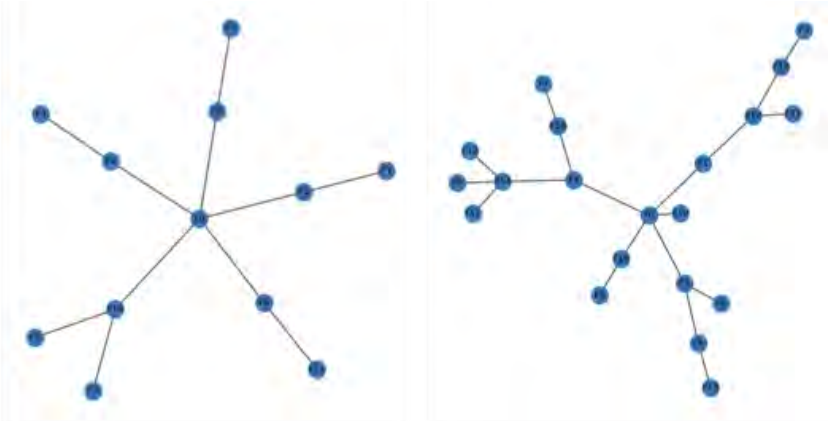
All simulations were conducted in Python 3.9, using NetworkX 2.6 for graphing operations, and NumPY 1.21 for numerical computations. The complete code implementation can be accessed on request from the author.

Findings

The GA weights were programmed with three presets: (1) Secrecy (prioritising resilience and a low number of connections), (2) Efficiency (prioritising fast flows of communication through numerous parties), and (3) Neutral (a balance of both).

Under the secrecy presets, a situation which reflects high surveillance pressures to the covert organisation, the simulated group forms into a centralised network. There seems to be a focus on the individual actor, with small groups placed very separately around one or two individual nodes. The networks generated are highly resilient, however, score low on efficiency of communication, as information must pass through numerous individuals at a time.

Figure 2.



Secrecy preset (n = 12)

Secrecy preset (n = 20)

This topology is the result of heavily penalising edge density. The algorithm is optimising towards a solution that removes any links not strictly necessary for maintaining connectivity.

By minimising $d_i(i, j)$ in:

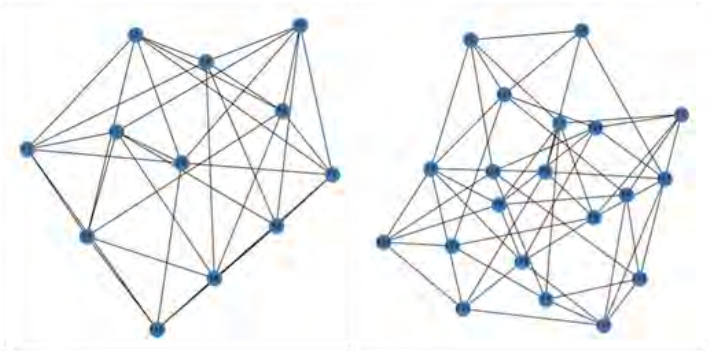
$$P_{edge}(G) = -\left| \frac{(m)^*(\sum d_i(i,j))}{n(n-1)} \right|$$

While preserving connections across all nodes, star-like topologies emerge, since a single hub requires only $n - 1$ edges to connect an entire network. The cascade resilience term reinforces these star-like structures. If all betweenness is concentrated in the hub, then each peripheral node carries near-zero load $b_i = 0$, causing:

$$\lim_{b_i \to 0} R(G) = 1 - \sum_{i \in V} \left(\frac{b_i}{\sum_{j \in V} b_j} \right)^2 = 1$$

When moving to an efficiency preset, the networks generated are almost fully connected. Information flows rapidly through the entire network, with little distinction between the “leader” and the “follower”. While scoring high on communication, these networks are fragile; compromising one or two individuals has a disproportionate effect on the rest of the network.

Figure 3.



Efficiency preset (n = 12) Efficiency preset (n = 20)

These results are the immediate consequences of the high weighting towards the global efficiency term, which is maximised as the path length between each node

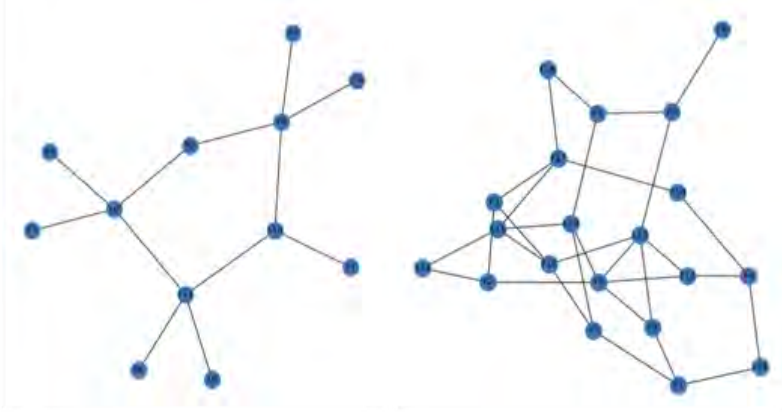
pair approaches 1:

$$\lim_{d(i,j)\to 1} E(G) = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d(i,j)} = \frac{1}{n(n-1)}$$

The disappearance of role hierarchy is also explicable. The externality term $D(G)$ that rewards proximity to leaders becomes relatively insignificant when the efficiency term dominates. Since all the nodes are already very close (d (i, j) = 1), the marginal gains from adesignated leader is small, and the solution gains little from preserving a leadership structure that would otherwise constrain connectivity through $P_{Leaders}$ (G).

Finally, the neutral preset creates hybrid “small world” structures. These clusters are formed around a central individual/group, scoring highly on both efficiency and robustness. This outcome reflects the direct tension between the edge cost penalty (which discourages full connectivity), and the efficiency and resilience terms that discourage extreme sparsity. Total score is therefore maximised at the intermediate point where dense local clusters provide high within-cluster efficiency, with sparse intercluster connections limiting d (i, j) and containing the edge cost penalty. Leaders and followers had near identical degree distributions, suggesting that a classic hierarchical structure actually proves a disadvantage. This configuration shows an adaptive balance, reflecting the operational strategies of covert organisations under moderate surveillance pressures.

Figure 4.



Neutral preset (n = 12) Neutral preset (n = 20)

Discussion

These findings, in several respects, corroborate important existing expectations around covert network design, but it is important that these alignments remain suggestive rather than formally validated against real-world network data.

The emergence of compartmentalised structures under high secrecy priorities reflects observations such as Crenshaw's (1981) emphasis on fragmentation, and Rappoport's (2013) analysis of cellular forms. Such findings also align with Enders and Jindapon's (2009) predictions that increased exposure drives organisations towards lower connectivity networks; a sentiment mirrored in empirical cases such as Al-Qaeda's post 9/11 decentralisation, in which new counter terrorism interest led to a structural shift away from centralised command, towards more compartmentalised, semi-autonomous cells (Vasiliev, 2022).

By generating hybrid "small world" networks when efficiency and secrecy priorities are balanced, the algorithmic output provides a theoretical foundation for structures that have been empirically observed. Examples such as the AQAP's regional cell structure (Cronin, 2006) and the Taliban's shura system (Afsar, Samples and Wood, 2008) both exhibit the clustered hub topology that the neutral preset converges on. However, this must be acknowledged as illustrative parallels, and not direct structural comparisons. The fact that this genetic algorithm naturally converges towards these network structures suggests that small world systems may represent apossible rational equilibrium under mixed operational pressures.

Certain results do, however, depart from traditional assumptions. Prioritising operational efficiency does not produce hierarchical organisations, as hypothesised by Kilberg (2012). Instead, the algorithm produces decentralised but highly connected networks. This finding is consistent with the periods of groups like FARC, who achieved high coordination without the presence of a rigid hierarchy, before later phases saw greater centralisation (Brittain, 2010).

Policy implications

The findings presented above are predicated on the assumption that covert network design is a rational venture; covert groups are not organised at random, instead, there is a certain adaptation present in response to operational pressures and constraints. For policy makers, this reflects the principle that counterterrorism measures must also be differentiated and adapted in response to such recognised constraints. The modelled outputs are suggestive of some sort of organising rationale; however, the model only takes into account a very small set of constraints, and a relatively small search space. Therefore, the policy implications shown here must be taken as academic forays and cannot be made actionable as a result of the influence of this singular computational exercise.

When a network prioritises secrecy, represented by high levels of surveillance or state coercion, they tend to form highly compartmentalised, centralised structures. In such environments, leadership decapitation and targeted strikes, at a certain level, yield limited disruption in the system, as operational arms are fragmented. Instead, counterterrorism efforts may find infiltration, the disruption of logistical nodes, and restrictions of resource flows more effective. Intercepting these intermediaries can destabilise an otherwise resilient, low-connectivity, structure. The model above may be able to be adapted to understand the through flow of an individual node in the network, and which targeted intervention would produce the most disruption to the organisation.

Conversely, when groups prioritise operational efficiency, networks tend to become decentralised and horizontally connected. Here, extensive peer-to-peer communication enables high tempo and coordinated attacks. Although efficient in communication, these configurations are vulnerable to surveillance and infiltration. Policy makers may therefore choose to focus on interventions that increase signals intelligence, network mapping, and exploitation of network wide communication channels to identify key methods to disrupt operational tempo. Resources would theoretically be best focused on coordinated, whole network disruptions, instead of individual, small scale, node decapitations.

Finally, in a more realistic and balanced setting, the algorithm produces hybrid “small world” structures; clusters around semi central hubs. These structures represent the biggest challenge for policy makers, as they score highly across the board. This suggests a need for graduated intervention frameworks, combining selective targeting of coordinators with broader efforts to disrupt inter-cluster communication channels.

Limitations

This study's findings must, however, be interpreted with some methodological constraints in mind. The genetic algorithm optimises 10-50 node networks. This may make it most suitable to low level operational cell analysis, as there may be reductions in stochastic accuracy at higher levels, or with more federated structures. A single cohesive network is a rather idealised and naive representation of whole, modern, covert organisations. In the future, implementations of search size limiting to only pareto efficient solutions could be implemented, such as the optimisations performed in Lindelauf, Borm, and Hamers (2011). This optimisation allows further convergence to some sort of optima.

A further limitation is that the confirmatory findings found in the discussion section are, while encouraging for the salience of this methodology, not strong evidence that this model is correct. The fitness function was the quantitative derivation of qualitative evidence found, partly, in the same literature that generated the empirical observations. There is, therefore, a risk of circularity in the fact that the model may reproduce known structures well, because it was designed with those structures in mind. Future work in this field should incorporate more out of sample data to establish stronger external validity.

Additionally, drawing policy implications from static solutions may be a novel academic pursuit, but it must be recognised that counter-terrorism policy, by its nature, involves a multitude of stakeholders, all with different identities, beliefs, and moral agency; factors that remain incredibly hard to quantify. The most effective policies will therefore use insights from studies such as this one, in conjunction with

human judgement and contextual understanding. It is imperative to realise that behind every modelled network, lies an incredibly complex social fabric that must necessarily be seen and understood.

Conclusion

This study has attempted to model the structural design of covert networks using a genetic algorithm informed by both network externalities and operational trade-offs. The results have shown how varying priorities produce distinct topologies; a reflection of the rational adaption of terrorist actors to environmental constraints.

The findings reaffirm that no single counterterrorism strategy can be universally applied across all networks. Instead, effective interventions must recognise the endogenous relationship between state pressure and terrorist adaptation: how does policy influence the evolutionary trajectory of a network in the long run?

The genetic algorithm used in this study offers a replicable framework for future research. For example, an extension may incorporate dynamic feedback mechanisms, heterogeneous actor preferences built from known human intelligence profiles, or predictive network construction using observed, yet incomplete, network data. Conceptually, the study represents an addition to the more nuanced perspective of representing terrorists as rational individuals, evolving towards high performing configurations within constrained environments. This emphasises a shift from static to evolutionary thinking, stressing anticipation, feedback, and resilience to attacks, rather than reactive suppression.

Ethical statement

The genetic algorithm and simulations reported here are intended solely for the academic analysis of theoretical trade-offs in covert network design. No operational recommendations for real world actors are to be insinuated, and where parameter choices and outputs are described, they are to be taken at a purely conceptual level only

Reference List:

Abrahms, M. (2008). What Terrorists Really Want: Terrorist Motives and Counterterrorism Strategy. International Security, 32(4), pp.78–105. DOI: https://doi.org/10.1162/isec.2008.32.4.78

Afsar, S., Samples, C. and Wood, T. (2008). The Taliban: An Organizational Analysis. [online] Dtic.mil. Available at: https://apps.dtic.mil/sti/html/tr/ADA485136/.

Borgatti, S.P. and Everett, M.G. (2006). A Graph-theoretic perspective on centrality. Social Networks, 28(4), pp.466–484. DOI: https://doi.org/10.1016/j.socnet.2005.11.005.

Brittain, J.J. (2010). Revolutionary social change in Colombia: the origin and direction of the FARC-EP. London Pluto.

Brophy-Baermann, B. and Conybeare, J.A.C. (1994). Retaliating against Terrorism: Rational Expectations and the Optimality of Rules versus Discretion. American Journal of Political Science, 38(1), p.196. DOI: https://doi.org/10.2307/2111341.

Callahan, D., Shakarian, P., Nielsen, J. and Johnson, A.N. (2012). Shaping Operations to Attack Robust Terror Networks. [online] arXiv.org. Available at: https://doi.org/10.48550/arXiv.1211.0709.

Clauset, A. and Gleditsch, K.S. (2012). The Developmental Dynamics of Terrorist Organizations. PLoS ONE, [online] 7(11), p.e48633. DOI: https://doi.org/10.1371/journal.pone.0048633.

Crenshaw, M. (1981). The Causes of Terrorism. Comparative Politics, [online] 13(4), pp.379–399. DOI: https://doi.org/10.2307/421717.

Crenshaw, M. (2008). Theories of Terrorism: Instrumental and Organizational Approaches. Journal of Strategic Studies, 10(04), pp.13–31. DOI: http://dx.doi.org/10.1080/01402398708437313.

Cronin, A.K. (2006). How Al-Qaeda Ends: The Decline and Demise of Terrorist Groups. International Security, [online] 31(1), pp.7–48. DOI: https://doi.org/10.2307/4137538.

Ek, B., VerSchneider, C. and Narayan, D.A. (2015). Global efficiency of graphs. AKCE International Journal of Graphs and Combinatorics, 12(1), pp.1–13. DOI: https://doi.org/10.1016/j.akcej.2015.06.001.

Enders, W. and Jindapon, P. (2009). Network Externalities and the Structure of Terror Networks. Journal of Conflict Resolution, 54(2), pp.262–280. DOI: https://doi.org/10.1177/0022002709355439.

Enders, W. and Sandler, T. (2012). The political economy of terrorism. Cambridge ; New York: Cambridge University Press.

FATF (2018). Financing of Recruitment for Terrorist Purposes. [online] www.fatf-gafi.org. Available at: https://www.fatf-gafi.org/en/publications/Methodsandtrends/Financing-recruitment-terrorist-purposes.html.

Forrest, S. (1993). Genetic Algorithms: Principles of Natural Selection Applied to Computation. Science, 261(5123), pp.872–878. DOI:https://doi.org/10.1126/science.8346439.

Gutfraind, A. (2009). Understanding Terrorist Organizations with a Dynamic Model. Studies in Conflict & Terrorism, 32(1), pp.45–59. DOI: http://dx.doi.org/10.1080/10576100802627720.

Hassanat, A., Almohammadi, K., Alkafaween, E., Abunawas, E., Hammouri, A. and Prasath, V.B.S. (2019). Choosing Mutation and Crossover Ratios for Genetic Algorithms—A Review with a New Dynamic Approach. Information, 10(12), p.390. DOI: https://doi.org/10.3390/info10120390.

Horgan, J. (2005). Psychology of Terrorism. Routledge.

Kilberg, J. (2012). A Basic Model Explaining Terrorist Group Organizational Structure. Studies in Conflict & Terrorism, 35(11), pp.810–830. DOI: https://doi.org/10.1080/1057610x.2012.720240.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983). Optimization by Simulated Annealing. Science, 220(4598), pp.671–680. DOI: https://doi.org/10.1126/science.220.4598.671.

Kydd, A.H. and Walter, B.F. (2006). The Strategies of Terrorism. International Security, [online] 31(1), pp.49–80. DOI: https://doi.org/10.2307/4137539.

Landes, W.M. (1978). An Economic Study of U. S. Aircraft Hijacking, 1961-1976. The Journal of Law and Economics, 21(1), pp.1–31. DOI: https://doi.org/10.1086/466909.

Latora, V. and Marchiori, M. (2001). Efficient Behaviour of Small-World Networks. Physical Review Letters, 87(19), pp.1–4. doi:https://doi.org/10.1103/PhysRevLett.87.198701.

Li, Q. (2005). Does Democracy Promote or Reduce Transnational Terrorist Events? Journal of Conflict Resolution, 49(2), pp.278–297. doi:https://doi.org/10.1177/0022002704272830.

Lindelauf, R., Borm, P. and Hamers, H. (2009). The influence of secrecy on the communication structure of covert networks. Social Networks, 31(2), pp.126–137. DOI: https://doi.org/10.1016/j.socnet.2008.12.003.

Maragatham, S. (2024). Analysis of Graph Inverses and Algebraic Connectivity in a Systematic Way. Communications on Applied Nonlinear Analysis, 32(3), pp.525–538. DOI: https://doi.org/10.52783/cana.v32.2013.

Rapoport, D.C. (2013). The Four Waves of Modern Terror: International Dimensions and Consequences. [online] ResearchGate. Available at: https://www.researchgate.net/publication/286896869_The_four_waves_of_modern_terror_International_dimensio ns_and_consequences.

Snoek, J., Larochelle, H. and Adams, R.P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. arXiv:1206.2944 [cs, stat]. [online] Available at: https://doi.org/10.48550/arXiv.1206.2944.

Song, X., Zhao, P., Yin, R., Zu, Y. and Zhang, Y. (2024). Cascading failure model and resilience-based sequential recovery strategy for complex networks. Reliability Engineering & System Safety, [online] 253, p.110488. DOI: https://doi.org/10.1016/j.ress.2024.110488.

Vasiliev, A. (2022). The evolution of Al-Qaeda: between regional conflicts and a globalist perspective. Vostok. Afro-aziatskie obshchestva: istoriia i sovremennost, (1), p.141. DOI: https://doi.org/10.31857/s086919080015708-5.

Authors Bio:

As a recent Economics and Management graduate from the University of Bristol, my work aims to reconcile quantitative modelling approaches with problems in counter terrorism policy. Pursuing a masters level degree at the University of St Andrews in Terrorism and Political Violence, I believe that the model of the rational individual can be applied to those that pursue more nefarious goals, and in turn, there can be the generation of more effective, equitable, and efficient counter terrorism policy.

Law

Meet the Subject Editor

Senali Galappaththi
Law

Senali is currently pursuing an LLM in Law, Innovation and Technology at the University of Bristol and is a qualified Attorney-at-Law in Sri Lanka. Drawing on previous legal and research experience, she served as the Law Subject Editor for the Student Research Journal. It was an enriching experience for her to review research across a broad range of legal topics and contribute to the editorial review process, which required careful attention to detail, critical analysis, effective communication, and editorial leadership. Her personal research interests include data protection, privacy, and AI regulation. She hopes to further develop her expertise through professional experience in data protection, privacy, and information governance, with a view to building a career in these fields.

Is the Paris Agreement Enough? Human Rights and State Obligations in the Climate Crisis

Author: SosefoSepasitiano Patelisio

ORCID: 0009-0002-7811-3139

Programme: LLMInternational Law & International Relations

Affiliation: Law School

Email: sepapatelisio@gmail.com

Abstract

Climate change presents profound challenges not only to the environment but also to the protection of fundamental human rights, raising critical questions about the scope of State's legal obligations under international law. This article argues that compliance with the Paris Agreement alone is insufficient to discharge State's broader human rights obligation in addressing climate change. While some States contend that the climate treaty regime constitutes the exclusive legal framework governing climate obligations this article demonstrate that such a position misconstrues the relationship between treaty law and general principles of international law. It contends that principles such as a due diligence, prevention of transboundary harm, and the obligation not to cause significant environmental damage to continue to apply alongside the climate treaty regime. Drawing on jurisprudence from the International Court of Justice, regional human rights courts, and United Nations bodies, the article highlights the growing recognition that climate change directly threatens rights including the right to life, territorial integrity, and self-determination.

Ultimately, the article concludes that climate treaties operate within, rather than replace, the wider framework of international law, requiring States to adopt comprehensive measures that safeguard both environment and human rights.

Keywords: Climate change, Paris Agreement, human rights, general principles of international law, obligation of states in respect of climate change.

My deepest love and appreciation to my parents,Kosilio&PatiniaPatelisio,whosesacrifices,prayers,andunwavering support have been the foundation of everythingIhaveachieved.Tomysiblings,Patelisio,Polini&MikeNansen, Fapiano, & Neomai. Thank you for all your loveandassistance.
To Seini Mahe, Saane & Anatule Mahe, 'AmeliaFangupo,Solomone&LataFangupo,Tevita&MoliketiFangupo.To Penisimani & Toli'a Fifita, Sioeli & Sinifi Fifita,Solomone&SesikaFifita,'Okusitino,SolomoneFifita,Lilieta&Temo Mayanavanua, Elnathan, Jessica & MalakaiTu'ipulotu,Heilala,Samisoni,Kisione,'Aseli,Sela,Heilala&Hengihengi.
To Olivia & Mosese Latu, & Luseane Manu.
With heartfelt thanks to everyone acknowledgedabovefortheirlove,support,prayers,andencouragementthroughout this journey.

As climate change continues to pose an existential threat to both the environment and human well-being, it has triggered a contentious international debate over the scope of state's legal obligations in responding to its impacts. In the Advisory proceedings of *Obligations of States in respect of Climate Change*[1], Germany argued that compliance with the Paris Agreement is sufficient to fulfil state's human rights obligations related to climate change. The argument is central on the belief that the Paris Agreement, with its legally binding elements regarding mitigation, adaptation and finance, is enough to provide a comprehensive framework for addressing climate change, and thus sufficient to cover human rights obligations.[2] However, this position is flawed, as it disregards the additional obligations imposed by the general principles of international law (GPIL) in addressing climate change under the realm of international law and human rights.

This article will argue that fulfilling obligations under the Paris Agreement does not suffice for States to meet their obligations under human rights conventions in the context of climate change. It will first argue that the Paris Agreement should not be treated as the exclusive legal framework for climate action. It will then argue that state's climate obligations extend beyond the Paris Agreement, particularly under GPIL, which imposes duties such as due diligence, the prevention principle and the obligation not to cause significant transboundary harm. Lastly, it will examine the right to territorial integrity and the right of self- determination, highlighting the increasing recognition of climate change's impact on the enjoyment of fundamental human rights. Finally, the article will conclude by demonstrating that a narrow reliance on the Paris Agreement is insufficient to discharge state's broader legal obligations in combatting climate change and safeguarding human rights.

I. The Paris Agreement is not the primary legal framework for climate change

In international law, it is common that more than one rule addresses a specific issue; such parallelism occurs frequently.[3] Lindroos (2005) argues when multiple legal norms apply to the same issue, the principle of *lex specialis* serves to resolve any potential conflicts by giving priority to the more specific rule over the general one.[4]

In this context, Germany argued that the Paris Agreement and the United Nations Framework on Climate Change Convention (UNFCCC)—as specialized legal instruments

[1] UNSC RES 2023 (4 April 2023) UN Doc A/RES/ 2023.
[2] Written Statement of Germany, para 27.
[3] *Southern Bluefin Tuna (Australia and New Zealand v. Japan),* (2000), vol. 119, I.L.R., para. 52.
[4] Anja Lindroos, 'Addressing Norm Conflicts in a fragmented Legal System' (2005) 74(1) *NJIL* 27.

on climate change—should take precedence over broader human rights obligations when addressing climate-related issues.[5] Germany maintained that any potential normative conflict should be resolved by giving priority to the more specific legal framework.[6] Similarly, Australia argued that principles outside the treaty regime do not supplant or impose obligations upon States in relation to climate change.[7] Bangladesh[8] even contended that the UNFCCC regime should, on the basis of the *lex specialis* principle, displace, limit or modify what are considered to be the *leges generals* in these proceedings, i.e. important GPIL such as prevention of significant transboundary harm, the due diligence principle and the prevention of harm.

However, these arguments are misleading as the application of *lex specialis* is not automatic. States cannot invoke it merely because two or more legal regimes address the same subject matter. For *lex specialis* to apply, two key elements must be met; there must be an actual inconsistency between the two provisions[9]; or a clearly expressed intention that one is meant to exclude the other. As clarified by the International Law Commission;

> " it is not enough that the same subject matter is dealt with by two provisions; there must be some actual inconsistency between them, or else a discernible intention that one provision is to exclude the other"[10]

In relation to the first element, there is no inherent inconsistency between the GPIL, such as the prevention of transboundary harm, prevention of significant harm to the environment, due diligence and, on the other hand, the treaty regime addressing climate change. A state party to the climate treaty regime can fully comply with its obligations under that regime while simultaneously fulfilling the broader principles of GPIL.[11] It is not a matter of conflict but rather of complementarity, as the climate change legal regime seeks to prevent environmental harm that crosses borders, aligning with international legal obligations to prevent such damage.[12]Thus, these principles and the climate change regime are not only compatible but are harmoniously applied together.

---

[5] Written Statement of Germany, para 86.
[6] Ibid.
[7] Written Statement of Australia, para 2.2.
[8] Written Statement of Saudi Arabia, para. 5.8.
[9] Written Statement of Bangladesh, para.11.
[10] ARSIWA Commentary, Art. 55, Yearbook of the International Law Commission, 2001, vol. II, para. 4.
[11] Written Statement of Nauru, para. 11.
[12] Ibid.

This was echoed by Egypt:

> "the existence of treaties dealing with the subject of climate change does not preclude the application of the general rules of international law"[13]

However, Kuwait[14] argued that applying general principles such as transboundary harm could impose additional duties not expressly agreed to within the climate treaties, thereby violating the principle of State consent in treaty law.

Nevertheless, this counterargument fails because it mischaracterizes the relationships between climate treaties and GPIL. Under Article 31(3) (c) of the Vienna Convention on the Law of Treaties, treaties must be interpreted in light of relevant rules of law, unless explicitly excluded—which the Paris Agreement does not do. Moreover, Meguro (2020) argues, the notion that climate should be addressed exclusively through the treaty regime ignores the multidimensional nature of the crisis.[15] Climate change affects not only the environment but also human rights, sovereignty and global peace.[16] Climate treaties do not exist in a legal vacuum—they operate within the wider matrix of international law and must be interpreted as part of an integrated system of obligations.[17]

This point was raised by the Cook Islands:

> "the UNFCCC, including the Paris Agreement, is not a *lex specialis* regime that regulates all climate change action, and singularity provides for all of State's obligation in respect of climate change, and legal consequences under State's obligations, to the exclusion of all general rules of international law"[18]

Thus, the climate change treaty regime does not function as *lex specialis* in a manner that would displace or limit State's obligations under GPIL. Instead, the climate regime exists within the overarching framework of international law, complementing rather than overriding its general principles.

The second requirement–that there must be a clear intent to exclude GPIL —is plainly unmet. Nowhere in the UNFCCC, Kyoto Protocol or Paris Agreement is there any explicit

---

[13] Written Statement of Egypt, para. 71.
[14] Written Statement of Kuwait, para. 3.
[15] Maiko Meguro, 'Litigating climate change through international law: Obligations and rights strategy' (2022) 33(4) *Leiden Journal of International Law* 31.
[16] Ibid.
[17] *Meguro* (n 12).
[18] Written Statement of Cook Islands, para. 135.

provision that displaces core principles of GPIL.[19] International treaties cannot be assumed to implicitly override fundamental principles of customary international law (CIL) without express provisions to that effect. As Thirlway aptly described:

> "if a treaty provides that the parties, in their mutual relations, shall not be obliged to do something which CIL would otherwise have required of them, it is trite law that the obligation so set aside does not form part of the law between the parties. On the other hand, where the treaty is silent, CIL continues to apply"[20]

This position is not theoretical—it is supported by International Court of Justice (ICJ) jurisprudence. In *Construction of a Road in Costa Rica along the San Juan River,* Nicaragua argued that the 1858 Costa Rica-Nicaragua Treaty of Limits[21] constituted *lex specialis* concerning the obligation to consult under CIL.[22] However, the Court rejected the claim that a treaty could displace procedural obligations under CIL merely by containing some relevant provisions. It ruled that:

> "the fact that the 1858 Treaty may contain limited obligations concerning notification or consultation in specific situations does not exclude any other procedural obligations with regard to transboundary harm which may exist in treaty or customary international law"[23]

Likewise, in *ELSI,* the Court refused to accept that a fundamental principle of CIL could be silently waived, reaffirming that any departure must be clear and express. The Chamber rejected the notion of silent derogation of CIL, holding that it was:

> "unable to accept that an important principle of customary international law should be held to have been tacitly dispenses with, in the absence of any words making clear an intention to do so"[24]

Therefore, there is no discernible inconsistency between the climate change treaty regime and the relevant GPIL. Both legal frameworks operate in parallel, without contradicting each other, and there is no provision in either body of law that suggests any inherent conflict. Thus, it cannot be argued that the climate change treaties exempt States parties

---

[19] ARSIWA Commentary (n 10).
[20] Hugh Thirlway, Law and Procedure of International Court of Justice: *Fifty Years of Jurisprudence* I (2013), p. 127.
[21] *Costa Rica–Nicaragua, Treaty of Limits*, 15 April 1858, 118 C.T.S. 439.
[22] Ibid.
[23] Ibid 108.
[24] *Elettronica Sicula* S.p.A., Judgment, I.C.J. Reports 1989, para. 50.

from obligations that would otherwise be required under GPIL, nor is there any indication that the treaties were designed to exclude the application of GPIL.

II. Additional Climate Obligations extend beyond the Paris Agreement

Since it has been established that there is neither an actual inconsistency between the Paris Agreement and the relevant GPIL, nor a clear intention within the treaty text to exclude the application of broader norms, it follows that the principle of *lex specialis* does not preclude the operation of GPIL. If the Paris Agreement was the only source of obligations to combat climate change then any State that is not party to the climate change regime— or which has withdrawn from it— would have no legal obligations to protect the climate from anthropogenic greenhouse gas (GHG) emissions.[25] To limit these obligations solely to treaty law is to ignore the broader framework of international law, which not only reinforces these duties but also establishes accountability for States that fail to act.

a. Duty of Due diligence

The first obligation beyond the Paris Agreement, in fulfilling human rights obligations, is the duty of due diligence. The duty requires states to take all reasonable steps and measures to prevent significant harm to other states or the environment.[26] States are obligated to mitigate and remedy human rights impacts of climate change, including those arising from business activities within their territory, or jurisdiction.[27]

The Office of the High Commissioner for Human Rights (OHRHR) emphasized this;

> "The obligations of States in the context of climate change and other environmental harms extend to all rights-holders and to harm that occurs both inside and beyond boundaries."[28]

The *Pulp Mills on the River Uruguay (Argentina v. Uruguay),* clarified this;

> "It is 'every State's obligation not to allow knowingly its territory to be used for acts contrary to the rights of other States"[29]

---

[25] Written Statement of Costa Rico, para.14.
[26] Working Group on the issue of human rights and transnational corporations and other business enterprises, Information Note on Climate Change and the Guiding Principles on Business and Human Rights (June 2023).
[27] Ibid.
[28] OHCHR, *Understanding Human Rights and Climate Change*, Submission of the OHCHR to the 21st Conference of the Parties to the UNFCCC (2021), 3.
[29] *Pulp Mills on the River Uruguay (Argentina v. Uruguay), Judgment, I.C.J. Reports 2010 (I),* para. 101.

However, Australia contended that due diligence is a non-self-standing legal obligation but merely a standard of conduct that applies within the framework of other substantive obligations.[30] In circumstances where a State is required to achieve a particular outcome, due diligence is restrained in directing how the State should act to meet that goal.[31] The Seabed Disputes Chamber of ITLOS supports this, stating that an obligation 'to ensure' 'is not an obligation to achieve... rather, it is an obligation to deploy adequate means, to exercise best possible efforts, to do the utmost, to obtain this result.[32]

However, this interpretation is flawed. To suggest that due diligence only exists in relation to other obligations undermines its preventive function. The *Pulps* case affirmed that States have an ongoing duty to act with care and vigilance to prevent transboundary harm-regardless of whether a specific result is ultimately achieved.[33]Framing due diligence merely as a standard of effort ignores its legal weight as a proactive obligation and dilutes the States' responsibility to take meaningful action.

Fulfilling due diligence is important for upholding human rights. The link between due diligence and human rights was proclaimed by the UN Resolution 76/300 which;

> "Reaffirm[ed] that States have the obligations to respect, protect and promote human rights, including in all actions undertaken to address environmental challenges, and to take measures to protect the human rights of all"[34]

Similarly, in its commentary on Article 6 of the International Covenant on Civil and Political Rights (ICCPR), which protects the right to life, the United Nations Human Rights Committee emphasized that States must take legislative and other measures to

> "ensure that all activities taking place in whole or in part within their territory and in other places subject to their jurisdiction, but having a direct and reasonably foreseeable impact on the right to life of individuals outside their territory...are consistent with article 6"

In *Cáceres et al v Paraguay*, the UN Human Rights Committee held that Paraguay violated Articles 6 of the ICCPR. This was because of its failure to impose environmental laws

---

[30] Written Statement of Australia, para. 4.13.
[31] Ibid.
[32] *International Tribunal for the Law oftheSea (ITLOS), Responsibilities and Obligations of States Sponsoring Persons andEntities to Activities in the Area (Advisory Opinion) (2011) ITLOS Rep 10*, para. 110.
[33] *PulpMills* (n 28).
[34] UNSC RES 2022 (28 July 2022) UN Doc S/RES/ 76/300.

regulating the usage of poisonous agrochemicals, which led to the contamination of drinking water and food resources. The contamination resulted in the author's hospitalization and the unexplained death of Mr. Cacerres. Thus, Paraguay's failure to act amounted to a breach of its obligations.[35] Likewise, in *Poma Poma v Peru*, the UN Human Rights Committee found a violation of Article 27 (rights of minority groups to preserve their cultural, religious and linguistic identity) of ICCPR. Peru's water operations produced environmental harm that disrupted the author's traditional livelihood of llama herding. No prior consultation was conducted with the author or her indigenous community. The lack of participation led the Committee to find Peru in breach of its human rights obligations. Both examples demonstrate the importance of due diligence in upholding human rights and for the safety of communities.

Although the duty of due diligence is not explicitly stated in Resolution 76/300, the obligation to protect and take measures to uphold human rights plainly aligns with its core principles. Resolution 76/300 aligns with various human rights conventions that recognize environmental protection as a human right such as the Additional Protocol to the American Convention on Human Rights in the area of Economic, Social, and Cultural Rights[36], the Charter of Fundamental Rights of the European Union[37], the Arab Charter on Human Rights[38] and the African Charter on Human and Peoples' Rights.[39] These instruments uphold that states have obligations beyond the Paris Agreement, comprising both negative obligations (to refrain from harming the environment) and positive obligations (to take proactive measures for its protection) in environmental governance.

b. Duty of Prevention

The next obligation is the prevention of transboundary harm to the environment. The prevention principle is outlined in Principle 21 of the 1972 Stockholm Declaration on the

[35] UN Doc CCPR/C/126/D/2751/2016 (20 September 2019) para 7.1-7.9.
[36] "Everyone shall have the right to live in a healthy environment and to have access to basic public services…The State Parties shall promote the protection, preservation and improvement of the environment" (Article 11).
[37] "A high level of environmental protection and the improvement of the quality of the environment must be integrated into the policies of the Union and ensured in accordance with the principle of sustainable development" (Article 37).
[38] "Everyone shall have the right to an adequate standard of living for himself and his family, ensuring well-being and a decent life, including adequate food, clothing, housing, services and a right to a safe environment. The State Parties shall take appropriate measures within their available resources to ensure the realization of this right" (Article 38).
[39] "All people shall have the right to a general satisfactory environment favourable to their development" (Article 24)

Human Environment[40] and Principle 2 of the 1992 Rio Declaration on Environment and Development.

The Duty of Prevention principle is one of general application. As Kiss (1989) observed, the ICJ in *Corfu Channel*[41] established a general principle of international law: every state has an obligation not to knowingly allow its territory to be used in a manner that violates the rights of other states.[42]It obliges States 'to exercise best possible efforts, to do the utmost'[43] to prevent activities which take place in its territory, or in any area under its jurisdiction, from causing significant harm to other States.

Importantly, the *Trail Smelter*'s declaration—that a State has no right to permit activities within its borders that cause serious injury to another, especially when such harm is proven by clear and convincing evidence[44]—serves as a compelling precedent affirming the human right obligation to prevent transboundary environmental harm under international law.

However, Australia disputed that the prevention principle, usually applied to localized transboundary harm, cannot be directly applied to climate change due to its complex aggregate nature.[45] Unlike conventional pollution cases, climate change concerns emissions from multiple sources over time, presenting causation and attribution difficult.[46] This reasoning, while recognizing the complexity of climate change, incorrectly assumes that the diffuse nature of harm excuses States from their human rights obligations. Complexity does not negate responsibility. The prevention principle is a cornerstone of GPIL and has evolved to address environmental threats— including those with far-reaching cumulative impacts. The ICJ has repeatedly affirmed States duty to prevent transboundary harm, regardless of whether it arises from a single source or a global phenomenon.[47] Accepting Australia's argument would create a legal vacuum in which States could evade responsibility for contributing to climate change simply because it is a collective problem.

---

[40] United Nations Conference on the Human Environment, Stockholm Declaration on the Human Environment, available in (1972) 11 ILM 1416 (June 5–16, 1972), Principle 21.
[41] (United Kingdom v. Albania), Merits, Judgment, I.C.J. Reports 1949.
[42] Alexandre Kiss, Droit international de l'environnement (1989), 73.
[43] *Pulp Mills* (n 28) 101.
[44] *Trail Smelter Arbitration*, Decisions of 11 March 1941, vol. III, RIAA, 1965.
[45] Written Statement of Australia, para 4.10.
[46] Ibid.
[47] *Pulp Mills on the River Uruguay* (n 32).

The European Court of Human Rights (ECtHR) has further affirmed the prevention duty as a human rights obligation by interpreting provisions of the European Convention on Human Rights (ECHR). The Court acknowledged that states must ensure a healthy environment as part of their duty to uphold the right to life (Article 2) and the right to private and family life (Article 8). In *Tatar v Romania*[48], the Court set out a "positive obligation to take all reasonable and suitable measures to protect the rights [of the applicants to a private life]" which entails "the paramount duty to put in place a legislative and administrative framework for the effective prevention of damage to the environment and to human health"[49]. Similarly, in *Duarte Agostinho and Others v Portugal*[50], the applicants argue that Portugal's recurring forest fires since 2017 are a direct consequence of climate change, causing them harm and posing serious health risks. They claim that the 33 respondent states have violated Articles 2 and 8 of the ECHR by failing to take adequate climate action, as outlined in the 2015 Paris Agreement. This case was brought before the Grand Chamber, which had not yet issued its judgement at the time of writing. Thus, the growing body of jurisprudence underscores the prevention duty as a human rights obligation in addressing environmental harm.

Environmental harm can have severe impacts on human rights, hindering the enjoyment of the human rights.[51] Thus, the prevention principle becomes essential in safeguarding human rights. The ECtHR has affirmed this principle, acknowledging that environmental degradation can threaten the rights to life[52], respect for private and family life[53], and property[54]. Accordingly, the prevention duty provides a robust legal basis for safeguarding human rights to individuals both within and beyond their borders.

c. Obligation not to cause significant transboundary harm

Another obligation is the duty to not cause significant transboundary harm. A state's jurisdiction rests on its effective control over activities conducted within its territory or authority. It is responsible for preventing those activities from causing harm beyond its

[48] ECHR, Judgment of 27 January 2009
[49] Ibid 88.
[50] *Duarte Agostinho and Others v. Portugal and 32 other States*, No. 39371/20
[51] *Kawas Fernández v. Honduras. Merits*. Judgment of April 3, 2009. 196. 148.
[52] *Öneryildiz v. Turkey* [GS], No. 48939/99. Judgment of November 30, 2004, 71.
[53] *López Ostra v. Spain*, No. 16798/90. Judgment of December 6, 1994, 51.
[54] *Papastavrou and Others v. Greece*, No. 46372/99. Judgment of April 10, 2003, 33.

borders, particularly when such harm affects the rights of individuals in other states.[55] The Inter-American Court of Human Rights (IACtHR) clarified this;

> 'States have an obligation to avoid transboundary environmental damage that can affect the human rights of individuals outside their territory'[56]

In *Carême v. France*[57] the ECtHR reaffirmed the duty not to cause transboundary harm as part of France's positive obligations under the ECHR[58]. The applicant alleged that France has taken insufficient steps to prevent climate change and this failure entailed a violation of Article 2 and Article 8 under the Convention.[59] While the Court did not find a violation in this case because the Claimant's residence was located in an area subject to flooding by 2040,[60] the Court recognized the duty not to cause significant transboundary harm, reflecting the growing recognition that environmental protections are integral to fulfilling human rights obligations.

Thus, the Courts have indicated that there are certain activities that involve significant risks to the health of the individual, and therefore, States have a specific obligation to regulate them such as the introduction of monitoring and oversight mechanisms.[61]

III. The Right to Territorial Integrity (TI) and The Right to Self-Determination

Every State is entitled to be respected for its TI—"respect for territorial sovereignty is an essential pillar of international relations".[62] As the Court has affirmed, "the respect of territorial integrity is an important part of the international legal order and is enshrined in the Charter of the United Nations."[63] Respect for sovereignty not only affirms a State's rights over its own territory but also imposes a corollary duty: no State may use its territory, or permit its use, to carry out acts that infringe upon the rights of other States. This obligation is intrinsic to the very notion of sovereignty, for one State cannot invoke its sovereign rights to undermine the sovereignty of another. Sovereignty, therefore, entails a

---

[55] Inter-American Court of Human Rights, *The Environment and Human Rights*, Advisory Opinion OC-23/ 15 November 2017, 102.
[56] Ibid.
[57] *Carême v. France* (no. 7189/21).
[58] Ibid, 73.
[59] *Carême* (n 45) 3.
[60] *Carême* (n 44) 79.
[61] African Commission on Human and Peoples' Rights, *Social and Economic Rights Center (SERAC) and Center for Economic and Social Rights (CESR) v. Nigeria*. Communication 155/96. Decision of October 27, 2001, para. 53.
[62] *Corfu Channel* (n 42) 35.
[63] Accordance with *International Law of the Unilateral Declaration of Independence in Respect of Kosovo*, Advisory Opinion, I.C.J. Reports 2010, p. 437, para. 80.

duty to ensure that the rights of other States—particularly their TI— are safeguarded within one's own jurisdiction.[64]

The duty to respect a State's territorial sovereignty is not confined to its land territory alone. As the Court in *Military and Paramilitary Activities in and against Nicaragua* affirmed*, "*the principle of respect for State sovereignty"—enshrined in Article 2(1) of the UN Charter—entails an obligation on all States to "respect the territorial sovereignty of others".[65] This obligation extends beyond land to encompass a State's "internal waters, territorial sea, and the airspace above its territory".[66]

However, Australia argued that TI, focused on preserving state sovereignty and borders, is unrelated to human rights obligations in addressing climate change, which are grounded in individual protections like the rights to life and health.[67] However, this view overlooks how transboundary climate impacts can directly threaten the TI of vulnerable states and the rights of their people. As affirmed in ICJ jurisprudence[68], no state may use its territory in a way that causes harm to others. Climate change does just that, undermining both TI and human rights, such as the rights to self-determination. Thus, the two are inherently connected in today's global climate reality.

The right to TI is intrinsically linked to the right to self-determination, describing it as a "corollary" and a "key element in the exercise" of that right under international law.[69] It is through the peremptory norm of general international law[70] that people are entitled to freely determine their political states and pursue their economic, social and cultural development.[71] The right to self-determination is a fundamental human right enshrined in Article 1 common to the UN Charter, ICCPR[72] and the International Covenant on Economic, Social and Cultural Rights (ICESCR).[73]

---

[64] *Island of Palmas (Netherlands, United States of America)* (1928), vol. II, R.I.A.A., p. 839.
[65] Merits, I.C.J.Reports 1986,para.213.
[66] Ibid 212.
[67] Written Commentsof Australia,para4.14-4.15.
[68] *MilitaryandParamilitary Activitiesinand againstNicaragua* (n 64).
[69] *Legal Consequences of the Separation of the Chagos Archipelago from Mauritius in 1965*, Advisory Opinion, I.C.J. Reports, p. 134, para. 160.
[70] Draft Conclusions on Identification and Legal Consequences of Peremptory Norms of General International Law (Jus cogens), Yearbook of the International Law Commission, 2022, vol. II, Part Two.
[71] Common Article 1, paragraph 1, International Covenant on Civil and Political Rights, 16 December 1966, 999 U.N.T.S. 171, and International Covenant on Economic and Social Rights, 16 December, 999 U.N.T.S. 3.
[72] 16 December 1966, 999 U.N.T.S. 171.
[73] 16 December 1966, 993 U.N.T.S. 3.

Every State, under the principle of self-determination is obligated to refrain from 'any action which would dismember or impair, totally or in part, the territorial integrity or political unity of sovereign and independent States conducting themselves in compliance with the principle of equal rights and self-determination of peoples'[74]

In 2009, the United Nations Human Rights Council recognized the right to self-determination as one of the rights most vulnerable to the impacts of climate change. Unlike most individual human rights, self-determination—as enshrined in Common Article 1 of both the ICCPR and ICESCR— is a collective right. It imposes obligations not just on states towards individuals within their territory, but also their conduct toward other states[75], making it a right with corresponding inter-state duties. Additionally, the right under Common Article 1 of the Covenants holds a distinct position, preceding individual rights listed in Part II of both covenants. Unlike those individual rights, Article 1 is not constrained by the territorial limitations set out in Article 1. Instead, it applies beyond a State's own territory, without restriction. As the court has affirmed, the obligations under this provision extend to the actions of a State wherever it implements authority, emphasizing its broader extraterritorial reach.[76]

However, the U.S. disputed extending the right to self-determination to climate change, contending there is no historical foundation for observing GHG emissions as violating this right—especially not in relation to peoples outside a state's own borders.[77] However, the U.S. misses the legal recognition of this right. The Court has constantly declared that this right is an obligation *erga omnes*[78], and the International Law Commission has recognized its standing as a norm of jus cogens.[79] Accordingly, all States— despite of whether they are party to any particular treaty—are obliged to respect and guarantee the right, with no derogations allowed. That is so regardless of whether and when a State has signed and ratified any related human rights treaty.

---

[74] United Nations General Assembly resolution 2625 (XXV), Declaration on Principles of International Law concerning Friendly Relations and Co-operation among States, 24 October 1970.
[75] CCPR General Comment No 12: Article 1 (Right to Self-determination), 13 March 1984, para. 5.
[76] Application of the International Convention on the Elimination of all Forms of Racial Discrimination (Georgia v. Russian Federation), Provisional Measures, Order of 15 October 2008, I.C.J. Reports 2008, para. 109.
[77] Written Comments of the United States of America, para. 4.64.
[78] *East Timor (Portugal v Australia)* (Judgment) [1995] ICJ Rep 90, [29].
[79] International Law Commission, Report of the International Law Commission on the Work of Its Seventy-third Session, UN GAOR, 73rd sess, Supp No 10, UN Doc A/77/10 (2022), 16 ('ILC Report on Seventy-third session').

Conclusion

In conclusion, while the Paris Agreement plays a crucial role in global climate governance, it cannot be regarded as the sole legal framework governing states' obligation in addressing climate change and fulfilling human rights obligations. GPIL imposes further duties such as due diligence, the prevention principle and the obligation not to cause significant transboundary harm. Climate change creates direct threats to basic human rights such as the right to territorial integrity and to the right of self-determination for vulnerable communities whose livelihoods are threatened. This highlights the need for state action beyond the Paris Agreement. Depending exclusively on the Paris Agreement to fulfil climate-related human rights obligations would not only undermine legal accountability but also weaken well-established principles of international law. A thorough approach that incorporates GPIL, human rights law and environmental principles is necessary to safeguarding that states wholly encounter their obligations in tackling climate change and safeguarding fundamental human rights.

References

Cases

1. African Commission on Human and Peoples' Rights, *Social and Economic Rights Center (SERAC) and Center for Economic and Social Rights (CESR) v. Nigeria*. Communication 155/96. Decision of October 27, 2001.
2. Application of the International Convention on the Elimination of all Forms of Racial Discrimination (*Georgia v. Russian Federation*), Provisional Measures, Order of 15 October 2008, I.C.J. Reports 2008.
3. *Budayeva and Others v. Russia*, European Court of Human Rights (ECHR), No. 15339/02.
4. *Carême v. France* (App no 7189/21 (ECtHR, 28 March 2024).
5. *Case of Kawas Fernández v. Honduras. Merits,* reparations and costs. Judgment of April 3, 2009. Series C No. 196. 148.
6. *Certain activities carried out by Nicaragua in the border area (Costa Rica v. Nicaragua*) Judgment of December 16, 2015, para. 153.
7. *Costa Rica–Nicaragua, Treaty of Limits*, 15 April 1858, 118 C.T.S. 439.
8. *Corfu Channel Case (UK v Albania) (Merits)* [1949] ICJ Rep 4.
9. *Duarte Agostinho and Others v. Portugal and 32 other States*, No. 39371/20.
10. *East Timor (Portugal v Australia)* (Judgment) [1995] ICJ Rep 90.
11. *Elettronica Sicula* S.p.A., Judgment, I.C.J. Reports 1989.
12. Inter-American Court of Human Rights, *The Environment and Human Rights*, Advisory Opinion OC-23/17 (English translation), 15 November 2017, 102.
13. *International Law of the Unilateral Declaration of Independence in Respect of Kosovo*, Advisory Opinion, I.C.J. Reports 2010.
14. International Tribunal for the Law of the Sea, *Responsibilities and Obligations of States Sponsoring Persons and Entities with respect to Activities in the Area (Advisory Opinion)* (2011) ITLOS Rep 10.
15. *Island of Palmas (Netherlands, United States of America)* (1928), vol. II, R.I.A.A.
16. *Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, I.C.J. Reports 1996.*
17. *Legal Consequences of the Separation of the Chagos Archipelago from Mauritius in 1965*, Advisory Opinion, I.C.J. Reports, p. 134, para. 160.
18. *López Ostra v. Spain*, No. 16798/90. Judgment of December 6, 1994, 51.
19. *Military and Paramilitary Activities in and against Nicaragua* Merits, I.C.J. Reports 1986.
20. *Öneryildiz v. Turkey* [GS], No. 48939/99. Judgment of November 30, 2004, 71.

21. *Papastavrou and Others v. Greece*, No. 46372/99. Judgment of April 10, 2003, 33.
22. *Pulp Mills on the River Uruguay (Argentina v. Uruguay), Judgment,* I.C.J. Reports 2010.
23. *Southern Bluefin Tuna (Australia and New Zealand v. Japan),* (2000), vol. 119, I.L.R.,
24. *Tatar v Romania* ECHR, Judgment of 27 January 2009.
25. *Trail Smelter Arbitration*, Decisions of 16 April 1938 and 11 March 1941, vol. III, RIAA, 1905–82, 1965.

Journals

26. Anja Lindroos, 'Addressing Norm Conflicts in a fragmented Legal System' (2005) 74(1) *NJIL* 27.
27. ARSIWA Commentary, Art. 55, Yearbook of the International Law Commission, 2001, vol. II, para. 4.
28. Kiss, A. Droit international de l'environnement (1989), 73.
29. Meguro M, 'Litigating climate change through international law: Obligations strategy and < rights strategy' (2022) 33(4) Leiden Journal of International Law, 31 https://www.cambridge.org/core/journals/leiden-journal-of-international-law/article/litigating-climate-change-through-international-law-obligations-strategy-and-rights-strategy/C0FD6771513D997AAAF33E109BF09880> accessed 13 February 2025.
30. OHCHR, *Understanding Human Rights and Climate Change*, Submission of the OHCHR to the 21st Conference of the Parties to the UNFCCC (2021), 3.
31. Thirlway H, Law and Procedure of International Court of Justice: *Fifty Years of Jurisprudence* I (2013), p. 127.

Web Pages

32. Australia, 'Written Statement of Australia' (*Obligations of States in Respect of Climate Change,* 22 March 2024)< https://www.icj-cij.org/sites/default/files/case-related/187/187-20240326-wri-02-00-en.pdf > accessed 3 March 2025
33. Bangladesh, 'Written Statement of the People's Republic of Bangladesh' (*Obligations of States in Respect of Climate Change,* 22 March 2024)< https://www.icj-cij.org/sites/default/files/case-related/187/187-20240325-wri-01-00-en.pdf > accessed 3 March 2025.
34. China, 'Written Statement of China' (*Obligations of States in Respect of Climate Change,* 22 March 2024)< https://www.icj-cij.org/sites/default/files/case-related/187/187-20240322-wri-19-00-en.pdf > accessed 3 March 2025.

35. Colombia, 'Written Statement of Colombia' (*Obligations of States in Respect of Climate Change,* 11 March 2024)< https://www.icj-cij.org/sites/default/files/case-related/187/187-20240311-wri-01-00-en.pdf > accessed 3 March 2025.
36. Cook Islands, 'Written Statement of the Cook Islands' (*Obligations of States in Respect of Climate Change,* 22 March 2024)< https://www.icj-cij.org/sites/default/files/case-related/187/187-20240815-wri-01-00-ene.pdf > accessed 3 March 2025.
37. Egypt, 'Written Statement of Egypt' (*Obligations of States in Respect of Climate Change,* 22 March 2024)< https://www.icj-cij.org/sites/default/files/case-related/187/187-20240322-wri-24-00-en.pdf > accessed 3 March 2025.
38. Freestone D, 'Request for an Advisory Opinion submitted by the Commission of Small Island States on Climate Change and International Law, Case 31 (2024) <https://itlos.org/en/main/cases/list-of-cases/request-for-an-advisory-opinion-submitted-by-the-commission-of-small-island-states-on-climate-change-and-international-law-request-for-advisory-opinion-submitted-to-the-tribunal/> accessed 31 January 2025.
39. Germany, 'Written Statement of Germany' (*Obligations of States in Respect of Climate Change,* 22 March 2024 <https://www.icj-cij.org/sites/default/files/case-related/187/187-20240326-wri-01-00-en.pdf > accessed 3 March 2025.
40. Kingdom of Saudi Arabia, 'Written Statement of the Kingdom of Saudi Arabia' (*Obligations of States in Respect of Climate Change,* 21 March 2024)< https://www.icj-cij.org/sites/default/files/case-related/187/187-20240321-wri-08-00-en.pdf > accessed 3 March 2025.
41. Kuwait, 'Written Statement of Kuwait' (*Obligations of States in Respect of Climate Change,* 22 March 2024 < https://www.icj-cij.org/sites/default/files/case-related/187/187-20240322-wri-14-00-en.pdf > accessed 3 March 2025.
42. Nauru, 'Written Statement of Nauru' (*Obligations of States in Respect of Climate Change,* 22 March 2024 < https://www.icj-cij.org/sites/default/files/case-related/187/187-20240322-wri-36-00-en.pdf > accessed 3 March 2025.
43. Saudi Arabia, 'Written Statement of Saudi Arabia' (*Obligations of States in Respect of Climate Change,* 21 March 2024 < https://www.icj-cij.org/sites/default/files/case-related/187/187-20240321-wri-08-00-en.pdf > accessed 3 March 2025.
44. Working Group on the issue of human rights and transnational corporations and other business enterprises, Information Note on Climate Change and the Guiding Principles on Business and Human Rights (June 2023).

45. World Youth for Climate Justice, 'Climate Justice Proceedings at the ICJ' (Pacific Islands Students Fighting Climate Change, 27 December 2024) <https://www.pisfcc.org/> Accessed 2 April 2025.

International Instruments

46. HRC Res 41/21 (23 July 2019) UN Doc HRC/RES/41/21.
47. International Covenant on Civil and Political Rights, 16 December 1966, 999 U.N.T.S. 171.
48. International Covenant on Economic and Social Rights, 16 December, 999 U.N.T.S. 3.
49. International Law Commission, Report of the International Law Commission on the Work of Its Seventy-third Session, UN GAOR, 73rd sess, Supp No 10, UN Doc A/77/10 (2022), 16 ('ILC Report on Seventy-third session').
50. *Obligations of States in respect of Climate Change* UNSC RES 2023 (4 April 2023) UN Doc A/RES/ 2023.
51. United Nations General Assembly resolution 2625 (XXV), Declaration on Principles of International Law concerning Friendly Relations and Co-operation among States, 24 October 1970.
52. UNSC RES 2022 (28 July 2022) UN Doc S/RES/ /76/300.
53. UNSC RES 2023 (4 April 2023) UN Doc A/RES/ 2023.

Authors Bio:

Sosefo Sepasitiano Patelisio is a 2026 graduate of the University of Bristol, holding a Master of Laws (LLM) in International Law and International Relations. He currently serves as an Assistant Crown Counsel at the Attorney General's Office of Tonga, where his work focuses on legal advice, public international law, and government policy matters. Originally from the villages of Kolonga and Tatakamotonga in Tonga, he is a proud son of Kosilio and Patinia Patelisio, whose sacrifices, guidance, and unwavering support have shaped his journey. His research interests include international law, climate change law, and issues affecting Small Island Developing States in the Pacific. Beyond his legal career, he enjoys farming and planting crops, which keep him connected to his family, community, and cultural heritage.

Climate Refugees and International Legal Gaps: Deriving a Minimum Standard of Protection from Existing International Law

Authors:Thi Ngoc Mai Nguyen and Le Bao Ngan Pham
ORCIDs:0009-0002-0041-7319 | 0009-0007-4761-4033
Programmes:LLM International Law | MSc in International Business Law and Arbitration
Affiliations:University of Bristol Law School | University of Silesia in Katowice
Email:nguyenthingocmai.forwork@gmail.com and baonganpl287@gmail.com

Abstract

Climate change is increasingly driving cross-border displacement, but international law does not recognize a specific legal status for those forced to migrate due to environmental factors. Although the term “climate refugee” is widely used in academic and policy debates, it remains absent from formal refugee law. This article explores whether a minimum standard of protection can be derived from existing legal frameworks despite this gap. Using doctrinal analysis and treaty interpretation under the Vienna Convention on the Law of Treaties, it examines international refugee law, international human rights law, and complementary protection regimes. Although the 1951 Refugee Convention remains limited by its persecution-based structure, existing international law is not entirely silent on climate-induced displacement. Narrow interpretive pathways within refugee law, particularly through the “particular social group” category, may offer protection in specific contexts. However, more substantial protection increasingly emerges from international human rights law, especially through non-refoulement obligations linked to Articles 6 and 7 of the ICCPR. This article argues: despite the absence of a dedicated legal regime, an emerging minimum standard of protection can already be identified across overlapping international legal frameworks.

Keywords: Climate-induced displacement, Climate refugees, Minimum standard of protection, Non-refoulement, International human rights law

INTRODUCTION

The twenty-first century faces an unprecedented crisis of mass displacement driven by climate change, as rising sea levels, desertification, and extreme weather force millions to cross borders in search of safety.[1] Despite the magnitude of this phenomenon, a profound legal paradox persists at the heart of international law. While the socio-political term "climate refugee" has gained widespread currency, it remains a "legal ghost" within the current multilateral framework.[2] Specifically, Article 1A(2) of the 1951 Refugee Convention limits refugee status to individuals with a well-founded fear of persecution for reasons of race, religion, nationality, membership of a particular social group, or political opinion, none of which explicitly includes environmental or climate-related harm.[3][4] This gap leaves climate-displaced individuals in legal limbo, as they are neither formally recognized as refugees nor entirely excluded from protection, given the growing role of human rights law and complementary protection in preventing return to life-threatening conditions.

This paper examines whether existing international law can provide a minimum standard of protection for climate-displaced persons despite the absence of formal recognition. It argues that refugee law, human rights law, complementary protection, and soft law collectively reveal an emerging framework grounded in non-refoulement and human dignity, particularly through the interpretive possibilities of the 1951 Refugee Convention and Articles 6 and 7 of the ICCPR in light of *Teitiota v New Zealand*.

CONCEPTUAL AND DOCTRINAL CLARIFICATION

Legally, the refugee-migrant distinction is grounded in specific protections under international law. Article 1A(2) of the 1951 Refugee Convention defines refugees as

---

[1] United Nations High Commissioner for Refugees, 'Climate Change and Displacement' (UNHCR) https://www.unhcr.org/what-we-do/build-better-futures/climate-change-and-displacement accessed 28 February 2026.

[2] Nischala McDonnell, 'Falling through the Protection Gaps: Inappropriate Protection of Climate Displaced Persons in the International Refugee Legal Structure' (2024) 10 *Transnational Human Rights Review* 11, 13-15

[3] Convention Relating to the Status of Refugees (adopted 28 July 1951, entered into force 22 April 1954) 189 UNTS art 1A(2).

Sarah Locke, 'Who Defines the Refugee?' (2025) 28, 31-33.

individuals who, owing to a well-founded fear of persecution for reasons of race, religion, nationality, membership of a particular social group, or political opinion, are outside their country of nationality and unable or unwilling to avail themselves of its protection.[5] Therefore, refugees are entitled to particular forms of international protection, most notably the principle of non-refoulement under Article 33(1), which prohibits States from returning individuals to territories where their life or freedom would be threatened.[67] In contrast, “migrant” has no clear legal definition and broadly refers to people moving for various reasons, while refugees are distinct due to their specific protection under international law when state protection fails.[8] A strict refugee-migrant distinction, reinforced by narrow interpretations of the 1951 Convention, excludes many fleeing conflict, disasters, or climate harm from protection. In this context, climate-displaced persons are often described as “legal ghosts” because, despite facing severe threats to life, subsistence, or human dignity, they frequently fall outside the formal scope of refugee protection under existing international law.[9] Within this framework, the notion of a “minimum standard of protection” assumes central importance. This standard is not codified in a single instrument but derives from multiple branches of international law. Under the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights, it refers to core guarantees such as the right to life, the prohibition of torture and cruel, inhuman or degrading treatment, and protection against removal where a real risk of serious harm exists.[10] In refugee law, as reflected in the 1951 Convention, it includes fundamental guarantees such as non-refoulement, access to courts, and certain socio-economic rights. In international humanitarian law, particularly under the Geneva Conventions, it denotes basic assurances of humane treatment during armed conflict. Even in international investment law, the customary “minimum standard of

[5] Convention Relating to the Status of Refugees (adopted 28 July 1951, entered into force 22 April 1954) 189 UNTS 137 art 1A(2).

[7] ibid art 33(1).
Annick Pijnenburg and Conny Rijken, ‘Moving beyond Refugees and Migrants: Reconceptualising the Rights of People on the Move’ (2021) 23 Interventions 273, 276 <https://doi.org/10.1080/1369801X.2020.1854107>.

[8] ibid.

[9] Nischala McDonnell, ‘Falling Through the Protection Gaps: Inappropriate Protection of Climate Displaced Persons in the International Refugee Legal Structure’ (2024) 10 The Transnational Human Rights Review <https://doi.org/10.60082/2563-4631.1103> accessed 16 May 2026.

[10] Pierre Thielbörger, ‘The “Essence” of International Human Rights’ (2019) 20 German Law Journal 924, 930-933.

treatment" requires States to refrain from arbitrary conduct and denial of justice.[11] Although these branches of international law differ in structure and purpose, they share a common concern with preventing serious harm and limiting arbitrary State conduct. Across these regimes, human dignity functions as a foundational principle, while non-refoulement operates as a key protection against returning individuals to conditions threatening life or basic human dignity. Thus, the minimum standard operates less as a status regime than as a negative obligation to prevent serious harm. The resulting protection gap is not a legal vacuum but a fragmented system in which climate-displaced persons fall outside refugee status while remaining partially protected under overlapping legal regimes. Although this fragmentation creates inconsistency and reliance on State discretion, Article 31 of the Vienna Convention supports evolutionary interpretation, allowing refugee law to develop consistently with human rights and environmental law without creating obligations beyond the treaty text. Thus, the challenge lies not in the absence of legal norms but in achieving greater interpretative coherence across existing frameworks.

INTERNATIONAL REFUGEE LAW (1951) - LIMITS AND LATENT POSSIBILITIES

Structural Limits of the 1951 Convention

The starting point of any analysis remains the Convention relating to the Status of Refugees. Article 1A(2) defines a refugee as a person who, owing to a well-founded fear of being persecuted, is outside their country of nationality and unable or unwilling to avail themselves of its protection.[12] Persecution must be linked to one of five enumerated grounds: race, religion, nationality, membership of a particular social group, or political opinion.[13] However, climate change does not appear among these grounds. Nor do environmental degradation, natural disasters, or sea-level rise. The Convention's drafting history confirms that its drafters were concerned primarily with

[11] Martins Paparinskis, 'International Minimum Standard and the Protection of Property' in Martins Paparinskis, *The International Minimum Standard and Fair and Equitable Treatment* (1st ed, Oxford University PressOxford 2013) 271-259 <https://academic.oup.com/book/11589/chapter/160455070> accessed 21 February 2026.

[12] Hélène Lambert, 'Refugee Status, Arbitrary Deprivation of Nationality, and Statelessness within the Context of Article 1A(2) of the 1951 Convention and its 1967 Protocol relating to the Status of Refugees' (UNHCR, Legal and Protection Policy Research Series, Division of International Protection, Geneva 2014) 16.

[13] Maryellen Fullerton, 'A Comparative Look at Refugee Status Based on Persecution Due to Membership in a Particular Social Group' (1993) 26 *Cornell International Law Journal* 505, 511.

post-war political persecution. The structure of the definition reflects this historical context.[14]

The absence of climate change as a ground creates a structural limit, as refugee status requires a link to one of the five Convention grounds, which generalized environmental harm usually cannot meet.[15] Most climate-based asylum claims fail because environmental harm is seen as affecting entire populations, not targeting individuals on Convention grounds, and thus is not recognized as independent persecution.[16] This limitation does not make the Convention irrelevant: while climate change is not a ground, climate-related harm may still qualify when linked to discrimination or denial of protection under existing grounds.[17]

The Convention's object and purpose are protective. It seeks to safeguard individuals whose own state fails to secure their life or freedom. Where climate change exacerbates existing inequalities or state neglect, the analytical focus shifts from the environmental phenomenon to the nature of state protection. This distinction is central to identifying latent possibilities within the Convention framework.

Expansion Potential to Particular Social Group and Climate-Related Vulnerability

Among the five Convention grounds, membership of a particular social group (PSG) provides the greatest interpretive flexibility. Its content has progressively developed through judicial interpretation and guidance from the United Nations High Commissioner for Refugees.[18] A PSG is generally defined as a group sharing an immutable characteristic or a trait that is fundamental to identity or human dignity, with social perception often serving as an additional criterion. This interpretive flexibility becomes particularly relevant where climate-related vulnerability intersects with pre-existing patterns of discrimination, marginalization, or identity-based exclusion. Within this framework, climate-induced displacement may intersect with PSG in specific

---

[14] Sarah Locke, 'Who Defines the Refugee?: Analyzing the Limitations and Relevance of the 1951 UN Convention on Global Refugee Protection' (2025) 28.

Ramy Mohamed Moustafa, 'Recognizing Grave Socio-Economic Rights Abuses as Valid Refugee Claims under International Refugee Law' (2019) 25.

Heather Alexander and Jonathan Simon, 'Unable to Return in the 1951 Refugee Convention: Stateless Refugees and Climate Change' (2014) 26 *Florida Journal of International Law* 531, 536.

Ulrike Krause, 'Colonial Roots of the 1951 Refugee Convention and Its Effects on the Global Refugee Regime' (2021) 24(3) *Journal of International Relations and Development* 599, 602.

Christel Querton, 'The Interpretation of the Convention Ground of "Membership of a Particular Social Group" in the Context of Gender-related Claims for Asylum' (2012) 5.

contexts, particularly where vulnerability is closely tied to identity. Indigenous communities illustrate this dynamic: their deep cultural, spiritual, and economic connection to ancestral land may constitute a fundamental characteristic, and where environmental degradation disproportionately affects them alongside discriminatory denial of state protection, the harm may transcend generalized environmental hardship and potentially satisfy the persecution requirement under the Convention.[19] In such cases, the legal inquiry does not center on the climate event itself but on whether the state withholds protection because of group identity. Systematic neglect in disaster relief, exclusion from adaptation policies, or denial of relocation assistance due to entrenched marginalization may cumulatively meet the threshold of persecution. Similarly, other groups may qualify as PSGs when they share a common, fundamental characteristic linked both to their exposure to environmental harm and to discriminatory treatment, such as communities defined by traditional subsistence practices or by geographic isolation compounded by social stigma. The decisive element remains the nexus requirement: environmental harm must be connected to a discriminatory denial of protection for a Convention reason.

This approach does not create a new "climate refugee" category but works within existing refugee law, where climate change heightens vulnerability without replacing the persecution requirement. While the threshold remains high, serious and sustained denial of protection linked to a Convention ground may still allow limited protection, particularly through the PSG category.

The Principle of Non-Refoulement under Article 33

Article 33(1) of the 1951 Convention establishes the principle of non-refoulement. It prohibits States from expelling or returning a refugee to territories where their life or freedom would be threatened on account of a Convention ground. This principle constitutes the cornerstone of international refugee protection.[20]

Non-refoulement applies once an individual qualifies as a refugee under Article 1A(2). It does not extend to all migrants affected by environmental harm. The protection remains conditional on meeting the definitional criteria. Nevertheless, Article 33

[19] Scott Graham, 'Staging the Performances of the Privileged Social Group (PSG): Expanding the Philosophical Foundation of Critical Pedagogy' (2008) 6(2) *Policy Futures in Education* 263, 273.

Ellen F D'Angelo, 'Non-refoulement: The Search for a Consistent Interpretation of Article 33' (2009) 42 *Vanderbilt Journal of Transnational Law* 279, 279.

performs two important functions in the climate context. First, it reinforces the centrality of protection against life-threatening harm. The Convention recognizes that return is impermissible where fundamental interests are at risk. This logic aligns with broader human rights principles.[21]

In addition, non-refoulement operates as a minimum guarantee within refugee law. Even if broader socio-economic rights remain limited, the prohibition on return sets a clear protective floor. If climate-related harm, combined with persecution for a Convention reason, places an individual at risk, Article 33 prevents removal.[22] Non-refoulement also extends beyond formally recognized refugees through international human rights law, reinforcing that international law is not silent on climate- induced harm. Nevertheless, refugee law alone cannot comprehensively protect climate-displaced persons, as protection remains tied to persecution and nexus requirements. Still, through PSG analysis and assessments of state protection failure, the Convention retains limited interpretive space for certain climate-related claims.[23] In conclusion, the 1951 Convention does not recognize climate change as an independent ground and offers no universal solution. Nevertheless, its core principles, particularly PSG and non-refoulement, still provide limited but meaningful avenues for protection within an emerging minimum standard in international law.

INTERNATIONAL HUMAN RIGHTS LAW AS THE STRONGEST DOCTRINAL BASIS

Article 6 ICCPR - The Right to Life as a Right to Dignified Existence

Article 6(1) ICCPR affirms that every human being has the inherent right to life. Contemporary interpretation by the UN Human Rights Committee, especially in General Comment No. 36 (2018), has significantly expanded the normative scope of this provision. The right to life is no longer confined to protection against arbitrary killing by State agents. Rather, it encompasses the entitlement to conditions necessary for a

[21] Neil C Renic, *Asymmetric Killing: Risk Avoidance, Just War, and the Warrior Ethos* (Oxford University Press 2020).

Chiara Scissa, 'The Principle of Non-Refoulement and Environmental Migration: A Legal Analysis of Regional Protection Instruments' (2022) 3 *Diritto, Immigrazione e Cittadinanza* 9.

Geoff Gilbert and Tamara Wood, 'International Protection for People Displaced Across Borders in the Context of Climate Change and Disasters: A Practical Toolkit' (EJIL:Talk!, 4 March 2025) https://www.ejiltalk.o rg/int ernati onal - protec tion-f or- people - displaced- across- bor ders-i n-the -cont ext-of - climate-change-and-disasters-a-practical-toolkit/ accessed 28 February 2026.

life with dignity and imposes positive obligations upon States to take reasonable measures to prevent foreseeable threats to life.[24]

Importantly, these threats are not limited to acts of violence. The Human Rights Committee has explicitly recognized that environmental degradation, climate change, and natural disasters may jeopardize the effective enjoyment of the right to life. Where climate-induced phenomena, such as rising sea levels, extreme heat, prolonged drought, or ecosystem collapse, create life-threatening conditions, States must take preventive and protective measures.[25]

Under Human Rights Committee jurisprudence, States breach Article 6 if they return individuals to a real and foreseeable risk to life, based on a forward-looking and individualized assessment. This includes not only direct harm but also severe environmental conditions, meaning climate-induced uninhabitability can trigger State responsibility. This position was affirmed in *Teitiota v New Zealand*, where the Human Rights Committee recognised that environmental degradation and sea-level rise may expose individuals to life-threatening conditions capable of engaging non-refoulement obligations under Article 6 of the ICCPR.

Accordingly, the right to life functions as a normative anchor for a minimum standard of protection. While it does not establish a new refugee category, it prohibits removal where climate-related conditions pose a direct threat to survival or a dignified existence.

Teitiota v. New Zealand (2020)

The Human Rights Committee's decision in *Teitiota v. New Zealand* (2020) represents a landmark in the evolution of climate-related protection under international human rights law. The applicant, a national of Kiribati, argued that his removal from New Zealand would expose him to life-threatening conditions caused by sea-level rise, overcrowding, and environmental degradation.[26]

Although the Committee rejected the communication on evidentiary grounds, finding the risk insufficiently imminent under Article 6, it affirmed that climate change can

[24] Ginevra Le Moli, 'THE HUMAN RIGHTS COMMITTEE, ENVIRONMENTAL PROTECTION AND THE RIGHT TO LIFE' (2020) 69 International and Comparative Law Quarterly 735, 740-742.
Aristi Volou, 'The Protection of Socio-Economic Rights through the Canon of Civil and Political Rights: A Comparative Perspective' (2017) 5 Groningen Journal of International Law 147, 163-165.

[26] *TeitiotavNewZealand [2019]HumanRightsCommitteeCommunicationNo2728/2016, CCPR/C/127/D/2728/2016 [2.1-2.10]*.

trigger non-refoulement obligations under the ICCPR. It recognized that environmental degradation and sea-level rise may create life-threatening conditions, making return a potential violation of Article 6.[27]

Teitiota is significant because it confirms that climate-related harm can, at least in principle, trigger non-refoulement obligations under the ICCPR. Although the communication ultimately failed due to insufficient evidence of an imminent risk to life, the Committee nevertheless acknowledged that environmental degradation and climate-related harm may, in certain circumstances, render return incompatible with Article 6 obligations. In this way, Teitiota supports the argument that international law already contains the normative foundations of a minimum standard of protection for climate-displaced persons, even in the absence of formal recognition under the 1951 Refugee Convention.[28]

Article 7 ICCPR as a Prohibition of Ill-Treatment in Slow-Onset Climate Scenarios

Complementing Article 6, Article 7 ICCPR prohibits torture and cruel, inhuman, or degrading treatment or punishment. Unlike Article 6, Article 7 is absolute and non-derogable. It does not require proof of intent to harm, nor does it require a risk of death. The focus instead lies on the severity of suffering and the preservation of human dignity.

This provision is particularly relevant in slow-onset climate scenarios. Gradual processes such as desertification, salinization of agricultural land, prolonged drought, and depletion of freshwater resources may not immediately threaten life. However, they can gradually erode the basic conditions necessary for subsistence. Where individuals are returned to environments in which access to food, potable water, shelter, or healthcare has collapsed to a degree incompatible with human dignity, such return may amount to exposure to inhuman or degrading treatment.[29]

The Human Rights Committee has affirmed that States must not remove individuals where substantial grounds indicate a risk of treatment contrary to Article 7. While the evidentiary standard is similar to Article 6, the focus is on severe suffering or degrading

[27] ibid 9.4-9.6.
[28] Le Moli (n 24) 737.
[29] Shirley Llain Arenilla and Cindy Hawkins Rada, 'Climate Change and Forced Migration' (2020) 11 Migraciones internacionales 1, 3-4.

conditions rather than risk of death. In some contexts, Article 7 offers broader protection by addressing extreme deprivation and vulnerability beyond threats to life. Nevertheless, doctrinal caution is necessary. Not every instance of socio-economic hardship amounts to ill-treatment within the meaning of Article 7. The boundary between ordinary poverty and treatment incompatible with human dignity remains context-dependent. However, where climate change renders living conditions fundamentally unsustainable, Article 7 operates as a complementary safeguard reinforcing the right to life.

A distinction should also be drawn between sudden-onset disasters and slow-onset environmental degradation. In cases of sudden disasters, the immediacy of harm may make it easier to demonstrate a foreseeable threat to life or physical integrity. By contrast, slow-onset processes such as desertification, salinization, or rising sea levels often involve cumulative and incremental harm, making it more difficult to satisfy evidentiary thresholds relating to imminence and individualized risk. This distinction illustrates one of the central practical limitations of applying existing non-refoulement frameworks to climate-induced displacement.

Together, Articles 6 and 7 of the ICCPR provide a strong basis for climate-related protection, establishing non-refoulement grounded in human dignity and showing that international law offers an adaptable minimum standard despite limits.

COMPLEMENTARY PROTECTION AND SOFT LAW

Complementary Protection in Regional Humanitarian Practices

Complementary protection refers to forms of international protection granted to individuals who do not meet the refugee definition but who nonetheless face serious harm upon return. Its legal basis varies across jurisdictions. In regional systems, especially within the European Union, complementary protection has been codified and institutionalized.

A key instrument is the Directive 2011/95/EU. This Directive establishes subsidiary protection for persons facing a real risk of serious harm, including the death penalty, torture, or serious and individual threats to life resulting from indiscriminate violence in situations of armed conflict. While environmental harm is not expressly listed, the

concept of “serious harm” has been interpreted in conjunction with human rights obligations.[30]

Across several European jurisdictions, complementary protection reflects a gradual shift from a persecution-based refugee model toward a human rights-based approach grounded in human dignity, focusing on whether return would expose individuals to a real risk of serious harm.[31] This approach can be seen in Sweden. Swedish law allows subsidiary protection where a person faces torture, inhuman or degrading treatment, or other serious harm. In cases concerning Somalia during famine and state collapse, Swedish courts examined whether general humanitarian conditions reached the threshold of serious harm. Applicants did not need to prove individualized persecution. The courts assessed whether return would expose them to intolerable living conditions in objective terms.[32]

A similar reasoning appears in Germany. Sections 60(5) and 60(7) of the Residence Act incorporate Article 3 of the European Convention on Human Rights and the principle of non-refoulement. The Federal Administrative Court has held that removal is unlawful if there is a real risk of serious harm to life or physical integrity. This applies even when the risk results from general conditions such as infrastructural collapse or extreme humanitarian crisis. The focus is on foreseeable consequences of return, not on identifying a persecutor.[33]

Comparable approaches can also be observed in Italy,[34] France,[35] and the Netherlands[36]: where courts and administrative bodies increasingly assess return decisions through the lens of human dignity and objective exposure to serious harm. In these jurisdictions, the analysis focuses less on identifying intentional persecution

[30] European Union Agency for Asylum, *Article 15(c) Qualification Directive 2011/95/EU: A Judicial Analysis* (EUAA) https://www.euaa.europa.eu/sites/default/files/public/Article-15c-Qualification-Directive-201195EU-A-judicial-analysis.pdf accessed 28 February 2026.

[31] Megan Bradley and Roberta Cohen, ‘Disasters and Displacement: Gaps in Protection’ (2010) 1(1) *Journal of International Humanitarian Legal Studies* 1, 7.

[33] F.G. v Sweden App no 43611/11 (ECtHR, 23 March 2016).
‘Germany - Federal Administrative Court, 20 August 2018, 1 B 18.18’ (Asylum Law Database) https://www.asylumlawdatabase.eu/en/case-law/germany-%E2%80%93-federal-administrative-court-20-august-2018-1-b-1818 accessed 28 February 2026.

[34] European Council on Refugees and Exiles, ‘Italy: Latest Immigration Decree Drops Protection Standards’ (ECRE, 28 September 2018) https://ecre.org/italy-latest-immigration-decree-drops-protection-standards/ accessed 28 February 2026.

[35] Office français de protection des réfugiés et apatrides, ‘Exclusion and Decision Not to Grant Refugee Status’ (OFPRA) https://www.ofpra.gouv.fr/en/exclusion-and-decision-not-grant-refugee-status accessed 28 February 2026.

[36] Elgafaji v Staatssecretaris van Justitie (Opinion of AG Maduro) Case C-465/07, EU:C:2008:650.

and more on whether prevailing humanitarian or environmental conditions would subject individuals to degrading treatment contrary to Article 3 ECHR or equivalent human rights protections.

Across these jurisdictions, the definition of refugee under the 1951 Convention remains unchanged. No European State formally recognizes climate refugees. Protection is extended through human rights mechanisms.[37] The legal question is no longer who persecutes the applicant. It is whether the return would expose the applicant to serious harm that violates fundamental human rights standards. Complementary protection extends protection beyond the narrow refugee definition without creating a new category of "climate refugee." Although general poverty or environmental degradation alone is insufficient, States may prevent return where environmental conditions pose a real risk of serious harm or threaten basic human dignity. While these protections remain limited, fact-specific, and inconsistent across jurisdictions, they demonstrate an emerging willingness to protect individuals beyond the 1951 refugee framework.

Soft Law Mechanisms and Normative Influence

Although soft law instruments do not create binding obligations, they influence legal interpretation, policy development, and the gradual consolidation of international norms. The Global Compact for Safe, Orderly and Regular Migration recognizes climate change, environmental degradation, and disasters as drivers of migration, encourages greater international cooperation, and calls for enhanced protection pathways for affected populations.[38] While non-binding, the Compact reflects an emerging international consensus that climate-related mobility constitutes a significant humanitarian and governance concern rather than a marginal issue.[39] Similarly, the UN Guiding Principles on Internal Displacement establish standards on non- discrimination, necessities, and protection against arbitrary displacement in disaster

[37] Sumudu Atapattu, 'Climate Change and Displacement: Protecting "Climate Refugees" within a Framework of Justice and Human Rights' (2020) 11(1) *Journal of Human Rights and the Environment* 86, 92.

[38] UN General Assembly, *Global Compact for Safe, Orderly and Regular Migration*, UN Doc A/RES/73/195 (19 December 2018) paras 18(h), 21.

Office of the United Nations High Commissioner for Human Rights, 'Global Compact for Safe, Orderly and Regular Migration (GCM)' (OHCHR) https://www.ohchr.org/en/migration/global-compact-safe-orderly-and-regular-migration-gcm accessed 28 February 2026.

contexts, reinforcing the human rights dimension of climate- and disaster-related displacement.[40]

Although soft law instruments such as the Global Compact for Safe, Orderly and Regular Migration remain non-binding and unevenly implemented, they still influence judicial reasoning and administrative practice by shaping interpretations of existing legal obligations. While they cannot independently create rights to admission or asylum, they contribute to an evolving understanding that climate-related harm may generate protection needs grounded in human dignity and vulnerability.

The Climate Regime and Indirect Protective Implications

The international climate regime does not regulate refugee status or grant cross-border protection to climate-displaced persons. Nevertheless, instruments such as the Paris Agreement and the Warsaw International Mechanism for Loss and Damage recognize climate vulnerability and displacement as matters of international concern rather than mere misfortune.[41] By emphasizing adaptation, cooperation, and human rights considerations, these frameworks indirectly reinforce expectations that States should prevent foreseeable climate-related harm.

Although the protective implications remain indirect and fragmented, the interaction between climate law, complementary protection, and human rights law increasingly supports a broader protective logic grounded in human dignity and non-refoulement. While no unified legal framework currently exists, these overlapping regimes suggest an emerging minimum standard of protection for climate-displaced persons.

CRITICAL EVALUATION AND CONCLUSION

International law still lacks a comprehensive framework for people displaced across borders by climate change. Nevertheless, existing legal regimes already provide important forms of protection. Refugee law offers only limited pathways because of its

[40] United Nations High Commissioner for Refugees, *Guidelines on International Protection No. 12: Claimsfor Refugee Status Relatedto SituationsofArmed Conflict and Violence underArticle 1A(2) of the1951 Convention and/or the1967 Protocolrelatingto theStatus of Refugees* (UNHCR, 2 December 2016) https://www.unhcr.org/sites/default/files/legacy-pdf/47949b3f2.pdf accessed 28 February 2026.

[41] Paris Agreement, adopted 12 December 2015, entered into force 4 November 2016, UNTS No 54113, preamble; Decision 2/CP.19, *Warsaw International Mechanism for Loss and Damage associated with Climate Change Impacts* (2013).

persecution-based structure; however, international human rights law, particularly through non-refoulement obligations under Articles 6 and 7 of the ICCPR, increasingly fills part of this gap. Complementary protection and soft law mechanisms further reinforce this trend. Current protection operates through a combination of refugee law, human rights law, and complementary protection mechanisms rather than through a single unified regime. As climate displacement intensifies, the challenge for international law will be less about creating protection from nothing and more about developing greater consistency across existing norms.

Bibliography

Annick Pijnenburg and Conny Rijken, 'Moving beyond Refugees and Migrants: Reconceptualising the Rights of People on the Move' (2021) 23 Interventions 273, 276 <https://doi.org/10.1080/1369801X.2020.1854107>.

Aristi Volou, 'The Protection of Socio-Economic Rights through the Canon of Civil and Political Rights: A Comparative Perspective' (2017) 5 Groningen Journal of International Law 147, 163-165.

Chiara Scissa, 'The Principle of Non-Refoulement and Environmental Migration: A Legal Analysis of Regional Protection Instruments' (2022) 3 *Diritto, Immigrazione e Cittadinanza* 9.

Christel Querton, 'The Interpretation of the Convention Ground of "Membership of a Particular Social Group" in the Context of Gender-related Claims for Asylum' (2012) 5.

Convention Relating to the Status of Refugees (adopted 28 July 1951, entered into force 22 April 1954) 189 UNTS 137 art 1A(2).

Elgafaji v Staatssecretaris van Justitie (Opinion of AG Maduro) Case C-465/07, EU:C:2008:650.

Ellen F D'Angelo, 'Non-refoulement: The Search for a Consistent Interpretation of Article 33' (2009) 42 *Vanderbilt Journal of Transnational Law* 279, 279.

European Council on Refugees and Exiles, 'Italy: Latest Immigration Decree Drops Protection Standards' (ECRE, 28 September 2018) https://ecre.org/italy-latest-immigration-decree-drops-protection-standards/ accessed 28 February 2026.

European Union Agency for Asylum, *Article 15(c) Qualification Directive 2011/95/EU: A Judicial Analysis* (EUAA) https://www.euaa.europa.eu/sites/default/files/public/Article-15c-Qualification- Directive-201195EU-A-judicial-analysis.pdf accessed 28 February 2026.

F.G. v Sweden App no 43611/11 (ECtHR, 23 March 2016).

Geoff Gilbert and Tamara Wood, 'International Protection for People Displaced Across Borders in the Context of Climate Change and Disasters: A Practical Toolkit' (EJIL:Talk!, 4 March 2025) https://www.ejiltalk.org/international-protection-for-

people-displaced-across-borders-in-the-context-of-climate-change-and-disasters-a-practical-toolkit/ accessed 28 February 2026.

'Germany - Federal Administrative Court, 20 August 2018, 1 B 18.18' (Asylum Law Database) https://www.asylumlawdatabase.eu/en/case-law/germany-%E2%80%93-federal-administrative-court-20-august-2018-1-b-1818 accessed 28 February 2026.

Ginevra Le Moli, 'THE HUMAN RIGHTS COMMITTEE, ENVIRONMENTAL PROTECTION AND THE RIGHT TO LIFE' (2020) 69 International and Comparative Law Quarterly 735, 740-742.

Heather Alexander and Jonathan Simon, 'Unable to Return in the 1951 Refugee Convention: Stateless Refugees and Climate Change' (2014) 26 *Florida Journal of International Law* 531, 536.

Hélène Lambert, 'Refugee Status, Arbitrary Deprivation of Nationality, and Statelessness within the Context of Article 1A(2) of the 1951 Convention and its 1967 Protocol relating to the Status of Refugees' (UNHCR, Legal and Protection Policy Research Series, Division of International Protection, Geneva 2014) 16.

Martins Paparinskis, 'International Minimum Standard and the Protection of Property' in Martins Paparinskis, *The International Minimum Standard and Fair and Equitable Treatment* (1st ed, Oxford University PressOxford 2013) 271-259 <https://academic.oup.com/book/11589/chapter/160455070> accessed 21 February 2026.

Maryellen Fullerton, 'A Comparative Look at Refugee Status Based on Persecution Due to Membership in a Particular Social Group' (1993) 26 *Cornell International Law Journal* 505, 511.

Megan Bradley and Roberta Cohen, 'Disasters and Displacement: Gaps in Protection' (2010) 1(1) *Journal of International Humanitarian Legal Studies* 1, 7.

Neil C Renic, *Asymmetric Killing: Risk Avoidance, Just War, and the Warrior Ethos* (Oxford University Press 2020).

Nischala McDonnell, 'Falling Through the Protection Gaps: Inappropriate Protection of Climate Displaced Persons in the International Refugee Legal Structure' (2024) 10 The Transnational Human Rights Review <https://doi.org/10.60082/2563-4631.1103> accessed 16 May 2026.

Office français de protection des réfugiés et apatrides, 'Exclusion and Decision Not to Grant Refugee Status' (OFPRA) https://www.ofpra.gouv.fr/en/exclusion-and- decision-not-grant-refugee-status accessed 28 February 2026.

Office of the United Nations High Commissioner for Human Rights, 'Global Compact for Safe, Orderly and Regular Migration (GCM)' (OHCHR) https://www.ohchr.org/en/migration/global-compact-safe-orderly-and-regular-migration-gcm accessed 28 February 2026.

Paris Agreement, adopted 12 December 2015, entered into force 4 November 2016, UNTS No 54113, preamble; Decision 2/CP.19, *Warsaw International Mechanism for Loss and Damage associated with Climate Change Impacts* (2013).

Pierre Thielbörger, 'The "Essence" of International Human Rights' (2019) 20 German Law Journal 924, 930-933.

Ramy Mohamed Moustafa, 'Recognizing Grave Socio-Economic Rights Abuses as Valid Refugee Claims under International Refugee Law' (2019) 25.

Sarah Locke, 'Who Defines the Refugee?: Analyzing the Limitations and Relevance of the 1951 UN Convention on Global Refugee Protection' (2025) 28.

Scott Graham, 'Staging the Performances of the Privileged Social Group (PSG): Expanding the Philosophical Foundation of Critical Pedagogy' (2008) 6(2) *Policy Futures in Education* 263, 273.

Shirley Llain Arenilla and Cindy Hawkins Rada, 'Climate Change and Forced Migration' (2020) 11 Migraciones internacionales 1, 3-4.

Sumudu Atapattu, 'Climate Change and Displacement: Protecting "Climate

Refugees" within a Framework of Justice and Human Rights' (2020) 11(1) *Journal of Human Rights andthe Environment* 86, 92.

Teitiota v New Zealand [2019] Human Rights Committee Communication No 2728/2016, CCPR/C/127/D/2728/2016 [2.1-2.10].

Ulrike Krause, 'Colonial Roots of the 1951 Refugee Convention and Its Effects on the Global Refugee Regime' (2021) 24(3) *Journal of International Relations and Development* 599, 602.

UN General Assembly, *GlobalCompactforSafe,OrderlyandRegularMigration* , UN Doc A/RES/73/195 (19 December 2018) paras 18(h), 21.

United Nations High Commissioner for Refugees, 'Climate Change and Displacement' (UNHCR) https://www.unhcr.org/what-we-do/build-better-futures/climate-change-and-displacement accessed 28 February 2026.

United Nations High Commissioner for Refugees, *GuidelinesonInternational Protection No. 12: Claims for Refugee Status Related to Situations of Armed Conflict and Violence under Article 1A(2) of the 1951 Convention and/or the 1967 Protocol relatingtothe Status ofRefugees* (UNHCR, 2 December 2016) https://www.unhcr.org/sites/default/files/legacy-pdf/47949b3f2.pdf accessed 28 February 2026.

Authors Bio-

Thi Ngoc Mai Nguyen is a graduate of the University of Bristol Law School, where she earned a Master's degree in International Law. Her research interests focus on developing national policies that comply with international legal obligations and enhance their effectiveness, addressing transnational crime through sustainable institutional frameworks, and advancing the implementation of legal obligations related to human rights protection and climate change mitigation. Beyond her academic pursuits, she is also a classical pianist.

Authors Bio-

Le Bao Ngan Pham holds an M.Sc. in International Business Law and Arbitration from the University of Silesia in Katowice (Poland). Her research interests include international economic law, environmental governance, ESG regulation, human rights, AI governance, digital governance, and comparative law. Her recent research examines the interaction between sustainability regulation, emerging technologies, cross-border data governance, and digital sovereignty, particularly from the perspective of developing countries.

The Law of State Responsibility is increasingly irrelevant; it cannot respond to the increasing power and influence of Non-State Actors.

Author: SosefoSepasitiano Patelisio

ORCID: 0009-0002-7811-3139

Programme: LLMInternational Law & International Relations

Affiliation: Law School

Email: sepapatelisio@gmail.com

Abstract

The International Law Commission's Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA) remain the core framework for determining when States can be held accountable for breaches of international law. Grounded in
the principle of State primacy, ARSIWA centres on attribution as the mechanism for linking conduct to a State. However, the growing influence and participation of non- State actors (NSAs), involving multinational corporations and armed groups, has revealed significant constraints in this framework. In many situations, States rely significantly on NSAs to pursue national objectives and evade international responsibilities, thus, revealing gaps in ARSIWA's attribution rules, particularly under Article 4-11. These provisions have been criticized as inadequate for addressing indirect state involvement or collaboration with armed groups. This article argues that traditional state responsibility is becoming less effective in addressing NSA-related misconduct. It examines key attribution principles, analyses Articles 4,6,7,8, 9, 10 and
11, and offers recommendations to strengthen accountability in this evolving landscape.

Keywords: Articles on Responsibility of States for International Wrongful Acts (ARSIWA), State Attribution, State Responsibility, Armed groups, Non-State Actors (NSA).

To my loving parents, Kosilio & Patinia Patelisio, for their unwavering love, prayers, and support.
To my siblings, Patelisio; Polini & Mike Nansen; Fapiano; and Neomai, for their encouragement andconstant support. To Fr. Visone Tangifua, Fr. Sione Panisi Vaohea, and Sr. Malia Simione, for their spiritualguidance and prayers. To Simaima Patelisio; Polini & Suliasi Katoa, Pateli & Susana; Moana & Weather Mafi;Katalina Pateli; 'Alekisio Na'a; and 'Emeliano & Senorita Pateli. To Teiko & Vei Mataele, Lea & VeronicaMataele, Save & Muli Mataele, Vili & Seini Mataele, Pinomi & Lavinia Mataele and Sione Mataele. To SialeLangi, Melefepaki Langi, Akata & Etuate Manu, Nia & Tangi Sikei, Kolio Taliauli, Tevita Faleofa, Malia&Kuli Fehoko, and Soane Hamala.Thank you for all the support, love and care —words cannot fully expressmy gratitude. And finally to the Kolonga Parish community, for their prayers, encouragement, and support throughout this journey.

The law of state responsibility, entrenched in the International Law Commission's (ILC) *Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA),* has long been established as a primary tool for holding states accountable in international law. Entrenched in the principle of State primacy, this framework establishes when and how a state can be held responsible for violations of its international obligations. However, with the growing influence and emergence of non-State actors (NSAs) — such as multi-corporations and armed groups— their utilization by states have exposed gaps in ARSIWA's applicability in holding states accountable for indirectly supporting or collaborating with NSAs. ARSIWA's reliance on crucial concepts such as attribution particularly under Articles 5,7,8,9,10 and 11 has often been criticized as inadequate in holding states accountable for working with and through NSAs to violate international law, thereby evading attribution that would be deemed unlawful under international law. This article argues that the law of State responsibility is becoming increasingly irrelevant in responding to the increasing power and influence of NSA to states. The focus of this article will be on these attribution challenges with a specific focus on armed groups. It will first provide an overview of the general principles of state attribution. Then it will examine Articles 4, 6 and 7 of ARSIWA followed by an analysis of Article 8 and finally Articles 9,10 and 11. Finally, the essay will provide recommendations aimed at reinforcing the legitimacy and effectiveness of state responsibility in the context of NSAs.

General Principles of State Attribution

According to Kelsen, a state is a corporate entity and thus, cannot carry out its functions and tasks without the intermediate involvement of others.[1] The State, being a moral person without any physical existence, can act only through persons whose conduct is performed on behalf of the State. Such persons are called organs or agents, according to their precise link to the State.[2] The acts or omissions of individuals acting on behalf of the State are, from a legal perspective, treated as acts of the State itself. This does not involve representation or mandate. The conduct is not regarded as that of the individual and then attributed to the State; it is solely and directly an act of the

[1] Hans Kelsen, *ThePureTheory of Law* (2nd edn, Cambridge University Press 2014) 290–312.
[2] R. Ago, Third Report, § 107.

State.[3] As the nineteenth-century German private law scholars observed in relation to other moral persons, the State acts through its organs as they were its own "hands and feet". The concept of attribution therefore concerns the legal process by which the conduct of individuals or groups is treated as the conduct of the State for the purposes of international responsibility. In other words, international law determines whether acts or omissions carried out on behalf of the State may legally be regarded as "acts of the State."[4]

Without attribution, it would be impossible to link wrongful actions of an individual or group to the state, leaving violations of international law unaccounted for. The search for attribution is the search for a relationship or connection between the state, as a legal person, and the person or group carrying out the conduct.[5] For, as the court observed in the *German Settlers in Poland Advisory Opinion*, 'states can act only by and through their agents and representatives'[6]

The ARSIWA is central to the concept of attribution as it provides the framework for determining when states can be held responsible for acts that violate international obligations. As provided by ARSIWA Article 2, attribution is a requisite for determining state responsibility for internationally wrongful acts (IWAs).[7] It is foundational in international law that state responsibility requires the attribution of conduct to the state.[8]

---

[3] Robert Kolb, *The International Law of State Responsibility: An Introduction* (Edward Elgar Publishing 2017).

[4] Onattribution, see the general developments in: Crawford, State, pp.113ff; Crawford, ILC's, pp. 91ff; Crawford, Pellet and Olleson, pp. 221ff; R. Ago, Third Report, §§ 106ff and Fourth Report, §§ 1ff. For earlier contributions, see e.g. Anzilotti, Teoria, pp. 121ff; Schoen, Völkerrechtliche, pp. 43ff. For a more developed treatment, see L. Condorelli, 'L'imputation à l'Etat d'un fait internationalement illicite', RCADI, vol. 189, 1984-VI, pp. 9ff. Some authors argue for a dropping of the notion of attribution on the grounds that it is inherent in the notion of duty and varies according to the primary obligation: I. Brownlie, 'International Law at the Fiftieth Anniversary of the United Nations', RCADI, vol. 255, 1995, pp. 93–94.

[5] Meg Kinnear, '*ARSIWAM ISDS, AND THE PROCESS OF DEVELOPING AN INVESTOR-STATE JURISPRUDENCE'* (Cambridge University Press, 2021) < https://www.cambridge.org/core/services/aop-cambridgecore/content/view/2C5A3E1EBC1BC50E2301E52D0530C887/S2633900521000473a.pdf/arsiwa_isds_and_the_process_of_developing_an_investorstate_jurisprudence.pdf> accessed 25 November 2024.

[6] German Settlers in Poland 'Advisory Opinion' (*International Court of Justice*, 23 November 2020) < https://jusmundi.com/en/document/decision/en-german-settlers-in-poland-advisory-opinion-monday-10th-september-1923> accessed 27 November 2024.

[7] Carlo Stefano, 'The Meaning of Attribution' (*Oxford University Press*, 2020) < https://watermark.silverchair.com/oso-9780198844648-chapter-2.pdf> accessed 25 November 2024.

[8] *German Settlers in Poland* (n 3)

Articles of Attribution in ARSIWA

Article 5: Attribution of acts of private parties

Article 5 of ARSIWA is a provision that provides that some acts or omissions by persons not being part of the organic structure of the State and not empowered by ordinary formal laws to represent the State, but in fact exercising State functions, are attributed to the State.[9]

When an entity is not an organ of the state (entity that performs legislative, executive or judicial responsibilities), the latter may still be held responsible for its conduct in certain circumstances.[10] In essence, the conduct is only attributable to the state when the entity is acting within the scope of the governmental authority it has been empowered to exercise. Acts that go beyond this scope is not attributable to the state. The Draft Articles clarified this,

> *'The fact that an entity can be classified as public or private..or is subject to State control to a greater or lesser extent..do not emerge as a decisive criteria for the purposes of attribution to the State. ..[T]he most appropriate solution is that is these entities are empowered to exercise specific functions which are akin to those normally exercised by the organs of the State'*[11]

The case of *Hyatt International Corporation v Iran*[12] illustrates this. There, the Tribunal held Iran liable for acts of expropriation carried out by a non-state charity, the Foundation of the Oppressed (FO). The Tribunal found that the FO was instrumentally controlled by the Government of the Islamic Republic of Iran and thus was liable for the actions of FO through attribution.

However, in circumstances, where any entity is not empowered by the law of the state to exercise governmental functions, the State cannot be held responsible for their

---

[9] ARSIWA Article 5 provides: The conduct of a person or entity which is not an organ of the State under article 4 but which is empowered by the law of that State to exercise elements of the governmental authority shall be consideredan act of the State under international law, provided the person or entity is acting in that capacity in the particular instance.

[10] ARSIWA Article5 provides: The conduct of a personor entity which isnot an organ of the Stateunderarticle 4 but which is empowered by the law of that State to exercise elements of the governmental authority shall beconsidered an act of the State under international law, provided the person or entity is acting in that capacity in the particularinstance.

[11] Draft Articles Commentary, Art. 7, 18.

[12] (1985) 9 Iran–US CTR 72, 94.

actions. In *EDF (Services) Limited v Romania*.[13], the Tribunal did not consider the commercial activities of two state-owned corporations, CN Bucarest Aeroporti Otopeni (AIBO) and Transportationuri Aerience Romane Tarom SA (TAROM) attributable to Romania. The Tribunal found that both corporations were not exercising delegated authority. Rather, they were pursuing private commercial activities of making profits and were not specific acts alleged by the Claimant.

One of the important features of article 5 is that conduct is attributable to a state if the entity is empowered by the State's domestic law. In this way, it ties and binds the entity's conduct to state authorization and hold states accountable. However, the concept of governmental authority is ambiguous, broad and lacks precise definition, sparking several questions. What constitutes governmental authority? What evidentiary burden does the plaintiff have to prove the accused was exercising governmental authority? How can the plaintiff determine that entity acted in that capacity at the wrongful time, especially when entities perform both public and private functions? It has been argued, including in *Noble Ventures, Inc. v. Romania* and by Boon, that there is no clear or consistent understanding in international law of what constitutes 'governmental authority'. [14] Crawford agrees and argues its lack of precision leads to several ambiguities as it can apply to 'public corporations, public agencies or in cases, private companies empowered by the State's domestic laws.'[15] Even if governmental authority is identified in the context of article 5, to exercise those powers invested by domestic law, in practice, this will be difficult to identify where exactly the boundaries lie.[16] Although the ARSIWA commentary provides some ease in identifying examples of functions associated with governmental authority —such as powers of detention and discipline, immigration control, quarantine and the identification of property for seizure —relying on historical governmental functions may fail to account for the evolving role of the state, as privatization and outsourcing

[13] ICSID Case No. ARB/05/13.
[14] Noble Ventures, Inc. v. Romania, ICSID Case No. ARB/01/11), 82.
[15] James Crawford, *First Report on State Responsibility* (International Law Commission, UN Doc A/CN.4/490, 24 April 1998) para 168. Kristen E. Boon 'Attribution in International Law' (*Cambridge University Press*, 2024) <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/14E496EF5304E7B788AA10779EF51675/9781009334679c2_25-46.pdf/attribution-in-international-law.pdf > accessed 28 November 2024.
[16] Ibid.

increasingly blur the lines between public and private functions.[17] Accordingly, Crawford argues that States could escape responsibility by merely indicating that these entities are not its de jure organs or agents.[18] Thus, although article 5 is a critical provision in ensuring state accountability when NSAs perform governmental functions, its ambiguity in defining 'governmental authority' requires refinements to maintain effectiveness in international law.

Article 7: Ultra vires and personal acts

Article 7 of ARSIWA addresses attribution of conduct by state organs acting *ultra vires* —beyond the scope of its authority.[19] Under this provision, the conduct of a state organ is attributed to the state even if it acts outside its given powers, as long as the organ is acting in an official capacity.

The case of *Gustave Caire v United Mexican States,*[20] demonstrated this when Jean Caire was killed by two Mexican officers. Although their actions were clearly *ultra vires* (as they engaged in attempted extortion, murder and had violated the orders of their superiors), their actions were attributed to Mexico because they were acting in their capacity as State officials, even if unlawfully and beyond their authority. The Commission held;

> *'the two officers, even they acted outside their competence…and even if their superiors countermanded an order, but involved the responsibility of the State, since they acted under cover of their status as officers and used means placed at their disposal on account of that status'*[21]

In domestic law, when an organ or entity exceeds its authority (ultra vires), the state may still be held responsible because the organ or entity is part of the state's institutional framework.[22] However, it is argued that Article 7 fails to extend this

---

[17] *Boon* (n 11) 13.

[18] James Crawford, 'State Responsibility' (Cambridge University Press, 2013) < https://www.cambridge.org/core/books/state-responsibility/EE846D9378B83A9DCC5E794FC086B07E > accessed 30 November 2024.

[19] Article 7 of ARSIWA provides, the conduct of an organ of a State or of a person or entity empowered to exercise elements of the governmental authority shall be considered an act of the State under international law if the organ, person, or entity acts in that capacity, even if it exceeds its authority or contravenes instructions.

[20] *Caire Claim Case*, France v Mexico [1929] 5 RIAA 516.

[21] Ibid.

[22] James Crawford,' Organs and entities exercising governmental authority' (*Cambridge University Press*, 2013) < https://www.cambridge.org/core/services/aop-cambridge.pdf> accessed 25 December 2024.

principle to private individuals, even when they are empowered by the state or operating with its implicit consent or knowledge. Crawford contends that when a private entity acts in a way that furthers the interests of the state but exceeds its legal mandate, Article 7 does not provide a sufficient mechanism to hold the state responsible for their actions.[23] In Crawford's view, this is because Article 7 delegates the attribution only when they are carried out by state organs or entities acting within their authorized scope.[24] If a private individual or entity exceeds its mandate, the state may be shielded from liability under article 7, even if the actions align with the state's objectives.

However, it would be misleading to claim that article 7 does not ensure state accountability. Article 7 does provide for this especially in cases when state agents act outside their authority or disobey orders. Nevertheless, difficulties arise where an individual acts in a private capacity, as state responsibility will not arise in such circumstances.

Like the *Caire* case, if the two officers were private entities rather than state officials acting under the guise of state authority, it is argued that Mexico would not be liable under Article 7. Private entities do not possess the inherent connection to the State required for attribution under Article 7. This highlights a critical limitation for attribution of Article 7, as it fails to address scenarios where private actors, potentially operating with implicit state support, committed wrongful acts without formal ties to the state. This gap leaves room for states to evade liability by delegating unlawful activities to ostensibly independent private actors, thereby undermining accountability in international law. Thus, the situation becomes more complex when the perpetrators are private individuals acting on behalf of the state, but without clear state direction or control. Thus, Article 7 fails to provide an adequate framework to determine whether the state should be accountable.

Article 8: Conduct directed or controlled by a State

Despite the limitations of Article 7, Article 8 provides a broader structure to hold states accountable. Article 8 stipulates that the conduct of a person or group of persons can be attributed to a state if they are acting on the instructions of the state, or under its

[23] Ibid.
[24] *Crawford* (n 18).

direction or control.[25] Article 8 posits two situations in which this is done: first, the private entity has been instructed by a state to perform the acts in questions; and, second, where the entity is under the direction or control of the state.

The former occurs in the context of private military operations hired by the state to engage in certain activities on its behalf under issued instructions.[26] Although the principle is clear, difficulties arises in its application. In *Application of the Convention on the Prevention and Punishment of the Crime of Genocide* (Bosnian Genocide)[27], the court addressed whether the actions of military groups operating within the territory of the Federal Republic of Yugoslavia (FRY) could be attributed to the state before its disintegration.[28] The lawsuit claimed that the murder of Bosnian Muslim men at Srebrenica by members of the Republika Srpska's official (RSO) and the Bosnian Serb Amry (VRS), should be attributed to FRY. The Court found that FRY was providing "considerable military and financial support" and that there were "close ties" between FRY and the RSO. Despite these ties, the ICJ ruled that while Bosnia and Herzegovina had established that FRY had supported the VRS and the Republika Srpska, and that Bosnian Serb Army's actions at Srebrenica constituted genocide, it had not proven that these acts could be attributed to the FRY under the effective control test.[29] The test required evidence of explicit instructions to commit the massacre as well as genocidal intent in order to meet the effective control standard.

The same approach was taken in *Military and Paramilitary Activities in and Against Nicaragua*. [30] There, it concerned whether the military activities conducted by the Contras during the Nicaraguan civil war could be attributed to the U.S. The Court held and found that although the U.S. responsible was responsible for training, financing and providing logistical support to the Contras[31], the U.S. could not be held responsible

---

[25] Art 8 ARSIWA.
[26] Lehnardt (2008), 148; Tonkin (2011), 114–17; Beaucillon, Fernandez and Raspail (2011), 406.
[27] *Bosnian Genocide*, 2007 I.C.J. 43, 208-09, 400-01.
[28] Ibid 64-65, 236-37.
[29] *Bosnian Genocide,* 2007, I.C.J para 413. "It has not proved that instructions were issued by the federal authorities in Belgrade, or by any other organ of the FRY, to commit the massacres, still less that any such instructions were given with the specific intent (dolus specialis) characterizing the crime of genocide . . . All indications are to the contrary: that the decision to kill the adult male population of the Muslim community in Srebrenica was taken by some members of the VRS Main Staff, but without instructions from or effective control by the FRY"
[30] *Military and Paramilitary Activities in and Against Nicaragua* [1984] ICJ Rep 392.
[31] Ibid 59, 61.

because the U.S. did not exercise a sufficient level of control over the contras such that any act performed by them could be attributed to the U.S. The Court held;

> '[T]here is no clear evidence of the U.S. having actually exercised such a degree of control in all its fields as to justify treating the contras as having acted on its behalf' [32]

Nicaragua had to prove the U.S. possessed *effective control* of the military operations in the course of which the alleged violations were committed.[33]

The effective control test was the first method developed by the courts to determine the attribution of NSAs to states under international law. In *Bosnia* and *Nicaragua,* the test was used but ultimately failed because there was no evidence that the state had direct and detailed control over the military activities. While the test demands clear evidence of a state's control over an activity, it also reduces the risk of wrongful attribution. This protects states from being unfairly held accountable for an activity they did not control. Importantly, it protects state's sovereignty and ensures equality. However, it is argued that the effective control test sets the bar too high, such that the attribution of terrorist acts to the state proves nigh impossible. Boon agrees and contends that the evidentiary burden on the plaintiff is merely impossible to prove.[34] Providing high evidence that the accused issued explicit instructions or commands over specific operations in which wrongful acts occurred, is close to impossible. Like *Nicaragua,* this rigidity often allows States to escape liability even when they provide substantial support to NSA, such as funding, training, or strategic guidance, as these actions alone do not meet the threshold for "effective control." The requirement to demonstrate a direct link between a State's involvement and specific wrongful acts makes it nearly impossible to hold States accountable in many cases, particularly when key evidence of issuing instructions and commands is classified or exclusively within the State's control. The reliance of such concealed information places an onerous burden on claimants, requiring them to rely on circumstantial evidence, patterns of behaviour or inference from indirect sources. Consequently, the test creates a legal gap, enabling States to use NSA as proxies to commit internationally wrongful acts without bearing responsibility. This has been a major critique of the

[32] *Nicaragua* 1984, 404.
[33] *Nicaragua*,(n 26).
[34] *Kristen* (n11).

effective control test, as it undermines the principles of accountability and justice in international law.

Furthermore, because the ICJ did not find effective control in either *Nicaragua* or *Bosnian Genocide,* it remains unclear exactly what set of facts would satisfy the effective control test given its high threshold. Cassese agrees and contends to exercise control involves wielding power or authority over a person, but what is the scope of this authority or control? To what degree of control should a state exercise to incur responsibility?[35] Theoretically, if a state were to use a non-state actor to carry out a targeted killing, this could be seen as an exercise of effective control over the NSA. However, the state's involvement would likely need to entail significant control over the military operation, likely exceeding the level of control exercised by the U.S. over the Contras or the FRY over the VRS. Thus, the ICJ's reasoning in *Nicaragua* and *Bosnian Genocide* could be interpreted to suggest that a state might arm, fund, support, train and facilitate the operations of an armed non-state group, and even encourage the group to carry out ethnic cleansing as a means of defeating the enemy, yet still avoid responsibility due to a lack of evidence that State agents directly instructed the commission of the specific massacre.

Overall Control Test

The second scenario in which a state will be held responsible for the acts of private entities is where the entity is under the direction or control of the state.

Despite the ongoing criticism about the rigid criteria of the control test, the courts developed a more flexible and pragmatic standard for attributing the actions of NSA to states. Unlike the effective control test, the overall control test broadens the scope by recognizing that a state can be held accountable if it exert influence over a NSA by providing resources or strategic direction, even without micromanaging every action. In *Prosecution v Duško Tadić,*[36]a Bosnian Serb politician was accused of committing grave breaches of international humanitarian law. The Appeals Chamber rejected the application of the *Nicaragua* effective control test to the facts of the case. It

[35] Antonio Cassese James Crawford,' The Nicaragua and Tadić Tests revisited in light of the ICJ Judgement on Genocide in Bosnia' (*Oxford University Press*, 2013) https://watermark.silverchair.com > accessed 25 December 2024.

[36] Tadid, Case No. IT-94-1-A, Appeals Chamber Judgment, 115.

emphasized that the purpose of Article 8 was to prevent States from escaping international responsibility by having private individuals carry out tasks that may not or should be performed by State officials.[37] As a result, it declared that for organized groups, it required that the group as a whole be under the overall control of the State.[38] Demonstrating that the state exercised a more general level of control over the non-state group is sufficient to attribute the group's action to the state. Thus, the Appeal Chamber held FRY had exercised overall control over the Bosnian Serb Army. once a NSA is determined to be under the overall control of a state, the state becomes responsible for all the actor's actions, including *ultra vires* acts committed by the group. Attribution under the overall control test is significantly straightforward than under the effective control test. The overall control test requires only that a state demonstrate a broad level of involvement such as arming, financing, equipping or training terrorist units. In contrast, the effective control test demands evidence of direct and specific control over each operation, making the overall control test a far more accessible standard for holding states responsible. However, overall control test has the potential to overextend State responsibility by attributing acts of NSA to a State based on general support rather than specific operational control. Unlike the effective control test, the overall control test allows for attribution if a State exerts a broad influence over an armed group, such as providing training, financing or logistical support, without requiring evidence of direct involvement in particular wrongful acts.

If the *effective control* test had been applied in *Tadić*, the outcome might have been different possibly exonerating Serbia from liability. In the context of *Tadić*, this would mean showing that Serbia not only provided general support to the Bosnian Serb forces but also issued direct instructions or maintained close oversight of the specific acts of aggression and violations of international law in question.

Article 9

Article 9 addresses situations where groups or individuals exercise elements of governmental authority in the absence of official authorities and without any actual legal authority to do so.[39] The exceptional nature of these circumstances is highlighted

[37] Ibid 117.
[38] *Tadid* (n32).
[39] ARSIWA Commentary, Art. 9.

by the phrase'in circumstance such as to call for'. In *Yeager v Islamic Republic of Iran*[40], the Tribunal dealt with claims by Thoman Yeager, against Iran following his detention during the 1979 Iranian Revolution. Yeager argued that his detention was attributable to Iran. The Tribunal found sufficient evidence that after 11 February 1979, the revolutionary Komitehs were acting on behalf of the new government and exercising elements of governmental authority in the absence of official authorities.[41] The Tribunal ultimately held Iran responsible for Yeager's mistreatment, emphasizing that the revolutionary groups were acting as substitutes for the absent official authorities, this meeting the criteria set out in Article 9.

However, it can be argued that the stringent conditions of article 9 makes it less effective in ensuring state responsibility. The requirement that the NSA perform governmental functions on their own initiative emphasizes the nature of the activity rather than the state's potential tacit approval. This undermines accountability like *Yeager* where private individuals like the Komitechs act as unauthorized enforcers, carrying out functions reserved for the state. Despite their actions resembling governmental authority, the absence of explicit state delegation creates a loophole for evading liability.

Article 10

Traditionally, a state cannot be held responsible for the acts of insurgents.[42]. Such a movement cannot align to the interests of the State against which it is fighting. The movement external to the formal structure of the state, exercises no constituted authority and cannot be considered to be under the state's direction or control. Vattel clarified this;

> *' A civil war breaks the bond of society and government,or at least suspends their force and effect'*[43]

In practice, what matters is whether the conduct of an insurrectional movement can be attributed to the state, and Article 10 provides that such conduct is not attributable unless the insurgents succeed in establishing a new state.[44] Thus, Bonn argues that

[40] *Yeager v Islamic Republic of Iran* (1987) 17 US–Iran CTR 91
[41] Ibid,104
[42] ARSIWA Commentary, Art. 10, 2, 3. Also Draft Articles Commentary, Art. 14, 4
[43] Vattel (1758), Bk III, Ch. XVIII, 293 (645).
[44] *Crawford* (n 18) 34

Article 10 functions as a negative attribution clause, accompanied by a peculiar form of contingent responsibility that arises only if the insurgency succeeds.[45] It fails to take into account state responsibility if the insurrections fails. Violent groups may then leverage this gap to achieve a new government, knowing their actions will be diminished into state responsibility. This creates inconsistency where the same acts (like war crimes) are treated unequally depending on the success and failure of the movement. Thus, state responsibility may arise on a limited basis where the insurrection succeeds in creating either a new government or even a new state. While article 10 is appropriate when considering a monolithic insurrectionist movement, difficulties emerge in situations where multiple insurgency groups act so as to bring about the creation of a new state, even where only one of those groups succeeds in forming a government.

Article 11

Article 11 addresses a scenario where NSAs may be attributed to a state—specifically, when the state acknowledges and adopts the conduct of the NSA as its own.[46] In *U.S. v Tehran*,[47] the Iranian government did not acknowledge the seizure of the U.S. embassy by militants at the outset. However, the ICJ found that Iran bore responsibility due to its subsequent endorsement of the militant's actions, by issuing public statements of its leadership and their failure to take action in stopping the occupation. The *Tehran* case demonstrates that private acts can be attributed to a state if the state endorses the act. This complements article 11 in expanding state attribution to a state illustrating that state responsibility can arise through acknowledgement. The case further confirms that explicit acknowledgement, rather than mere acquiescence is generally required for attribution under Article 11.

However, the precedent also presents an opportunity to avoid responsibility by refraining from making public statements that may be interpreted as acknowledgement. By focusing solely on explicit acknowledgment, the article allows states to avoid responsibility by remaining silent or issuing ambiguous statements. Like *Tehran*, where the ICJ relied on a broader interpretation of state conduct, Article

[45] *Boon* (n 11) 13
[46] ARSIWA Commentary, Art. 11, §1.
[47] Tehran Hostages, ICJ Rep. 1980 p. 3, 35

11’s narrow framework would fail to hold states accountable in similar scenarios. Thus, while article 11 illustrates that acknowledgement can lead to state responsibility, it also presents an opportunity for states to escape liability if they remain silent.

Reforming Attribution under ARSIWA: Strengthening Accountability for Non-State Actors

The increasing dependence of States on NSAs, particularly armed groups and private entities, stresses that the current attribution framework under ARSIWA fall short to address indirect forms of State involvement in IWAs. The preceding discussion has shown that the existing attribution rules, especially under Articles 5,7,8,9, 10 and 11, impose high evidentiary thresholds and rigid control requirements that enable States to evade responsibility while benefiting from the conduct of NSAs. To preserve the legitimacy and effectiveness of the law of State responsibility, reforms to the attribution framework are necessary.

Lowering the Evidentiary Threshold under Article 8

The effective control test established in *Nicaragua* and reaffirmed in the *Bosnian Genocide* case imposes an exceptionally high evidentiary burden on claimants. Requiring evidence that a State directed or controlled a particular operation frequently makes attribution practically impossible, markedly where relevant evidence becomes classified or inaccessible. As demonstrated throughout this article, States may finance, equip, train and strategically support armed groups while avoiding responsibility because direct operational instructions cannot be proven.

One possible reform would be to lower the evidentiary threshold under Article 8 by allowing attribution where substantial support and operational dependence are established. Rather than requiring proof of direct instructions for each wrongful act, international law could recognize attribution where a State provides sustained military, logistical or financial assistance that enables that conduct in question.

Re-structure the Overall Control Standard

A second reform would require codifying a modified version of the overall control test developed in *Tadić*. Unlike the effective control test, the overall control standard acknowledges that States can exercise decisive influence over organized armed groups without directing every specific operation. The current inconsistency between

the ICJ's effective control test and the ICTY's overall control approach has established uncertainty in international law concerning the proper threshold for attribution. Codification of an overall control standard within ARSIWA would promote greater legal clarity and consistency. However, to avoid overreaching State responsibility, the standard should require evidence of continued management, strategic guidance and material support between the State and the armed group. This would present a more balanced structure competent of addressing indirect State involvement while still respecting the principles of State sovereignty.

Conclusion

In conclusion, while ARSIWA serves a foundational framework in holding states responsible under international law, articles 5 through 11 reveal significant gaps holding States accountable for the actions of NSA under the rules of attribution. The critique is central at the rigid, complex and high thresholds of the control tests, coupled with the high evidentiary burden on plaintiffs to prove and provide explicit classified evidence. Accordingly, States use NSA as tactics to carry out illegal activities for their military and strategic objectives while simultaneously evading responsibility under international law. This demonstrates the need to reform ARSIWA so that it better reflects the realities of modern welfare. Reforms such as lowering the evidentiary threshold under Article 8, codifying an overall control standard, and introducing rebuttable presumptions of State involvement could strengthen the effectiveness of the attribution regime. Until ARSIWA is re-moulded and reconstructed to reflect the realities of modern welfare, ARSIWA will continue to become fragile in holding states accountable through NSA.

References

Cases

1. *Bosnian Genocide*, 2007 I.C.J. 43, 208-09, 400-01.
2. *Caire Claim Case*, France v Mexico [1929] 5 RIAA 516.
3. Certain Questions Relating to Settlers of German Origin in the Territory Ceded by Germany to Poland (Advisory Opinion) [1923].
4. *EDF (Services) Limited v Romania* ICSID Case No. ARB/05/13, 8 October 2009, 197–8.
5. German Settlers in Poland 'Advisory Opinion' (*International Court of Justice*, 23 November 2020) < https://jusmundi.com/en/document/decision/en-german-settlers-in-poland-advisory-opinion-monday-10th-september-1923> accessed 27 November 2024 .
6. *Hyatt International Corporation v Iran* (1985) 9 Iran–US CTR 72, 94.
7. Nicaragua, 1986 I.C.J. 115.
8. *Noble Ventures, Inc. v. Romania*, ICSID Case No. ARB/01/11, 82
9. *Socony Vacuum Oil Company* (1955) 21 ILR 55
10. Tadid, Case No. IT-94-1-A, Appeals Chamber Judgment, 115.
11. Tehran Hostages, ICJ Rep. 1980 p. 3, 35
12. Yeager v Islamic Republic of Iran (1987) 17 US–Iran CTR 91

Journals

13. Antonio Cassese James Crawford,' The Nicaragua and Tadić Tests revisited in light of the ICJ Judgement on Genocide in Bosnia' (*Oxford University Press*, 2013) https://watermark.silverchair.com > accessed 25 December 2024
14. Boon K'Attribution in International Law' (Cambridge University Press, 2024) <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/14E496EF5304E7B788AA10779EF51675/9781009334679c2_25-46.pdf/attribution-in-international-law.pdf > accessed 28 November 2024
15. Crawford J,' Organs and entities exercising governmental authority' (*Cambridge University Press*, 2013) < https://www.cambridge.org/core/services/aop-cambridge.pdf> accessed 25 December 2024
16. Crawford J, 'State Responsibility' (Cambridge University Press, 2013) < https://www.cambridge.org/core/books/state-

responsibility/EE846D9378B83A9DCC5E794FC086B07E > accessed 30 November 2024

17. Kinnear M, '*ARSIWAM ISDS, AND THE PROCESS OF DEVELOPING AN INVESTOR-STATE JURISPRUDENCE'* (Cambridge University Press, 2021) < https://www.cambridge.org/core/services/aop-cambridgecore/content/view/2C5A3E1EBC1BC50E2301E52D0530C887/S2633900521000473a.pdf/arsiwa_isds_and_the_process_of_developing_an_investorstate_jurisprudence.pdf > accessed 25 November 2024
18. Kelsen H, The Pure Theory of Law (2nd edn, 1960; trans. Knight 1967), 290–312. Also German Settlers in Poland, (1923) PCIJ Ser. B No. 6, 22: 'States can only act by and through their agents and representatives.'
19. Stefano C, 'The Meaning of Attribution' (*Oxford University Press*, 2020) < https://watermark.silverchair.com/oso-9780198844648-chapter-2.pdf> accessed 25 November 2024
20. Vattel (1758), Bk III, Ch. XVIII, 293 (645)

Authors Bio

Sosefo Sepasitiano Patelisio is a 2026 graduate of the University of Bristol, holding a Master of Laws (LLM) in International Law and International Relations. He currently serves as an Assistant Crown Counsel at the Attorney General's Office of Tonga, where his work focuses on legal advice, public international law, and government policy matters. Originally from the villages of Kolonga and Tatakamotonga in Tonga, he is a proud son of Kosilio and Patinia Patelisio, whose sacrifices, guidance, and unwavering support have shaped his journey. His research interests include international law, climate change law, and issues affecting Small Island Developing States in the Pacific. Beyond his legal career, he enjoys farming and planting crops, which keep him connected to his family, community, and cultural heritage.

Life Science

Meet the Subject Editor

Heston Brocklehurst
Life Science

My name is Heston, and I'm an MScR student in plant sciences at the University of Bristol. Last year, I graduated with a BSc in Biology from the University of Bristol, where over the course of the degree, I became fascinated by the incredible diversity of plants and their uses as foods, medicines and materials. Protecting the astounding diversity on our planet is our duty and a primary cause of its destruction is through unsustainable agricultural practices. I believe that understanding plant ecology is essential to achieving sustainable agriculture and food security simultaneously. Learning about plant ecology also happens to be incredibly interesting.

The Effect of Dominance Rank on Anemone Use in a High-Density Captive Population of the Common Clownfish *Amphiprion ocellaris*

Authors: Johnny Magnay and Fazeed Kotta

ORCIDs: 0009-0000-0726-9406 | 0009-0009-3417-7262

Programmes: BSc Zoology | BSc Biological Sciences

Affiliation: School of Biological Sciences

Emails: ec22818@alumni.bristol.ac.uk and npnd0209@leeds.ac.uk

Abstract

In socially hierarchical species, social dominance often mediates access to resources. Clownfish conform to this trend, with host anemones a contested resource. However, the persistence of this relationship in captivity remains unknown. Our study investigates the relationship between dominance status and anemone use in a high- density captive population of the common clownfish (*Amphiprion ocellaris*). Using continuous focal sampling, we recorded dyadic agonistic interactions and time spent in anemones during 20-minute periods, inferring a dominance hierarchy using David's scores. Opposing our predictions, high ranking individuals exhibited significantly lower anemone use than low ranking, but not medium ranking individuals. Potential explanations include trade-offs between dominance related aggression and resource use, and captive conditions making anemones act predominantly as refuges from conspecific aggression. Our findings suggest potential implications of housing *captive A. ocellaris* at high densities, informing improvements to welfare. Future research should repeat our methodology across low- and high-density captive populations to disentangle our explanations, using video recordings to avoid activity related confounds.

Keywords *Amphiprion ocellaris*, clownfish, dominance hierarchy, resource use, animal welfare

We would like to thank our supervisor Dr Bex Pike for her continued guidance, constructive feedback, and support throughout the duration of this project, as well as Mark Iwachiw and all the Bristol Aquarium staff for providing the resources and support necessary to carry out this research.

Introduction

Whilst clownfish social hierarchies and anemone symbioses have been widely explored, the relationship between dominance status and anemone use in captivity has received little attention. This remains true in the common clownfish (*Amphiprion ocellaris*), a small protandrous hermaphrodite[1] native to the Indo-Pacific Ocean[2] where it forms a symbiotic relationship with sea anemones[3], maintaining a strict size-based social hierarchy consisting of a dominant female, a breeding male, and up to 3 subordinates[1]. Clownfish subordinates queue for breeding positions[4], remaining sexually immature to prevent eviction from the host anemone by the dominant pair[5]. Avoiding eviction is crucial, as host anemones provide clownfish with services including protection from predators[6], access to food, external parasite removal[7], and egg protection[2]. In this way, anemones act as an essential resource, often coming under intense competition[8,9]. In most socially hierarchical species, access to such localised resources is mediated by social dominance, with higher-ranking individuals gaining priority access[10]. Although research into this association in clownfish is limited, current evidence suggests that they comply with this trend. For example, Chen and Hsieh[11] found dominance mediated shelter use in *A. ocellaris*, and Huebner *et al*.[12] showed that larger *Amphiprion bicinctus* monopolise species of anemone providing greater protection from predators[13], restricting subordinates to less preferred species. Despite the recent popularity of these species in aquaria[14], no studies have assessed the relationship between clownfish dominance status and anemone use in captivity. Captive fish are often housed in confined conditions at unnaturally high densities[15], potentially increasing stress and disrupting ecological processes[16]. Understanding whether natural behaviours persist under such conditions is therefore essential for improving the management and welfare of captive clownfish. In this study, we investigate the association between dominance status and anemone use in a single high-density captive population of *A. ocellaris*, with a preliminary aim of inferring the dominance hierarchy. We hypothesise that the proportion of time spent in an anemone will increase with dominance rank. This prediction is based on the idea that anemones act as localised resources which will be controlled by more dominant individuals.

Methodology

(a) Experimental setup

This study received ethical approval from the UK Animal Welfare and Ethical Review Body (AWERB) under reference code UIN-24-069. Observations took place in a 690l display tank at Bristol Aquarium, England, 4 days per-week between the 7th October and 22nd November 2024, with a preceding pilot day. The tank housed 18 unrelated *A. ocellaris* (11 orange and 7 melanistic morphs) originating from captive breeding programmes and public donations, all acclimated for over a year, minimising the effects of observer presence. Semi-natural conditions were replicated using rocks, sand, 10-12 *Entacmea quadricolor* anemones which propagated during our study, and a mixed-species assemblage of fish and invertebrates (Appendix 1). Whilst *E. quadricolor* is not a natural host for *A. ocellaris*[17], this pairing is often used in behavioural studies (e.g. Wong *et al.*[18]), where symbioses form due to the absence of host anemones (but see Nguyen *et al.*[19]). Environmental conditions were maintained at 25°C, 89-96% dissolved oxygencontent, and 35PPT salinity.

(b) Sampling focal individuals

Prior to beginning data collection,we identified all observable clownfish and assigned each a unique identifier based onindividual stripe patterns, excluding an individual observed only 9 times and first identified after 2 weeks of sampling. Each day, 7 focal individuals were randomly selected from the total (with replacement) using the app 'Random Number Generator' by UXPAPPS-LTD[20], controlling for sampling bias. A pilot day ensured competency with fish identification and sampling methods.

(c) Recording behavioural parameters

Observations took place between 10am-2pm to control for temporal variation in activity, a trait observed in *A. ocellaris*[17], with start times varied between 10-10.30am to minimise confounding effects of the aquarium's routine. Although controlled for through random sampling, observations were paused for 5 minutes during feeding and cleaning events to minimize disturbance. To assess anemone use and infer a dominance hierarchy, we recorded 2 behavioural parameters simultaneously for each focal individual: (i) The duration of each contact with an anemone, excluding contact under 3 seconds due to observational constraints. 'Out-of-sight' conditions were

accounted for as all anemones remained in view for the study duration; (ii) The outcome (winner and loser) and partner ID of dyadic agonistic interactions, using an ethogram developed from preliminary observations and published literature (Table 1). Agonistic interactions were only recorded when an aggressive or threatening provocation elicited a submissive response, to avoid misidentification of conflict outcomes and agonistic partners. Whilst some studies consider clownfish 'shaking' behaviours a submissive response[11], there is ambiguity regarding their role as a dominant or submissive behaviour[2], so these were excluded from analysis.

Category	Behaviour	Description
Aggressive	Chasing	Aggressor rapidly approaches recipient for > 1 body length
	Lunging	Aggressor rapidly approaches recipient for < 1 body length
	Biting	Aggressor bites recipient
Threatening	Face-to-face swimming	Aggressor and recipient swim synchronously face-to-face at < 1 body length
	Side-by-side swimming	Aggressor and recipient swim synchronously side-by-side at < 1 body length
Submissive	Fleeing	Recipient rapidly retreats from aggressor for > 1 body length
	Flinching	Recipient rapidly retreats from aggressor for < 1 body length

Table 1: Ethogram of agonistic and submissive behaviours of the common clownfish Amphiprion ocellaris. *Adapted from Iwata and Manbo, Chen[21] and Hsieh[11], and preliminary observations. A focal individual may act as either the aggressor or recipient across interactions.*

After randomly allocating investigators to each parameter to control for investigator bias, we utilized continuous focal sampling by direct observation to measure both parameters during 20-minute observation periods, facilitating detailed recordings beyond that achievable using less targeted methods such as scan sampling. Whilst

video recordings are often preferred as they allow for longer sampling periods and more detailed behavioural assessments[22,23], the public setting of our study rendered direct observations a more suitable alternative. Inter-observer reliability was assessed using joint observations of both parameters. To remedy the lack of consideration for chance agreements in percentage agreement[24], we instead used the R package "psych"[25] to apply an unweighted Cohen's kappa[26] to the nominal dominance parameter (k=0.97, 95% CI[0.92-1], n=68), and an intra-class correlation coefficient ICC(2,1) to the continuous anemone use parameter (ICC=1, 95% CI[1-1], n=97), both indicating excellent agreement.

(d) Statistical analyses

Statistical analyses wereperformed using R (v4.2.1). To infer a potential dominance hierarchy, we calculatednormalised David's scores (normDS) using the package "EloRating"[27], using the*Dij*method (de Vries' correction[28]) to account for imbalances in dyadic interaction frequencies, controlling for the influence of spatial proximity and behavioural variation. NormDS were chosen because they consider an individual's success over all dyadic interactions and have been widely used in social animals[29,30]. We subsequently assessed hierarchy steepness (strength) by calculating the slope of a regression line fitted to a plot of normDS as a function of rank, as proposed by de Vries, Stevens and Vervaecke[31]. Additionally, we assessed hierarchy linearity by using the "aniDom" package[32] to calculate triangle transitivity (Ttri), the most reliable method when not all dyadic outcomes are known[33], suiting our data. As suggested by Sánchez-Tójar, Schroeder and Farine[34], we assessed hierarchy uncertainty and resolution by calculating Spearman's rank correlations between 2 halves of the dataset using the package 'aniDom'[32], and the ratio of interactions to individuals sampled, with rs=0.5 and 10-20 interactions/individual the recommended minimum for each respective metric.

To test the relationship between dominance rank and anemone use, the range of normDS values was divided into low, medium, and high categories of equal width, and individuals were assigned a category according to their normDS score. The percentage time spent in an anemone was then calculated for each 20-minute sample. Due to substantial variation in our data, we used non-parametric tests to reduce the impact of outliers, testing for differences in median anemone use between the

dominance categories using a 2-tailed Kruskal-Wallis test with a 5% significance level, followed by pairwise comparisons using Wilcox rank sum test with a Benjamini-Hochberg correction for multiple testing.

Results

Over 21 days of experimentation, we collected 136 data points for the anemone use parameter and recorded 1386 dyadic agonistic interactions. Each fish was randomly sampled between 6 and 11 times.

(a) Social dominance hierarchy

NormDS ranged between 4.18-12.22 and were different for all individuals, revealing the presence of a dominance hierarchy (Figure 1). The hierarchy was only of low to moderate steepness (0.39; Figure 1) but had a high transitivity which was significantly different from random (Ttri=0.88, p<0.001). Uncertainty was low (rs=0.80, 95% CI[0.63-0.93]), indicating high reliability, and a high ratio of 81.53 interactions to individuals sampled showed sufficient sampling effort. Low ranks received the most aggression per individual (195 instances) followed by high (44) medium (29) ranks.

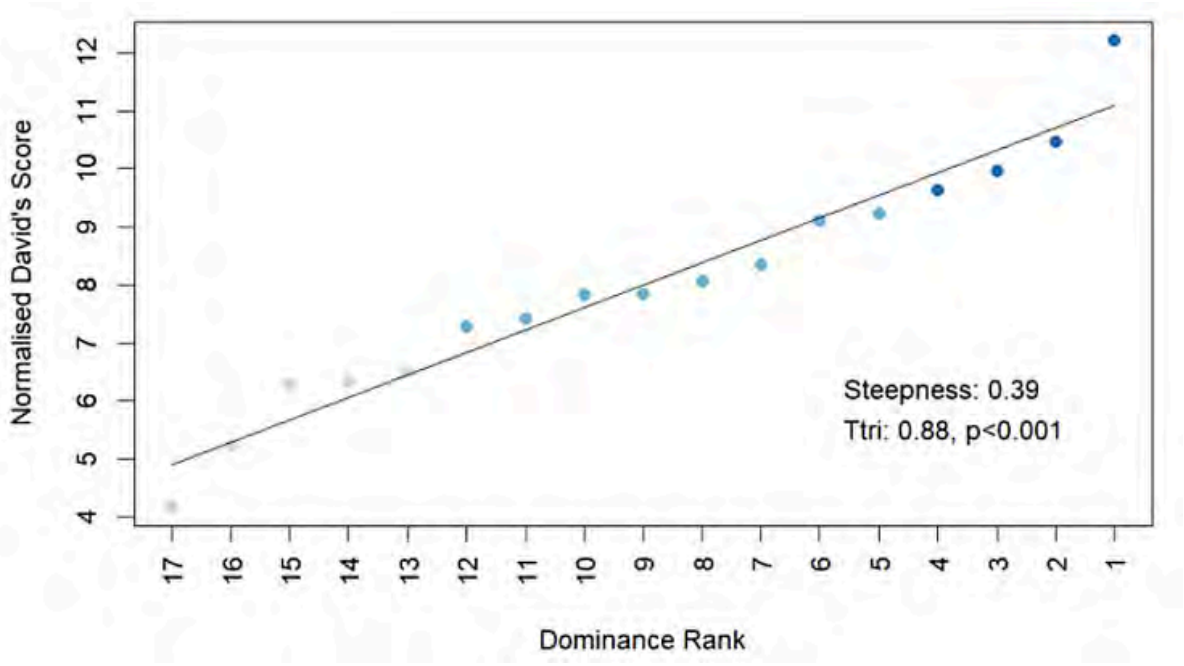


Figure 1: Normalised David's scores and associated dominance rank, showing overall steepness and linearity.*Nodes are coloured accordingtodominance category(low,medium, high), and a regression line y = 0.39 x + 4.51 depicts overall hierarchy steepness.*

(b) Anemone use

The population hada low median anemone use ($\tilde{x}$ =16.33%) with moderate variation (IQR[3.10-32.69]),and 8.09% of data falling as statistical outliers. The low, medium, and high ranks contained 5, 8, and 4 individuals respectively. Low ranking fish had the highest median and largest variability in anemone use ($\tilde{x}$ =27.17%, IQR[17.73- 62.33]) followed by the medium ($\tilde{x}$ =12.29%, IQR[1.06-33.29]) and high ($\tilde{x}$ =8.38%, IQR[1.83-18.31]) ranks. We found a significant difference in median anemone use between the dominance categories (χ^2 =20.02, df=2, p<0.0001; Figure 2), with pairwise comparisons revealing significantly higher anemone use in the low rank than the medium or high rank (Pairwise; low vs medium: p<0.01; low vs high: p<0.0001; Figure 2). However, no significant difference was found between medium and high ranks (Pairwise; medium vs high: p =0.230; Figure 2).

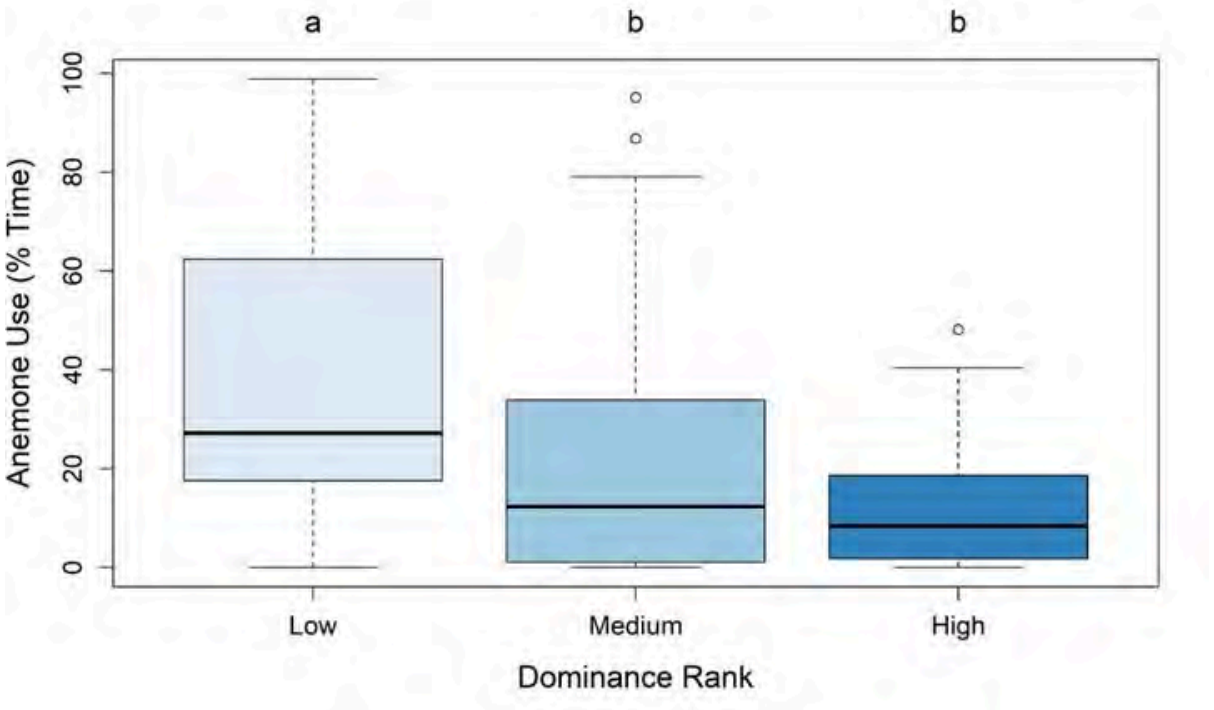


Figure 2: Percentage of time spent in an anemone by clownfish of different dominance ranks. *Different letters aboveboxes denotesignificantdifferencesbetweengroups(p<0.05), whereas shared letters indicate no significant difference.*

Discussion

To explorehow anemone use may be influenced by clownfish dominance status in a high-densitycaptive population, we inferred a dominance hierarchy and tested for differencesin anemone use between low, medium, and high-ranking individuals, predictingapositive association due to greater social dominance granting priority access tothese localised resources.

(a) Dominance hierarchy

The distinct David's scoresand high linearity of our study population imply a single linear dominance hierarchyacross all 17 individuals (Figure 1), with observations suggesting it largely follows the natural size-based structure[35], although future research could test this experimentally by measuring clownfish size. However, it was larger and moderately weaker than wild clownfish hierarchies [1,35], likely due to the high-density captive environment. In the wild, clownfish reject arrivals when their anemone reaches capacity[36], regulating group size and maintaining steep, size-based distances between ranks which reduce conflict[35]. In captivity, new arrivals cannot disperse, increasing population density and potentially reducing inter-rank steepness. Supporting this, we observed a high frequency of agonistic interactions, a trait often used to identify weak hierarchies[22], although this may instead be explained by greater encounter frequencies in larger populations[37].

We controlled for variation in dyadic interaction frequency by incorporating de Vries' correction in our David's score calculation, but technical limitations prevented us from accounting for acoustic dominance signals, an important regulator of clownfish hierarchies[38]. Despite this, our uncertainty and resolution metrics confirm that our methods were sufficient to replicate the present hierarchy, making it a suitable foundation for investigating the effects of dominance status on anemone use in our population.

(b) Anemone use

Opposing our predictions, low ranking individuals showed significantly higher anemone use than both medium and high ranks (Figure 2). This contrasts with previous research finding that high dominance grants priority resource use in *A.*

ocellaris[11] as well as a range of other socially hierarchical species (crayfish[39], chickens[40], primates[10,41]. There is little research agreeing with our findings; although interestingly, our results align with a study of crayfish, which, using a larger population than Martin and Moore[39], observed higher shelter-use in less dominant individuals[42]. By comparing these studies, we can infer a potential explanation for our results; group-size dependent trade-offs between dominance-related aggression and resource use. Such trade-offs have been found in *A. ocellaris*[43], where in small groups, high ranking individuals expend little effort reinforcing their dominance, allowing them to dominate localised resources, as observed by Chen and Hsieh[11]. In larger groups, however, greater competition may lead high ranking individuals to perform higher frequencies of dominance-reinforcing aggression[42], trading-off resource use, and leaving low- ranks better able to exploit resources, potentially contributing to the high anemone use among low-ranking individuals in our study. Opposing this idea, a study of *Amphiprion clarkii* found no link between group size and average aggression rates[37]. However, they did not consider rank-specific changes in aggression, limiting the applicability of their results.

Another possible explanation for our findings is the removal of ecological pressures including predation risk and food scarcity in captivity, which may reduce predator avoidance and competitive behaviours[44] and minimise the reliance of clownfish on their host anemone's ecological services[7,9]. This effect might be enhanced by our use of the non-host anemone *E. quadricolor* which can offer lower fitness benefits than natural hosts[17]. Under these reduced ecological pressures, anemones may instead act predominantly as refuges from conspecific aggression, similar to shelter-use in other captive organisms[45]. Our results support these ideas, as low-ranking individuals both received the most aggression and spent the most time in anemones, and medium and high ranks were similarly low in both measures (Figure 2). To disentangle this explanation from that regarding trade-offs between aggression and resource use, our methodology should be replicated using both low- and high-density captive *A. ocellaris* populations under the same conditions. If our findings are primarily driven by group-size-dependent trade-offs, we would expect dominant individuals to exhibit the greatest anemone use in small groups, whereas larger groups should replicate our observations. Due to our use of a single clownfish population, caution is needed when extrapolating these results to other captive populations of *A. ocellaris*. However, as a

general recommendation, aquaria should ensure clownfish are housed at low population densities with sufficient shelter, preferably in the form of natural host anemone species. This should minimise aggression and reinforce natural behaviours, a key aspect of maintaining welfare[46].

Although our data exhibited substantial variation, the presence of outliers did not affect the significance of our results, reinforcing the robustness of our statistical methodology. Despite this, these outliers remain responsible for the greater variation observed in low-ranking fish, hence ecological explanations for this trend are not discussed. Both sampling and investigator bias were controlled for, as were factors relating to fluctuating environmental conditions, spatial and temporal variation, observer presence, and the aquarium's routine. Whilst Hussey and LaPlante[43] reported differences in aggression-maintenance trade-offs between orange and melanistic *A. ocellaris* morphs, their small sample size (n=3) introduces confounds of individual variation in behaviour, a trait present in *A. ocellaris* and related species[18,47], so our use of both morphs is unlikely to detract from our results. However, our use of a 3 second threshold for recording anemone use due to observational constraints may have underestimated the anemone use of highly active individuals. If activity increases with dominance rank in *A. ocellaris*, as observed in laboratory mice[48], this may prove confounding. Videorecording-based observations are recommended for future studies tominimise these confounds.

Conclusion

Our study explored the relationship between dominance status and anemone use in captive *A. ocellaris*. We inferred a single linear dominance hierarchy, allowing assessments of dominance-mediated anemone use. Contrary to our predictions, high ranking individuals spent significantly less time in anemones than low ranking, but not medium ranking individuals. This may reflect group-size-dependent trade-offs between dominance-related aggression and resource use, as well as reduced ecological benefits provided by anemones in captivity where they might instead act as refuges from conspecific aggression. These findings suggest potential implications of housing *A. ocellaris* under high-density captive conditions, hence we recommend aquaria maintain low population densities with many host anemones to encourage

natural behaviours and improve welfare. Our study serves as a foundation for research seeking to better understand how clownfish dominance status interacts with other behavioural traits in captive environments. Future research should repeat our methodology across multiple low- and high-density captive A. ocellaris populations to disentangle our explanations, using video recordings to avoid the activity-related confounds which may limit our study.

References

1. Moyer J,Nakazono A. Protandrous Hermaphroditism in Six Species of the Anemonefish Genus Amphiprion in Japan. *The Ichthyological Society of Japan*. Preprint posted online 1978:2. doi:10.11369/jji1950.25.101

2. Allen GR. *The Anemonefishes: Their Classification and Biology*. T. F. H. Publications; 1972.

3. Elliott JK, Elliott JM, Mariscal RN. Host selection, location, and association behaviors of anemonefishes in field settlement experiments. *Marine Biology*. 1995;122(3):377-389. doi:10.1007/BF00350870

4. Buston PM. Territory inheritance in clownfish. *Proc R Soc Lond B*. 2004;271(suppl_4). doi:10.1098/rsbl.2003.0156

5. Rueger T, Barbasch TA, Wong MYL, Srinivasan M, Jones GP, Buston PM. Reproductive control via the threat of eviction in the clown anemonefish. *Proc R Soc B*. 2018;285(1891):20181295. doi:10.1098/rspb.2018.1295

6. Fautin D. *The Anemonefish Symbiosis: What Is Known and What Is Not. Symbiosis.* 10 (1-3). https://kuscholarworks.ku.edu/entities/publication/ed3447a6-e5c6-4e01-a398-d22b5cbe3b19

7. Mariscal RN. The nature of the symbiosis between Indo-Pacific anemone fishes and sea anemones. *Marine Biology*. 1970;6(1):58-65. doi:10.1007/BF00352608

8. Fricke HW. Mating System, Resource Defence and Sex Change in the Anemonefish Amphiprion akallopisos. *Zeitschrift für Tierpsychologie*. 2010;50(3):313-326. doi:10.1111/j.1439-0310.1979.tb01034.x

9. Fautin D. *Anemonefish Recruitment: The Roles of Order and Chance. Symbiosis.* 1992. https://www.researchgate.net/publication/43091532_Anemonefish_recruitment_The_roles_of_order_and_chance

10. Overduin-de Vries AM, De Vries H, Vermande MM, Reijntjes AHA, Sterck EHM. Both aggressive and affiliative behaviour facilitate resource access in high-ranking

female long-tailed macaques (Macaca fascicularis). *Behav*. 2020;157(3-4):267-287. doi:10.1163/1568539X-bja10001

11. Chen TH, Hsieh CY. Fighting Nemo: Effect of 17α-ethinylestradiol (EE2) on aggressive behavior and social hierarchy of the false clown anemonefish Amphiprion ocellaris. *Marine Pollution Bulletin*. 2017;124(2):760-766. doi:10.1016/j.marpolbul.2016.12.042

12. Huebner L, Dailey B, Titus B, Khalaf M, Chadwick N. Host preference and habitat segregation among Red Sea anemonefish: effects of sea anemone traits and fish life stages. *Mar Ecol Prog Ser*. 2012;464:1-15. doi:10.3354/meps09964

13. Howell J, Goulet TL, Goulet D. Anemonefish musical chairs and the plight of the two-band anemonefish, Amphiprion bicinctus. *Environ Biol Fish*. 2016;99(11):873-886. doi:10.1007/s10641-016-0530-9

14. Militz TA, Foale S. The “Nemo Effect”: Perception and reality of *Finding Nemo* ’s impact on marine aquarium fisheries. *Fish and Fisheries*. 2017;18(3):596-606. doi:10.1111/faf.12202

15. Torgerson-White L, Sánchez-Suárez W. Looking beyond the Shoal: Fish Welfare as an Individual Attribute. *Animals*. 2022;12(19):2592. doi:10.3390/ani12192592

16. Beaulieu M. A Bird in the House: The Challenge of Being Ecologically Relevant in Captivity. *Front Ecol Evol*. 2016;4. doi:10.3389/fevo.2016.00141

17. Nguyen HT, Tran AT, Ha LTL, Ngo DN, Dang BT, Geffen AJ. Host choice and fitness of anemonefish ***AMPHIPRION OCELLARIS*** (Perciformes: Pomacentridae) living with host anemones (Anthozoa: Actiniaria) in captive conditions. *Journal of Fish Biology*. 2019;94(6):937-947. doi:10.1111/jfb.13910

18. Wong MYL, Medina A, Uppaluri C, Arnold S, Seymour JR, Buston PM. Consistent behavioural traits and behavioural syndromes in pairs of the false clown anemonefish *Amphiprion ocellaris*. *Journal of Fish Biology*. 2013;83(1):207-213. doi:10.1111/jfb.12133

19. Nguyen HT, Zhao M, Wang T, Dang BT, Geffen AJ, Cummins SF. Sea anemone–anemonefish symbiosis: Behavior and mucous protein profiling. *Journal of Fish Biology*. 2024;105(2):603-618. doi:10.1111/jfb.15772

20. Random Number Generator. Published online 2016. https://play.google.com/store/apps/details?id=ru.uxapps.random&pcampaignid=web_share

21. Iwata E, Manbo J. Territorial behaviour reflects sexual status in groups of false clown anemonefish (Amphiprion ocellaris) under laboratory conditions. *acta ethol*. 2013;16(2):97-103. doi:10.1007/s10211-012-0142-0

22. Brena PF, Mourier J, Planes S, Clua EE. Concede or clash? Solitary sharks competing for food assess rivals to decide. *Proc R Soc B*. 2018;285(1875):20180006. doi:10.1098/rspb.2018.0006

23. Pini-Fitzsimmons J, Knott NA, Brown C. Heterarchy Reveals Social Organization of a Smooth Stingray (Bathytoshia brevicaudata) Population in a Provisioned Food Context. *Front Mar Sci*. 2021;8:641761. doi:10.3389/fmars.2021.641761

24. Kaufman AB, Rosenthal R. Can you believe my eyes? The importance of interobserver reliability statistics in observations of animal behaviour. *Animal Behaviour*. 2009;78(6):1487-1491. doi:10.1016/j.anbehav.2009.09.014

25. Revelle W. R package. Published online May 16, 2026. doi:10.32614/CRAN.package.psych

26. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 1960;20(1):37-46. doi:10.1177/001316446002000104

27. Neumann C. Animal Dominance Hierarchies by Elo Rating. Published online July 15, 2024. doi:10.32614/CRAN.package.EloRating

28. De Vries H. Finding a dominance order most consistent with a linear hierarchy: a new procedure and review. *Animal Behaviour*. 1998;55(4):827-843. doi:10.1006/anbe.1997.0708

29. Koren L, Mokady O, Geffen E. Social status and cortisol levels in singing rock hyraxes. *Hormones and Behavior*. 2008;54(1):212-216. doi:10.1016/j.yhbeh.2008.02.020

30. Jaeggi AV, Stevens JMG, Van Schaik CP. Tolerant food sharing and reciprocity is precluded by despotism among bonobos but not chimpanzees. *American J Phys Anthropol*. 2010;143(1):41-51. doi:10.1002/ajpa.21288

31. De Vries H, Stevens JMG, Vervaecke H. Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*. 2006;71(3):585-592. doi:10.1016/j.anbehav.2005.05.015

32. Farine DR, Sanchez-Tojar A. Inferring Dominance Hierarchies and Estimating Uncertainty. Published online March 6, 2021. doi:10.32614/CRAN.package.aniDom

33. Shizuka D, McDonald DB. A social network perspective on measurements of dominance hierarchies. *Animal Behaviour*. 2012;83(4):925-934. doi:10.1016/j.anbehav.2012.01.011

34. Sánchez-Tójar A, Schroeder J, Farine DR. A practical guide for inferring reliable dominance hierarchies and estimating their uncertainty. Morand-Ferron J, ed. *Journal of Animal Ecology*. 2018;87(3):594-608. doi:10.1111/1365-2656.12776

35. Buston P. Size and growth modification in clownfish. *Nature*. 2003;424(6945):145-146. doi:10.1038/424145a

36. Buston P. Forcible eviction and prevention of recruitment in the clown anemonefish. *Behavioral Ecology*. 2003;14(4):576-582. doi:10.1093/beheco/arg036

37. Yllan L, Heatwole S, Wong M, Rueger T. Effect of social context on behaviour in anemonefish hierarchies. *Animal Behaviour*. 2024;209:83-93. doi:10.1016/j.anbehav.2023.12.014

38. Colleye O, Parmentier E. Overview on the Diversity of Sounds Produced by Clownfishes (Pomacentridae): Importance of Acoustic Signals in Their Peculiar

Way of Life. Neuhauss SCF, ed. *PLoS ONE*. 2012;7(11):e49179. doi:10.1371/journal.pone.0049179

39. Martin AL, Moore PA. The Influence of Dominance on Shelter Preference and Eviction Rates in the Crayfish, *Orconectes rusticus*. *Ethology*. 2008;114(4):351-360. doi:10.1111/j.1439-0310.2008.01473.x

40. Banks EM, Wood-Gush DGM, Hughes BO, Mankovich NJ. Social rank and priority of access to resources in domestic fowl. *Behavioural Processes*. 1979;4(3):197-209. doi:10.1016/0376-6357(79)90001-9

41. Jarvey JC, Low BS, Azanaw Haile A, et al. Aggression rates increase around seasonally exploited resources in a primarily grass-eating primate. Le Roux A, ed. *Behavioral Ecology*. 2024;35(1):arad079. doi:10.1093/beheco/arad079

42. Fero K, Simon JL, Jourdie V, Moore PA. *Consequences of Social Dominance on Crayfish Resource Use.* Vol 144. Brill; 2007. https://www.jstor.org/stable/i407277

43. Hussey RK, LaPlante LH. Correlation between aggression and maintenance behaviors of captive Amphiprion ocellaris in a host anemone: A comparison between two variants. *BIOS*. 2024;95(1). doi:10.1893/BIOS-D-22-00004

44. Berger-Tal O, Blumstein DT, Swaisgood RR. Conservation translocations: a review of common difficulties and promising directions. *Animal Conservation*. 2020;23(2):121-131. doi:10.1111/acv.12534

45. Capelli GM, Hamilton PA. Effects of Food and Shelter on Aggressive Activity in the Crayfish *Orconectes Rusticus* (Girard). *Journal of Crustacean Biology*. 1984;4(2):252-260. doi:10.2307/1548022

46. Greggor AL, Vicino GA, Swaisgood RR, et al. Animal Welfare in Conservation Breeding: Applications and Challenges. *Front Vet Sci*. 2018;5:323. doi:10.3389/fvets.2018.00323

47. Schmiege PFP, D'Aloia CC, Buston PM. Anemonefish personalities influence the strength of mutualistic interactions with host sea anemones. *Mar Biol*. 2017;164(1):24. doi:10.1007/s00227-016-3053-1

48. Robbers Y, Tersteeg MMH, Meijer JH, Coomans CP. Group housing and social dominance hierarchy affect circadian activity patterns in mice. *R Soc open sci*. 2021;8(2):rsos.201985, 201985. doi:10.1098/rsos.201985

Appendix

Appendix 1: Species composition of the study tank

Common name	Species
Common clownfish	*Amphiprion ocellaris*
Bubbletip anemone	*Entacmea quadricolor*
Red spotted hawkfish	*Amblycirrhites pinos*
Sand goby	*Pomatoschistus minutus*
Thin stripe hermit crab	*Clibanarius vittatus*
Pacific cleaner shrimp	*Lysmata amboinensis*
Fire shrimp	*Lysmata debelius*
Asterina starfish	*Asterina sp*

Authors Bio -

Jonny Magnay graduated from the University of Bristol's School of Biological Sciences in 2025 and has since begun a Postgraduate Research degree at the University of Glasgow. His research interests lie in applied marine ecology, with a focus on understanding how anthropogenic stressors influence ecological interactions, and on developing novel, evidence-based approaches to mitigate human impacts and support ecosystem restoration. Through his research, he aims to contribute to the conservation and recovery of marine habitats while advancing the understanding of our impacts on the natural world.

Authors Bio -

Fazeed Kotta is currently a postgraduate researcher (MRes) at the University of Leeds within the School of Biology. He graduated from the University of Bristol, with a BSc in Biology, and his research interests are animal behaviour and evolution, both the adaptations themselves and the genetics behind them. Previous research efforts include a review of the current understanding of the basal phylogenetics of Hexapods. He is currently studying Pinnipeds' milk and their underlying mechanisms and evolutionary pathways, and the implications for Mammalian diversity and evolution.

Investigating Natural Variation in Photosynthetic Temperature Responses Across *Arabidopsis Thaliana* Ecotypes and *Arabidopsis Lyrata*

Author: Danica D Cruz 0009-
ORCID: 0002-0870-9455
Programme: BSc Biology
Affiliation: SchoolofBiological Sciences
Email: danidcruz18@gmail.com

Abstract

As temperatures warm, understanding the traits that determine plant heat resilience has become increasingly important. This study examined physiological, biochemical and anatomical responses to warming (20°C vs. 30°C) in three *Arabidopsis thaliana* ecotypes (Col-0, Cvi, Ei-2) and the related perennial *Arabidopsis lyrata* MN-47. Photosynthetic parameters (α, ETRmax) differed among the ecotypes and *Arabidopsis lyrata* MN-47 but showed little change with warming. In contrast, PSII thermal tolerance (T_{50}) varied strongly: MN-47 maintained the highest T_{50} across temperatures, and Ei-2 was the only ecotype to exhibit a clear acclimatory increase at 30°C. Pigment concentrations and leaf structural traits also differed among ecotypes but were generally insensitive to warming. Together, these results show that variation in PSII thermal tolerance occurs independently of the biochemical and anatomical traits measured. The contrasting responses of MN-47 and Ei-2 point to mechanisms acting directly on PSII stability. Integrating gas-exchange and transcriptomic analyses will help identify these pathways and improve understanding of plant thermal resilience.

KEYWORDS: FOOD SECURITY, CROP IMPROVEMENT, GENETIC VARIATION

Introduction

Rising temperatures increasingly constrain plant performance, making it essential to understand how photosynthetic traits respond to heat[1]. *Arabidopsis thaliana* exhibits extensive natural variation across its global range, with contrasting climates and selective pressures driving local adaptation and genetic differentiation among populations[2]. This diversity provides a valuable system for examining how photosynthetic function responds to warming. The related perennial *Arabidopsis lyrata*, which often occupies cooler environments and differs in life history strategy, offers an additional evolutionary strategy that may reveal alternative mechanisms of heat tolerance[3].

Col-0, the standard laboratory reference ecotype, was originally collected near Columbia, Missouri, and is primarily valued for its well-characterised genome and extensive genetic resources rather than for any specific environmental adaptation. Cvi, collected from the Cape Verde Islands off northwest Africa (one of the geographically most isolated *A. thaliana* populations) has previously been associated with tolerance to photo-oxidative stress and altered stomatal behaviour, likely reflecting adaptation to high light and arid conditions. Ei-2 originates from the Eifel region of Germany; while its geographic context suggests a cooler temperate climate, specific photosynthetic characterisation of this ecotype and published climate data for its collection site are limited in the literature.

The *A. lyrata* accession MN-47 was collected in Minnesota, USA, and differs from *A. thaliana* principally in its perennial life history strategy[3]; published photosynthetic characterisation of MN-47 specifically at elevated temperatures is limited, though *A. lyrata* populations more broadly show natural variation in thermal stress resistance[3]. These four lines were selected to represent contrasting geographic origins and life history strategies, providing a framework within which to examine whether photosynthetic thermal responses reflect environmental background.

Heat stress can impair photosynthesis through several pathways, including reduced PSII efficiency, pigment degradation, and modifications to leaf structural traits such as leaf mass per area (LMA) and stomatal density[4–6]. Since these processes occur at physiological, biochemical, and anatomical levels, a multi-trait framework is required to

characterise plant responses to warming[7,8]. The extent to which such traits covary, and whether they collectively explain natural variation in thermal tolerance, remains uncertain.

This study asked three questions:

1. Do the four lines differ in photosynthetic performance, specifically initial rate (α) and maximal electron transport rate (ETRmax), when grown at 20°C and 30°C?
2. Do biochemical traits such as chlorophylls and carotenoids and anatomical traits such as LMA and stomatal density exhibit similar or divergent responses to growth temperature?
3. Do these traits account for variation in PSII thermal tolerance (T_{50})?

To address these questions, chlorophyll fluorescence imaging, pigment quantification, and leaf anatomical measurements were used to characterise naturally occurring variation in photosynthetic responses to elevated temperatures across *Arabidopsis* ecotypes and *A. lyrata* MN-47.

Methods

Plant material and growth conditions

Seeds of *A. thaliana* ecotypes (Col-0, Cvi, and Ei-2) and *A. lyrata* MN-47 were cold-stratified for 48 hours at 4°C, then sown into 8 cm pots filled with a 3:1 mixture of Levington F2 compost and silver sand. Plants were grown in controlled-environment cabinets at 20°C under a 16-hour photoperiod (120 µmol m^{-2} s^{-1}). Seedlings were transplanted midweek to ensure one plant per pot. After one week, half of the plants from each line remained at 20°C, while the other half were transferred to 30°C. Plants remained in their assigned temperature treatment until the time of sampling, ensuring that physiological, biochemical and anatomical measurements all reflected the same growth conditions.

Physiological trait measurements

Chlorophyll fluorescence and light-response curves (LRCs)

Photosynthetic performance was measured on two-week-old plants using an IMAGING-PAM fluorometer (Heinz Walz GmbH). Plants were dark-adapted for 20 min prior to measurement. LRCs were generated following 15 min pre-illumination at 798 µmol m^{-2} s^{-1} to stabilise photosynthesis. Actinic light was increased stepwise, with saturating pulses applied at each step. Two areas of interest (AOIs) were averaged per plant, yielding one biological replicate. Four LRC runs produced 16 replicates per line.

The effective quantum yield of PSII was calculated as:

$$\Phi PSII = (Fm' - F')/Fm'$$

where F' is steady-state fluorescence and F_m' is maximal fluorescence under actinic light.

Electron transport rate (ETR) was calculated as:

$$ETR = PAR \times \Phi PSII \times 0.84 \times 0.5$$

where PAR is incident irradiance at each light step, 0.84 is assumed leaf absorptance, and 0.5 reflects equal distribution of excitation energy between PSI and PSII.

αwas obtained as the slope of the linear portion of the ETR-PAR curve (first three light steps). ETRmax was defined as the maximal ETR prior to curve saturation.

T_{50}

T_{50} was assessed using detached leaves exposed to controlled temperature increments (Figure 1). Two fully expanded leaves per plant were placed between glass slides, sealed in two zip-lock bags, and laid flat in a temperature-controlled water bath. Baseline Fv/Fm

was recorded at 25°C. Temperature was increased from 41°C in 2°C intervals, with 5 min equilibration before each measurement. One AOI per leaf was recorded. Three independent runs produced 24 replicates per line.

Maximum PSII efficiency was calculated as:

$$Fv/Fm = (Fm - F_0)/Fm$$

where F_0 is the minimal fluorescence after dark adaptation and Fm is the maximal fluorescence induced by a saturating pulse.

T_{50} was defined as the temperature at which Fv/Fm declined to 50% of its initial value[9].

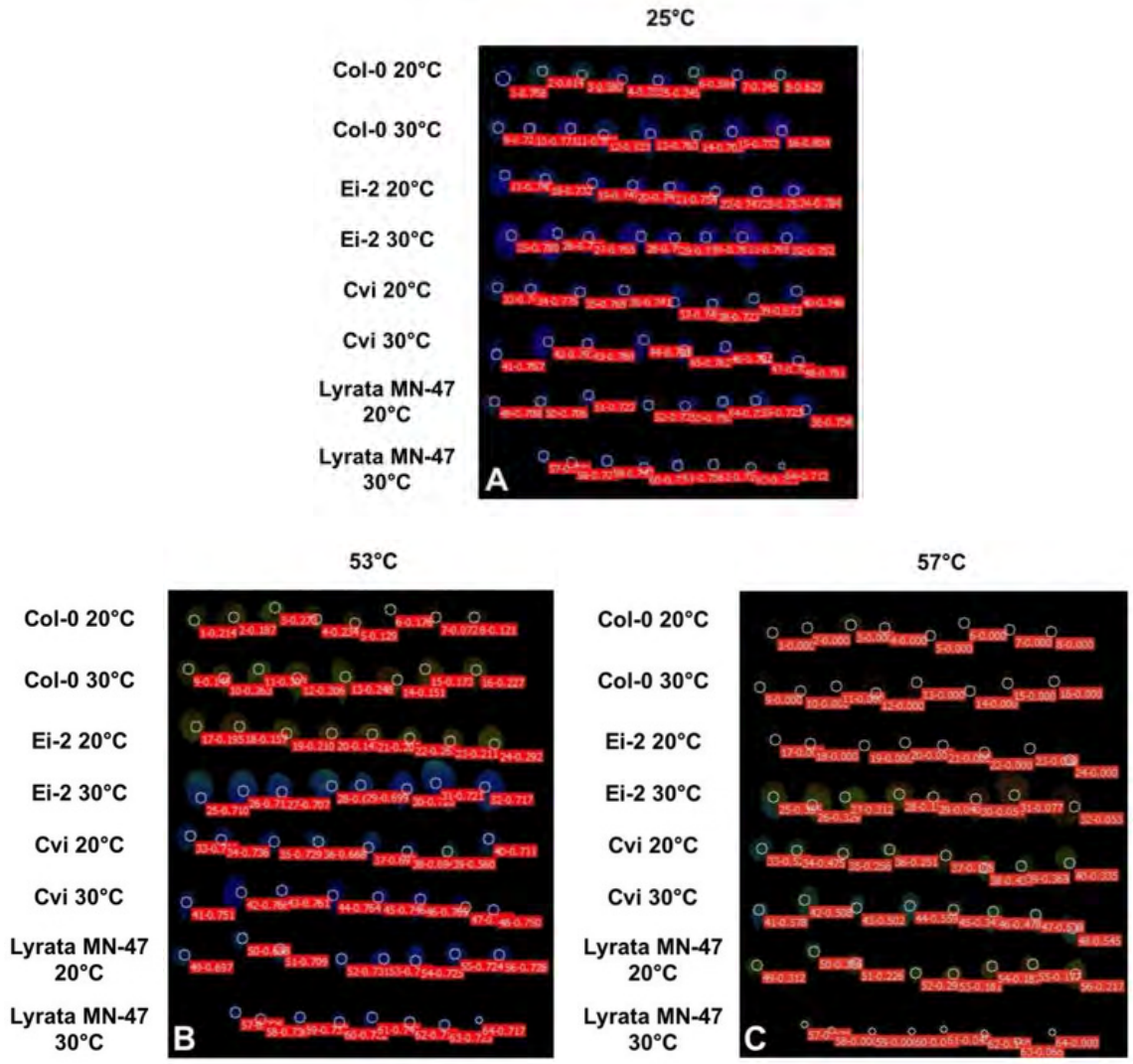


Figure 1. ***Chlorophyll fluorescence images illustrating PSII thermal tolerance in three Arabidopsis thaliana ecotypes (Col-0, Cvi, Ei-2) and Arabidopsis lyrata (MN-47).***

Detachedleavesfrom plantsgrown at 20°C or 30°C were exposed to increasing temperatures, and Fv/Fm was quantified using PAM imaging. Values (in red) represent measurements from circular areas of interest. Each row shows eight leaves for a single line, displayed under their respective growth temperatures. **(A)** Fv/Fm measured at 25°C, representing baseline PSII efficiency prior to high heat exposure. **(B)** Fv/Fm measured after heating leaves to 53°C, where declines in PSII efficiency become apparent in several lines. **(C)** Fv/Fm measured after heating to 57°C, showing severe reductions in Fv/Fm across all lines.

Biochemical trait measurements

Pigment quantification

Pigment concentrations were measured in three-week-old *A. thaliana* ecotypes (Col-0, Cvi, Ei-2) only. One fully expanded leaf was sampled per plant, with four replicates per ecotype. Fresh tissue was weighed and extracted in 1.2 mL of 80% acetone. After centrifugation, absorbance of the supernatant was recorded at 470, 646 and 663 nm.

Chlorophylls were calculated using standard trichromatic equations:

$$ChlA\ (mg\ g^{-1}\ FW) = (12.21(A_{663})-2.81(A_{646})) \times V / (FW \times 1000)$$

$$ChlB\ (mg\ g^{-1}\ FW) = (20.13(A_{646})-5.03(A_{663})) \times V / (FW \times 1000)$$

$$Total\ Chl\ (mg\ g^{-1}\ FW) = (17.32(A_{646})+ 7.18(A_{663})) \times V / (FW \times 1000)$$

Carotenoids were calculated as:

$$Carotenoids\ (mg\ g^{-1}\ FW) = (1000(A_{470}) - 1.28(ChlA) - 56.7(ChlB)) / 245$$

where *A* is absorbance at the indicated wavelength, *V* is extraction volume (mL), and *FW* is fresh weight (g).

Anatomical trait measurements

For anatomical analyses, one fully expanded leaf was sampled from each one-month-old plant, yielding four replicates per line. *A. lyrata* MN-47 was sampled only at 20°C and therefore was not included in the temperature difference comparison.

Leaf mass area

Each leaf was photographed alongside a ruler scale, and leaf area was quantified using ImageJ[10]. Samples were dried at 60°C to constant mass.

LMA was calculated as:

$$LMA = \text{dry mass (g)} / \text{leaf area (cm}^2\text{)}$$

Stomatal density

Stomatal impressions were taken from the abaxial leaf surface using clear nail varnish and transparent tape. Peel impressions were mounted on glass slides and imaged at 10× magnification. Three non-overlapping fields of view were counted per leaf, and stomatal density was expressed as stomata mm^{-2}.

Statistical Analysis

All analyses were performed in R version 4.5.1[11]. For physiological traits (α, ETRmax), the effects of line, temperature and their interaction were tested using two-way ANOVA. For biochemical (chlorophylls, carotenoids) and anatomical traits (LMA, stomatal density), which were measured in *A. thaliana* ecotypes only, the effects of ecotype, temperature and their interaction were tested. Model assumptions were checked using Shapiro–Wilk tests (normality) and Levene's tests (homogeneity). Significant ANOVA terms (α = 0.05) were explored with Tukey's HSD tests.

PSII thermal tolerance (T_{50}) was analysed using a linear mixed-effects model to account for run-to-run variability:

$$T_{50} \sim line \times temperature + (1|Run)$$

Fixed effects were tested using Type III ANOVA, and post-hoc comparisons used Tukey-adjusted estimated marginal means.

Δ-values (30°C − 20°C) were calculated from line means (physiological traits) or ecotype means (biochemical and anatomical traits) for descriptive visualisation only and were not statistically tested.

Results

Physiological traits

Alpha (α)

α was significantly affected by line ($F_{3,98}$ = 12.07, $p < 0.001$) and temperature ($F_{1,98}$ = 21.61, $p < 0.001$), with no line × temperature interaction ($p > 0.05$). At both temperatures, Cvi had significantly higher α than Col-0 ($p < 0.01$), whereas Ei-2 and *A. lyrata* MN-47 did not differ from Col-0 ($p > 0.05$) (Figure 2A). Across lines, α values decreased at 30°C, though individual line-level changes did not reach significance in post-hoc comparisons (Figure 2B).

ETRmax

ETRmax was significantly affected by line ($F_{3,98}$ = 7.12, $p < 0.001$) and temperature ($F_{1,98}$ = 10.20, $p < 0.01$), with no line × temperature interaction ($p > 0.05$). Cvi had significantly higher ETRmax than Col-0 at both temperatures ($p < 0.05$), whereas Ei-2 and *A. lyrata* MN-47 did not differ from Col-0 ($p > 0.05$) (Figure 2C). ETRmax increased at 30°C across lines, though individual line-level changes did not reach significance in post-hoc comparisons (Figure 2D).

$\underline{T_{50}}$

T_{50} was significantly affected by line ($F_{3,148}$ = 64.87, p < 0.001), temperature ($F_{1,148}$ = 13.89, p < 0.001), and their interaction ($F_{3,148}$ = 5.47, p < 0.01). *A. lyrata* MN-47 had significantly higher T_{50} than all other lines at both temperatures (p < 0.001), and Cvi was significantly higher than Col-0 at 20°C (p < 0.001). At 30°C, Ei-2, but not Cvi, was significantly higher than Col-0 (p < 0.001 and p > 0.05, respectively). Ei-2 was the only ecotype to show a significant temperature-dependent increase (p < 0.001) (Figure 2E). Δ-values showed corresponding numerical increases (Figure 2F).

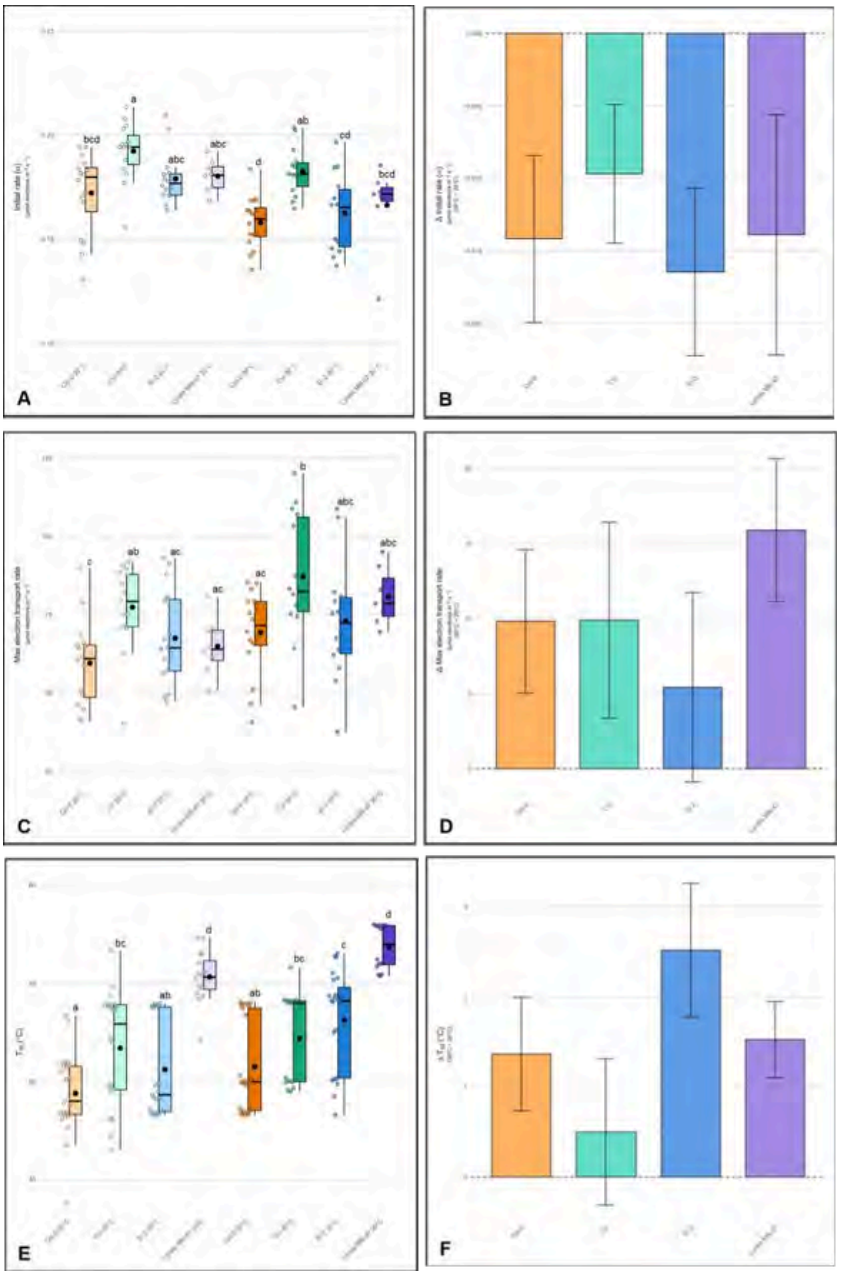


Figure 2. ***Physiological trait responses of three Arabidopsis thaliana ecotypes (Col-0, Cvi, Ei-2) and Arabidopsis lyrata (MN-47).*** Each line is represented by a distinct colour, with lighter pastel shades indicating plants grown at 20°C and darker shades of the same colour indicating plants grown at 30°C; intermediate shades are used in the Δ panels (B, D, F) to represent the temperature difference. **(A)** Initial rate (α). **(C)** Maximal electron transport rate (ETRmax). **(E)** PSII thermal tolerance (T_{50}). Boxplots show trait values for the four lines at each temperature. Boxplots display the mean (black dot), median (horizontal line), interquartile range (box), and data range (whiskers). Raw data points represent individual biological replicates. Significant differences among groups are indicated using compact letter displays (CLDs), where

groups sharing a letter are not significantly different (Tukey's HSD, α = 0.05). **(B)** Δα. **(D)** ΔETRmax. **(F)** ΔT_{50}. Bar plots show Δ-means (30°C − 20°C), with whiskers indicating standard error (SE). Δ-values are for descriptive visualisation only and were not statistically tested.

Biochemical traits

Total chlorophyll concentration

Total chlorophyll was significantly affected by temperature ($F_{1,17} = 7.92$, $p < 0.05$), but not by ecotype or the interaction ($p > 0.05$). No significant pairwise differences occurred among ecotypes at either temperature ($p > 0.05$) (Figure 3A). Δ-values showed lower total chlorophyll at 30°C across ecotypes (Figure 3C).

Chlorophyll a:b ratio

Chlorophyll a:b ratio was significantly affected by ecotype ($F_{2,17} = 7.90$, $p < 0.01$) and by the ecotype × temperature interaction ($F_{2,17} = 11.12$, $p < 0.001$), while temperature alone was not significant ($p > 0.05$). At 20°C, Ei-2 had a significantly lower a:b ratio than Col-0 and Cvi (both $p < 0.01$). No significant pairwise differences were detected at 30°C ($p > 0.05$). Only Cvi showed a significant temperature-dependent decrease ($p < 0.05$). Ratios tended to converge at 30°C, with Cvi showing the clearest numerical decline (Figure 3B).

Carotenoid concentration

Carotenoids were significantly affected by ecotype ($F_{2,17} = 4.03$, $p < 0.05$), but not by temperature or the interaction ($p > 0.05$). No significant pairwise differences occurred at either temperature ($p > 0.05$) (Figure 3D). Carotenoid levels decreased slightly at 30°C for Cvi and Ei-2, with little change in Col-0 (Figure 3E).

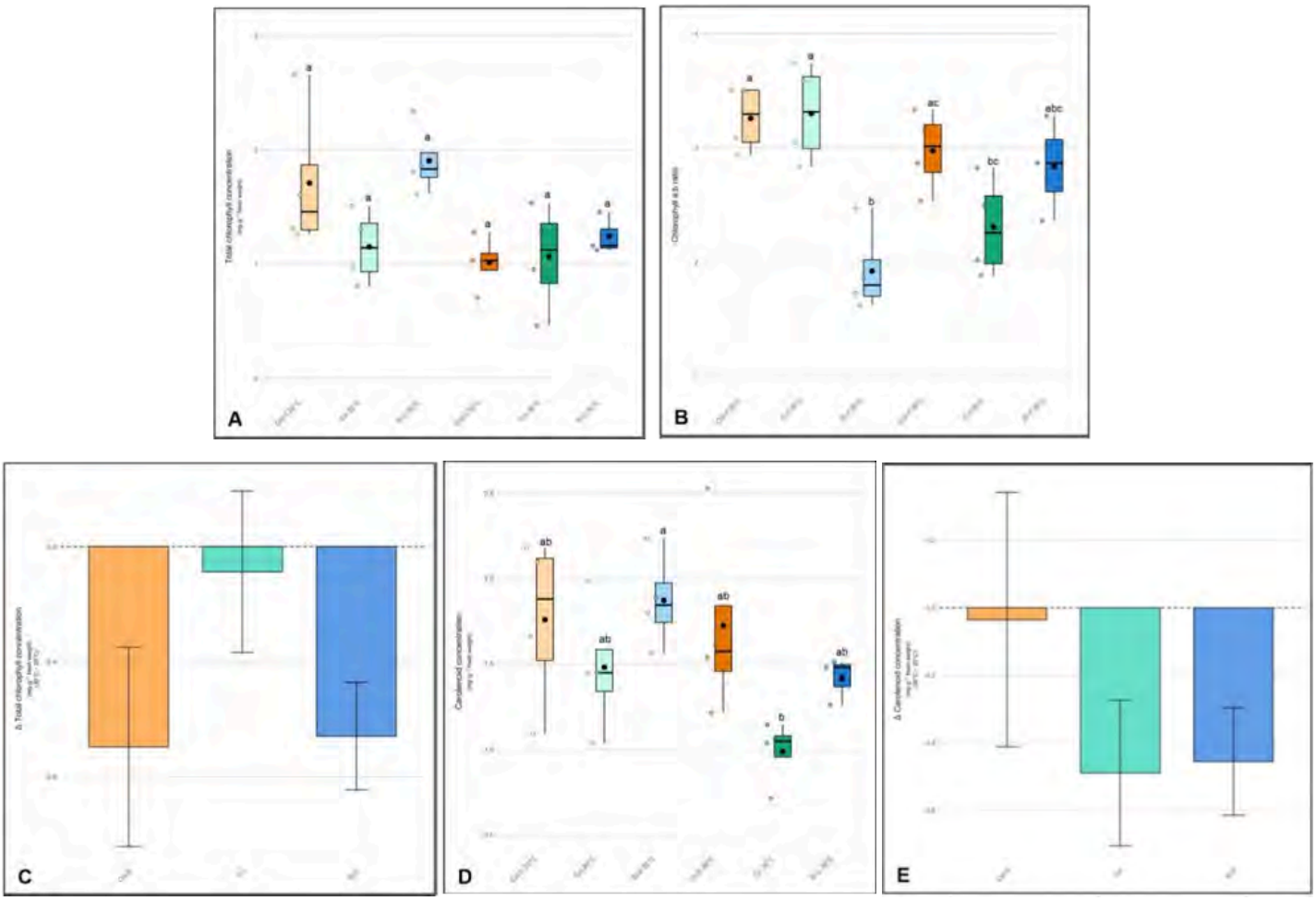


Figure 3. ***Biochemical trait responses of three Arabidopsis thaliana ecotypes (Col-0, Cvi, Ei-2).*** Each ecotype is represented by a distinct colour, with lighter pastel shades indicating plants grown at 20°C and darker shades of the same colour indicating plants grown at 30°C; intermediate shades are used in the Δ panels (C, E) to represent the temperature difference. **(A)** Total chlorophyll concentration. **(B)** Chlorophyll a:b ratio. **(D)** Carotenoid concentration. Boxplots show trait values for the three ecotypes at each temperature. Boxplots display the mean (black dot), median (horizontal line), interquartile range (box), and data range (whiskers). Raw data points represent individual biological replicates. Significant differences among groups are indicated using compact letter displays (CLDs), where groups sharing a letter are not significantly different (Tukey's HSD, $\alpha = 0.05$). **(C)** Δ total chlorophyll concentration. **(E)** Δ carotenoid concentration. Bar plots show Δ-means (30°C − 20°C), with whiskers indicating standard error (SE). Δ-values are for descriptive visualisation only and were not statistically tested.

Anatomical traits

Leaf mass area

LMA was significantly affected by temperature ($F_{1,17}$ = 6.01, p < 0.05), but not by ecotype or the interaction (p > 0.05). No significant pairwise differences occurred among ecotypes (p > 0.05) (Figure 4A). LMA decreased at 30°C for Cvi and Ei-2, with no change in Col-0 (Figure 4B).

Stomatal density

Stomatal density was significantly affected by line ($F_{3,21}$ = 3.24, p < 0.05) and by the line × temperature interaction ($F_{2,21}$ = 6.03, p < 0.01), while temperature alone was not significant (p > 0.05). At 20°C, *A. lyrata* MN-47 had significantly higher stomatal density than Ei-2 (p < 0.05). Although the interaction term indicated line-specific differences in how stomatal density varied with temperature, post-hoc tests did not detect significant temperature effects within any individual ecotype (all p > 0.05). Δ-values showed small, ecotype-specific numerical shifts (Figure 4D).

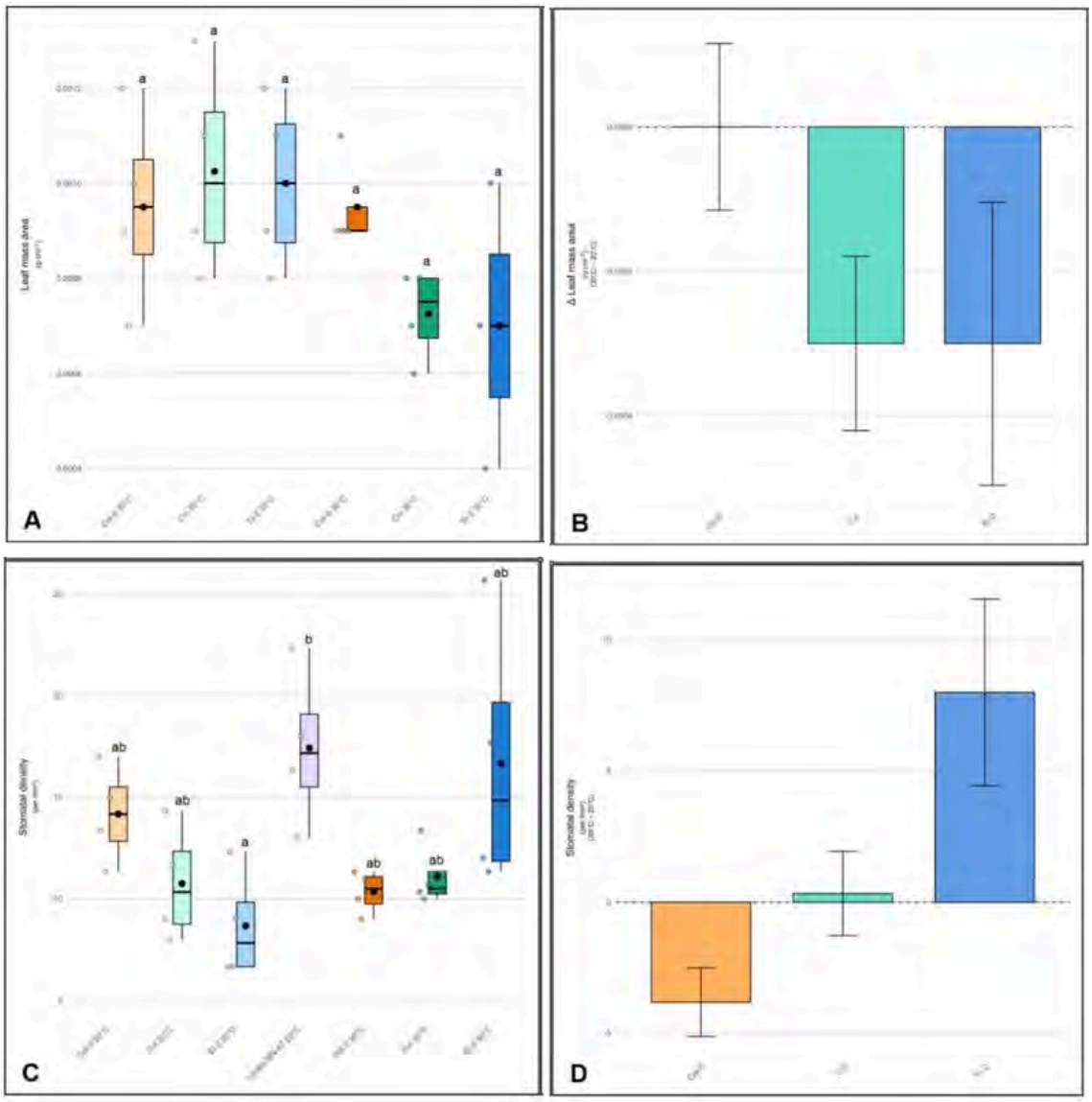


Figure 4. ***Anatomical trait responses of three Arabidopsis thaliana ecotypes (Col-0, Cvi, Ei-2)*****.** Each ecotype is represented by a distinct colour, with lighter pastel shades indicating plants grown at 20°C and darker shades of the same colour indicating plants grown at 30°C; intermediate shades are used in the Δ panels (B, D) to represent the temperature difference. **(A)** Leaf mass per area (LMA). **(C)** Stomatal density. Boxplots show trait values for the three ecotypes at each temperature. Boxplots display the mean (black dot), median (horizontal line), interquartile range (box), and data range (whiskers). Raw data points represent individual biological replicates. Significant differences among groups are indicated using compact letter displays (CLDs), where groups sharing a letter are not significantly different (Tukey's HSD, α = 0.05).

(B) ΔLMA. **(D)** Δ stomatal density. Bar plots show Δ-means (30°C − 20°C), with whiskers indicating standard error (SE). Δ-values are for descriptive visualisation only and were not statistically tested. Note: Stomatal density measurements include *A. lyrata* (MN-47) at 20°C only; *A. lyrata* MN-47 was not measured at 30°C.

Discussion

Overview of findings

This study revealed that the four lines differed in photosynthetic parameters (α, ETRmax) and PSII thermal tolerance (T_{50}), whereas pigment and anatomical traits showed comparatively small or non-significant temperature responses across the *Arabidopsis* lines studied. Since all leaves measured developed fully within their assigned treatment, the observed patterns represented genuine responses to the warming regime rather than artefacts of developmental history.

Physiological responses

α and ETRmax differed consistently among ecotypes, with Cvi exhibiting higher photosynthetic capacity than Col-0 at both temperatures. Flood *et al*.[2] reviewed evidence for natural genetic variation in photosynthetic traits across *A. thaliana* and crop species, demonstrating that such variation is widespread; however, the genetic factors underlying these differences remain largely unknown. Prinzenberg *et al*.[12] measured photosynthetic efficiency (ΦPSII) across a diverse panel of *A. thaliana* accessions under low-temperature conditions, finding substantial variation in cold-dependent photosynthetic responses and identifying associated quantitative trait loci; notably, some accessions from colder climates showed relatively maintained photosynthetic efficiency under cold, suggesting a degree of local photosynthetic adaptation. The present study extends this to elevated temperatures, finding that constitutive differences in α and ETRmax among ecotypes persisted across both growth temperatures without a temperature-dependent shift in rank order.

The higher photosynthetic capacity of Cvi is consistent with its Cape Verde Islands origin, where high light and arid conditions may have favoured elevated electron transport rates. In contrast, Ei-2 and Col-0, from cooler European and temperate North American climates respectively, showed comparatively lower α and ETRmax. Warming to 30°C produced only small shifts in these traits, indicating that linear electron transport remained stable under moderate heat. Yamori *et al*.[8] reviewed temperature responses of photosynthesis across C3, C4 and CAM plants, demonstrating that photosynthetic electron transport capacity generally remains stable over moderate temperature ranges, with limitations becoming more apparent as temperatures approach and exceed the species-specific photosynthetic optimum. Baker[1] similarly notes that PSII efficiency is relatively robust under moderate warming, with thermal damage becoming acute only when sustained temperatures exceed the thermal stability of PSII protein complexes. That neither α nor ETRmax showed significant line × temperature interactions in the present study supports this view: moderate warming of 10°C was sufficient to reveal variation in PSII thermal tolerance (T_{50}) but not to differentially disrupt upstream electron transport capacity across lines.

In contrast, T_{50} revealed clear differences in heat tolerance among the four lines. MN-47 consistently exhibited the highest T_{50} values across temperatures, indicating that its PSII complexes withstand higher temperatures than those of the *A. thaliana* ecotypes. This is consistent with Wos and Willi[3], who investigated thermal acclimation in *A. lyrata* through repeated heat pre-exposure, identifying differentially expressed genes under heat treatment and finding that the magnitude of inducible heat resistance varied among genotypes. The present study cannot directly determine whether MN-47's constitutive PSII stability reflects a similar inducible mechanism, as transcriptomic data were not collected, but MN-47's consistently elevated T_{50} (rather than an acclimatory increase specifically at 30°C) suggests constitutive PSII stability rather than inducible acclimation, consistent with a perennial life history strategy in which maintaining photosynthetic function across variable seasons confers a selective advantage. Ferguson *et al*.[9], who developed T_{50} as a screening metric for variation in PSII heat tolerance in rice, similarly found that T_{50} captured meaningful between-line variation not reflected in baseline

photosynthetic parameters, mirroring the disconnect between T_{50} and α or ETRmax observed here.

Ei-2 was the only ecotype to increase its T_{50} at 30°C, demonstrating short-term acclimation of PSII thermal tolerance. This is particularly notable given Ei-2's origin in the cooler Eifel highlands, which might not immediately predict a capacity for upward thermal acclimation; this may instead reflect phenotypic plasticity as a strategy for coping with unpredictable temperature fluctuations, a trait that has been documented in other cool- origin *A. thaliana* ecotypes[2]. Cvi showed no acclimatory shift despite its high α and ETRmax, an unexpected finding given its warm, arid geographic origin, which might otherwise suggest prior adaptation to heat. Col-0 maintained the lowest thermal tolerance, consistent with the expectation that extended laboratory cultivation may have reduced selection pressure on heat resilience traits. Since these differences occurred independently of changes in photosynthetic capacity, the results reinforce that performance under non-stressful conditions does not reliably predict heat resilience[7]. Ei- 2's acclimatory improvement appears to arise from mechanisms acting directly on PSII stability rather than broader photosynthetic remodelling.

Biochemical responses

Pigment traits showed limited sensitivity to warming. Total chlorophyll content declined modestly at 30°C in all ecotypes. Bita and Gerats[7] reviewed heat stress responses in plants and noted that chlorophyll degradation under warming is a commonly observed but highly context-dependent response, with the magnitude of loss varying with temperature severity, duration, and species. The modest, statistically significant decline observed here (without significant ecotype-specific differences) suggests that 30°C represents a threshold at which mild pigment degradation begins uniformly rather than in an ecotype- specific manner. This contrasts with the strong ecotype-level differentiation seen in T_{50}, indicating that pigment loss and PSII thermal stability are decoupled under the moderate warming regime applied here. Carotenoid concentrations differed among ecotypes but

were unaffected by temperature, indicating that bulk photoprotective pigment pools were maintained across treatments.

The chlorophyll a:b ratio was the only pigment variable showing an ecotype-specific response. Ei-2 exhibited a significantly lower ratio at 20°C than Col-0 and Cvi, and only Cvi showed a warming-induced decrease. A reduced ratio may reflect adjustments in antenna size or composition, but neither baseline differences nor temperature-related changes corresponded to T_{50} patterns. Since pigment data were not collected for *A. lyrata* MN-47, no conclusions can be drawn regarding how pigment composition relates to its high PSII thermal tolerance. Thus, pigment profiles did not explain the contrasting physiological responses observed among the *A. thaliana* ecotypes.

Anatomical responses

Anatomical traits were likewise stable under warming. LMA declined slightly at 30°C across ecotypes. Poorter *et al*. [5]) conducted a large meta-analysis of LMA variation across species and environments, finding that LMA varies strongly with temperature and that higher temperatures are generally associated with lower LMA; however, the paper notes that the physiological regulation underlying this response remains unclear. The direction of the effect observed here is consistent with their findings, though the magnitude was small and did not differ among ecotypes, suggesting that the 10°C warming increment was sufficient to produce the expected structural response but not to reveal ecotype-level differences in structural plasticity. The modest effect size may reflect the relatively limited temperature range and short developmental timescale of the present study. Stomatal density differed among ecotypes but showed no meaningful or consistent temperature-dependent shifts. Although MN-47 exhibited higher stomatal density than Ei-2 at 20°C, anatomical traits were not measured for MN-47 at 30°C, preventing assessment of its temperature response. The limited anatomical dataset therefore does not support a structural basis for line-level differences in T_{50}.

Since the leaves analysed were the youngest fully expanded leaves and therefore developed entirely within their temperature treatments, these findings reflected genuine structural stability under moderate warming. This contrasts with the strong ecotype-dependent patterns observed in PSII thermal tolerance.

Integration and limitations

Across all traits measured, PSII thermal tolerance was the trait most strongly shaped by line and temperature. MN-47's consistently high T_{50} and Ei-2's unique acclimatory increase occurred without accompanying changes in pigments, anatomy, or photosynthetic capacity, suggesting that the mechanisms responsible are thermally specific rather than part of a broader physiological remodelling. That these responses align, at least partially, with the geographic origins and life history strategies of each line lends ecological plausibility to the findings, though direct causal links between climate of origin and thermal tolerance require further genetic dissection. The absence of pigment data for MN-47 and the lack of anatomical measurements for MN-47 at 30°C limit direct comparisons between this species and the *A. thaliana* ecotypes. Nonetheless, the available data still indicate that mechanisms underpinning thermal tolerance likely lie outside the biochemical and leaf-structural traits assessed here, potentially involving differences in PSII repair dynamics, thylakoid membrane stability, or early activation of heat-responsive transcriptional pathways[4]).

The main limitation of the study was the low replication for pigment and anatomical traits (n = 4) and incomplete sampling of MN-47 for these measures, both of which reduce power to detect subtle or species-specific effects. However, the absence of strong or directional trends across *A. thaliana* ecotypes suggests that structural and pigment traits remained comparatively stable under the conditions applied. Moderate warming may therefore expose variation in PSII stability before eliciting broader physiological adjustments, helping explain why T_{50} responded strongly while other traits remained relatively unchanged.

Conclusion

This study showed that warming revealed clear phenotypic differences in PSII thermal tolerance among lines, while photosynthetic capacity, pigment composition, and leaf structural traits remained comparatively stable. MN-47 maintained the highest T_{50}, and Ei-2 was the only ecotype to display an acclimatory increase, yet these responses occurred without accompanying shifts in electron transport, pigments, or anatomy. These patterns indicate that the processes determining thermal resilience lie beyond the traits measured here and are likely driven by molecular mechanisms acting directly on PSII function. Future work combining gas-exchange measurements with RNA-seq would help identify the physiological and regulatory pathways that underpin these contrasting thermal responses.

As global temperatures continue to rise, understanding which traits and mechanisms confer heat resilience becomes increasingly urgent. The variation in PSII thermal tolerance documented here, across lines sourced from habitats ranging from the Cape Verde Islands to Minnesota, suggests that natural plant populations already harbour meaningful diversity in their capacity to withstand warming. Investigating this variation in *Arabidopsis*, as a genetically tractable system, provides a foundation for identifying molecular targets that could be leveraged in breeding programmes aimed at improving crop thermal resilience. Whether these strategies represent more broadly transferable mechanisms remains an open question with direct relevance to predicting how plant populations will fare under projected climate scenarios.

References

1. Baker NR. Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo. *Annu Rev Plant Biol*. 2008;59(1):89-113. doi:10.1146/annurev.arplant.59.032607.092759

2. Flood PJ, Harbinson J, Aarts MGM. Natural genetic variation in plant photosynthesis. *Trends Plant Sci*. 2011;16(6):327-335. doi:10.1016/j.tplants.2011.02.005

3. Wos G, Willi Y. Thermal acclimation in *Arabidopsis lyrata*: genotypic costs and transcriptional changes. *J Evol Biol*. 2018;31(1):123-135. doi:10.1111/jeb.13208

4. Allakhverdiev SI, Kreslavski VD, Klimov VV, Los DA, Carpentier R, Mohanty P. Heat stress: an overview of molecular responses in photosynthesis. *Photosynth Res*. 2008;98(1-3):541-550. doi:10.1007/s11120-008-9331-0

5. Poorter H, Niinemets Ü, Poorter L, Wright IJ, Villar R. Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New Phytol*. 2009;182(3):565-588. doi:10.1111/j.1469-8137.2009.02830.x

6. Delgado D, Alonso-Blanco C, Fenoll C, Mena M. Natural variation in stomatal abundance of *Arabidopsis thaliana* includes cryptic diversity for different developmental processes. *Ann Bot*. 2011;107(8):1247-1258. doi:10.1093/aob/mcr060

7. Bita CE, Gerats T. Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Front Plant Sci*. 2013;4. doi:10.3389/fpls.2013.00273

8. Yamori W, Hikosaka K, Way DA. Temperature response of photosynthesis in C3, C4, and CAM plants: temperature acclimation and temperature adaptation. *Photosynth Res*. 2014;119(1-2):101-117. doi:10.1007/s11120-013-9874-6

9. Ferguson JN, McAusland L, Smith KE, Price AH, Wilson ZA, Murchie EH. Rapid temperature responses of photosystem II efficiency forecast genotypic variation in rice vegetative heat tolerance. *Plant J*. 2020;104(3):839-855. doi:10.1111/tpj.14956

10. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676-682. doi:10.1038/nmeth.2019

11. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2025; https://www.R-project.org

12. Prinzenberg AE, Campos-Dominguez L, Kruijer W, Harbinson J, Aarts MGM. Natural variation of photosynthetic efficiency in *Arabidopsis thaliana* accessions under low temperature conditions. *Plant Cell Environ*. 2020;43(8):2000-2013. doi:10.1111/pce.13811

Author's Bio:

Hi, I'm Danica,a final-year Biology student. Growing up across Southeast Asia sparked my interest inthe challenges facing agriculture in a changing climate. During my degree, I became interested in how crops respond to environmental pressures and what determines agricultural resilience. A summer studentship at the John Innes Centre introduced me to research on plant–insect interactions in oilseed rape, while my final-year project investigated heat tolerance in Arabidopsis. This summer, I will be studying how drought affects growth and reproduction in peas. I hope to continue into postgraduate research on crop responses to environmental change and their implications for future food systems.

From Statistical Associations to Ecological Interactions: Rethinking Microbial Co-occurrence Networks in Soil Ecosystems

Author: Yuxuan Liao

ORCID: 0009-0009-1562-1049

Programme: MSc Business Analytics

Affiliation: Universityof Bristol Business School

Email: nc25042@bristol.ac.uk

Abstract

Microbial co-occurrence networks are now widely used tools for exploring potential interactions within soil microbial communities. These networks are typically inferred from sequencing data using statistical associations such as correlation or conditional dependence. However, a persistent challenge lies in the interpretation of these associations, as statistical co-occurrence does not necessarily imply direct ecological interaction. This review examines the conceptual gap between association-based network inference and the ecological interactions that shape soil microbial communities. Commonly used approaches for constructing microbial co-occurrence networks are first summarized, and the assumptions underlying these methods are critically examined. Key limitations that complicate ecological interpretation, including compositional effects, indirect associations, and environmental filtering, are then analyzed. Finally, emerging strategies for improving the ecological relevance of inferred networks are highlighted, including the integration of experimental validation, multi-omics data, and dynamical modeling frameworks. By clarifying the relationship between statistical association and ecological interaction, this review aims to provide

a clearer framework for interpreting microbial network analyses and to guide future research toward more mechanistic understanding of soil microbial community organization.

Keywords: Microbial co-occurrence networks, Ecological interactions, Network inference, Compositional effects, Multi-omics integration

Introduction

Microbial communities, defined here as groups of microorganisms that coexist and function within a shared environment, are fundamental to terrestrial ecosystems, particularly in soil. They play critical roles in nutrient cycling, including the transformation and recycling of carbon and nitrogen, as well as in the maintenance of broader ecosystem processes.[1-2] In soil ecosystems, microbial communities also contribute to organic matter decomposition, soil fertility, plant productivity, and the long-term sustainability of soil management.[3-4] These processes are closely linked to soil health, ecosystem service delivery, and sustainable agricultural production.[3] Understanding how soil microorganisms are organized and how they respond to environmental change is therefore important for predicting ecosystem stability and the effects of management interventions.[3-4] Because many soil functions emerge from relationships among microbial taxa rather than from individual organisms alone, understanding microbial interactions is essential for explaining community stability and ecosystem functioning.[1-2,5] However, identifying direct ecological interactions, that is, biological relationships in which one microorganism affects the growth, survival, or activity of another, remains challenging within complex and highly diverse soil microbial communities.[2]

Recent advances in high-throughput sequencing, a set of technologies that generate large volumes of microbial DNA or RNA sequence data, have enabled comprehensive characterization of soil microbial community composition, meaning which microbial taxa are present and how abundant they are within a sample.[1] Using these datasets, researchers have increasingly employed microbial co-occurrence networks, graph-based representations in which nodes represent microbial taxa and edges represent statistically inferred associations between taxa across samples, to infer potential ecological associations, or possible relationships suggested by patterns of co-variation, among soil microorganisms.[6-7] A central challenge, however, is that statistical associations, defined as patterns of co-variation in taxon abundances detected by statistical methods, identified in these networks do not necessarily

represent direct ecological interactions.[2,8] One important reason is the compositional nature of high-throughput sequencing data. Compositionality means that sequencing data usually describe the proportional, or relative, abundance of each taxon within a sample rather than its absolute cell count. As a result, an apparent increase in one taxon may mathematically reduce the relative abundance of others, generating spurious correlations, or statistical relationships that do not reflect real biological associations, when standard correlation measures are applied directly.[9,10] Despite these limitations, co-occurrence network analysis remains widely applied to explore microbial associations across diverse soil environments.[11]

In current soil microbial network studies, inferred co-occurrence edges are sometimes interpreted in explicitly ecological terms. Positive correlations may be described as cooperative or mutualistic relationships, whereas negative correlations may be interpreted as competition or exclusion.[12-13] For example, in a recent study of saline soils, the greater abundance of positive than negative links was interpreted as indicating that cooperative relationships were stronger than competitive relationships among bacterial communities.[14] Similar ecological language also appears in wetland soil microbial network studies, where co-occurrence links are discussed as positive or negative interactions among microbial nodes.[15] However, such interpretations remain hypotheses unless supported by experimental validation, because similar co-occurrence patterns can also arise from shared environmental preferences, dispersal limitation, indirect associations, or compositional artefacts.[12,16] Over-interpreting network edges may therefore misdirect subsequent mechanistic research, lead to inefficient use of experimental resources, and distort understanding of how microbial communities contribute to soil ecosystem functions.[12,16]

This review critically examines the interpretation of microbial co-occurrence networks in soil ecosystems. It summarizes methods for network construction, discusses key limitations, and explores emerging strategies for improving the ecological interpretation of microbial networks. By integrating these perspectives, this review aims to provide a clearer conceptual framework for interpreting microbial network

analyses and to guide future research toward a more mechanistic understanding of soil microbial community organization.

Methods for constructing microbial co-occurrence networks

Microbial co-occurrence networks are typically constructed from microbial community abundance data derived from high-throughput sequencing.[6] These datasets contain abundance profiles of microbial taxa across multiple soil samples, with taxa commonly represented as operational taxonomic units (OTUs) or amplicon sequence variants (ASVs).[17] Most high-throughput sequencing datasets provide relative abundances, meaning that microbial taxa are represented as proportions within each sample rather than as absolute cell counts. Absolute abundance data can be estimated using approaches such as quantitative PCR (qPCR) or spike-in standards, but these methods involve additional technical requirements, costs, and potential measurement biases. Consequently, relative abundance analysis remains widely used in microbial co-occurrence network studies because of its simplicity, scalability, and lower cost.

Abundance data serves as the basis for inferring co-occurrence networks through statistical association analysis. The general workflow is illustrated in Figure 1.

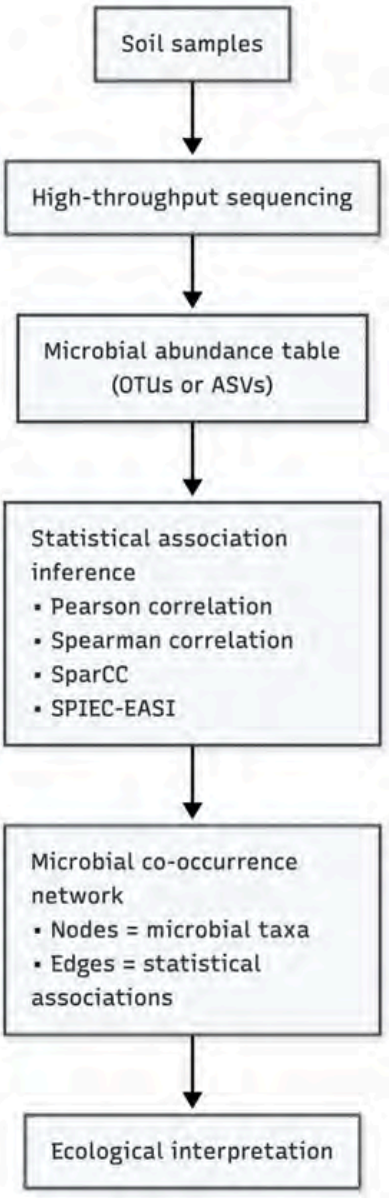


Figure 1. ***General workflow for inferring microbial co-occurrence networks from sequencing data.*** The workflow comprises three main steps: (i) generation of abundance tables; (ii) statistical inference of associations (see main text for method details); (iii) network construction where nodes represent taxa and edges represent inferred associations.

In microbial networks, each node typically represents one taxon identified in the dataset, whereas edges represent associations inferred from statistical relationships in abundance patterns across samples.[8,18]

The most widely used approach is statistical correlation analysis.[6] Commonly used measures include Pearson and Spearman correlations: Pearson is typically used to detect linear relationships between microbial taxa, whereas Spearman measures rank-based relationships and is less sensitive to outliers.[19] Due to their simplicity correlation-based methods remain widely used in exploratory studies.

However, correlation-based methods may capture indirect associations or be affected by the compositional nature of sequencing data.[9-10] Consequently, more advanced approaches have been developed for microbial network inference to account for these issues. Partial correlation methods attempt to identify potential direct associations by controlling for the influence of other taxa.[20] Bayesian network models provide a probabilistic framework for inferring conditional dependencies among microbial taxa, allowing the representation of joint probability distributions and the distinction between direct and indirect associations while explicitly accounting for uncertainty in high-dimensional data.[21]

Methods such as SparCC were designed to estimate microbial associations while accounting for the compositionality of microbiome data.[22] Similarly, SPIEC-EASI aims to infer more direct associations by using graphical modeling approaches that reduce the impact of indirect associations.[20] Several computational tools facilitate microbial network inference, including SparCC[22], SPIEC-EASI[20], and CoNet[23]. However, different inference methods rely on distinct statistical assumptions and may produce substantially different network structures, highlighting the importance of carefully selecting appropriate approaches.[19,22] Common statistical methods used are summarized in Table 1.

Table 1. Common statistical methods for inferring microbial co-occurrence networks

Method	Statistical principle	Strengths	Limitations	Example Implementations
Pearson Correlation	Measures linear correlation between microbial taxa based on abundance patterns across samples	Simple to implement and widely used in microbial network studies	Sensitive to compositional effects and indirect associations; assumes linear relationships	CoNet
Spearman Correlation	Rank-based correlation that measures monotonic relationships between taxa abundances	More robust to outliers and non-normal distributions	Still affected by compositionality and indirect associations	CoNet
SparCC	Estimates correlations from compositional microbiome data using log-ratio transformations and sparse correlation assumptions	Designed to account for compositional constraints in sequencing data	Assumes sparse correlations and may require large sample sizes	SparCC
SPIEC-EASI	Uses sparse inverse covariance estimation and graphical models to infer conditional dependencies among taxa	Reduces indirect associations and identifies potential direct interactions	Computationally intensive and sensitive to parameter settings	SPIEC-EASI
Bayesian Networks	Probabilistic graphical models that infer conditional dependencies among microbial taxa	Can capture complex probabilistic relationships	Requires large datasets and strong modeling assumptions	bnlearn

Table 1 summarises widely used statistical approaches for microbial network inference, including correlation-based methods (Pearson and Spearman), compositionality-aware methods (SparCC), graphical model-based approaches (SPIEC-EASI), and probabilistic models (Bayesian networks). Consequently, the choice of method should be guided by the characteristics of the data (e.g., compositionality, sparsity, and sample size) as well as the specific research questions being addressed. No single method can be considered universally optimal, and careful evaluation of methodological assumptions is essential for the robust interpretation of microbial co-occurrence networks.

The conceptual gap: statistical vs. ecological interactions

A major challenge in interpreting microbial co-occurrence networks is the conceptual gap between statistical associations and ecological interactions.[2] Statistical associations describe patterns of co-variation in microbial taxon abundances across samples.[6,8] By contrast, ecological interactions refer to relationships through which one taxon directly or indirectly affects another.[18] Because similar abundance patterns can arise from multiple ecological or statistical processes, a network edge does not necessarily indicate a true ecological interaction.[2] The conceptual relationship between statistical associations and ecological interactions is illustrated in Figure 2.

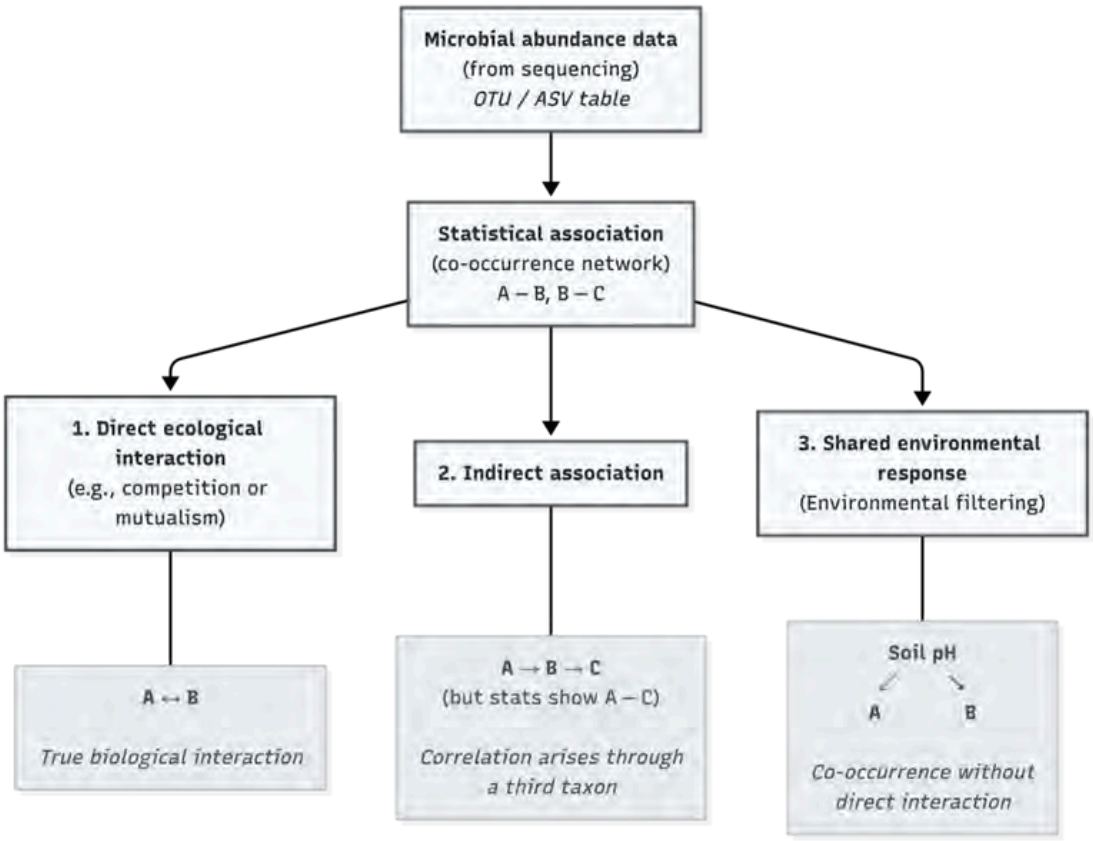


Figure 2. ***Conceptual differences between statistical associations and ecological interactions.*** The upper panel shows how abundance data generate statistical associations. The lower panels contrast direct interactions (left) with alternative mechanisms: indirect associations (middle) and environmental filtering (right). Compositional effects can further produce spurious correlations.

Together, these mechanisms demonstrate that observed correlations in microbial co-occurrence networks may arise from multiple processes rather than direct biological interactions, highlighting that statistical associations do not necessarily reflect true ecological relationships.

One important source of misinterpretation is the compositional nature of microbiome sequencing data.[9-10] Sequencing-based datasets usually provide relative rather than absolute abundances, meaning that taxon abundances are constrained within each sample.[10] As a result, an increase in the relative abundance of one taxon can artificially decrease the observed abundances of others, even when no direct

interaction exists. This can generate spurious correlations when standard correlation methods are applied.[19]

A second limitation is that many statistically inferred associations may be indirect rather than direct.[19-20] Two taxa may appear correlated because both are linked to a third taxon, rather than because they interact directly. For example, if taxon A influences taxon B, and B influences taxon C, correlation-based methods may incorrectly infer a direct A–C association (each character represent a different organism).

Environmental factors further complicate the ecological interpretation.[6,8] In soil ecosystems, variables such as pH, temperature, moisture, and nutrient availability strongly shape microbial community composition. Taxa that respond similarly to the same environmental conditions may therefore co-occur, even in the absence of direct biological interactions.[7-8] Such environmentally driven co-occurrence patterns may be mistakenly interpreted as evidence of direct microbial interactions.

In addition, most co-occurrence networks are inferred from cross-sectional data and therefore provide only static representations of community structure.[2,7] Such static networks cannot fully capture temporal dynamics, context-dependent interactions, or changes in relationships across environmental conditions. Longitudinal sampling and time-series analysis can partly address this limitation by tracking temporal changes in microbial community composition and inferred associations.[24-25] For example, recent plant–soil associated microbiome research has used longitudinal sampling to examine temporal changes in microbial associations.[26] However, strictly soil-focused longitudinal co-occurrence network studies remain relatively scarce, representing a critical gap for improving the ecological interpretation of microbial networks.

A final limitation is the lack of direct experimental validation for many inferred associations.[2,6] Because most co-occurrence networks are based on statistical inference, predicted associations should be treated as hypotheses rather than confirmed interactions. Without complementary approaches such as co-culture

experiments, perturbation studies, or multi-omics analyses, it is difficult to determine the biological meaning of individual network edges.

These challenges persist largely because microbiome sequencing datasets are high-dimensional and compositional, where taxon abundances are represented as relative rather than absolute values, a structure that introduces statistical dependencies and complicates the inference of ecological interactions.[10,19]

Emerging strategies for enhancing ecological interpretability

Several strategies have been developed to improve the ecological interpretability of microbial co-occurrence networks. These strategies include experimental validation, multi-omics integration, dynamic modeling approaches, machine learning techniques, and improved benchmarking frameworks.[2,5,7]

Experimental validation directly tests whether predicted associations correspond to biological interactions. Co-culture experiments, microbial manipulation, and perturbation studies can assess whether one microbial taxon affects another.[2,6] These approaches provide evidence that correlation-based network analyses alone cannot supply.

Multi-omics datasets, including metagenomics, metatranscriptomics, metabolomics, and proteomics, provide complementary information about microbial functions and activities.[5] Such integrative approaches help move microbial network studies from descriptive correlations toward functional interpretation.[2,5,7] By linking microbial taxa with functional genes, metabolic pathways, and metabolic products, multi-omics analyses can provide stronger evidence for potential ecological interactions.

Dynamic modeling approaches also offer opportunities to overcome the limitations of static co-occurrence networks and infer microbial interactions more mechanistically.[2,18] Unlike static networks, dynamic models incorporate temporal changes in taxon abundances. For instance, generalized Lotka-Volterra models can

infer interaction strengths from time-series microbial community data.[2] These models provide insights into how microbial populations influence each other's growth over time.

Machine learning techniques are increasingly applied to detect complex and potentially nonlinear interaction patterns that may not be captured by traditional correlation-based methods.[27] For example, random forests and neural networks can identify complex patterns within large and high-dimensional microbiome datasets.[27] Such approaches may improve the prediction of microbial interactions in highly complex ecosystems.[6,19] However, machine learning approaches typically identify statistical patterns rather than direct ecological mechanisms and therefore should complement, rather than replace, experimental and ecological validation.[2] Major strategies for improving ecological interpretation are summarized in Table 2.

Table 2. Emerging strategies for enhancing the ecological interpretability of microbial co-occurrence networks.

Strategy	Description	Example approaches	Key advantage
Experimental validation	Direct testing of predicted microbial interactions inferred from co-occurrence networks	Co-culture experiments, perturbation studies, microbial manipulation	Confirms whether statistically inferred associations correspond to real ecological interactions
Multi-omics integration	Integration of multiple molecular datasets to link microbial taxa with functional activities	Metagenomics, metatranscriptomics, metabolomics, proteomics	Provides functional context and helps interpret potential interactions
Dynamic modeling	Mathematical models that incorporate temporal changes in microbial populations	Generalized Lotka–Volterra models, time-series modeling	Captures dynamic ecological relationships and interaction strengths
Machine learning approaches	Data-driven algorithms used to detect complex patterns in high-dimensional microbiome data	Random forests, neural networks	Identifies nonlinear relationships and complex interaction patterns
Benchmarking and standardization	Development of standardized datasets and evaluation frameworks for network inference methods	Synthetic datasets, standardized pipelines	Improves reproducibility and comparability across studies

Table 2 shows that these strategies are complementary rather than interchangeable. Their combined value lies in providing experimental, functional, temporal, and reproducibility support for ecological interpretations that cannot be justified by correlation-based networks alone.

Improved data standardization and method benchmarking are also important for enhancing the reliability and reproducibility of microbial network analyses.[2,5,19] Differences in sequencing protocols, data processing pipelines, and network inference methods can strongly influence network structure. Developing standardized datasets and evaluation frameworks can help compare different network inference approaches. Such efforts will improve the reproducibility and ecological interpretability of microbial network studies.[2,5,19]

Conclusion

Microbial co-occurrence networks have become valuable tools for describing community structure, but their ecological interpretation is limited: statistical associations do not necessarily represent ecological interactions.[2,7] Despite these limitations, co-occurrence networks remain useful for generating hypotheses, identifying candidate interactions for further study, and providing an overview of community organization.[5-6,8] However, associations derived from sequencing data should not be directly interpreted as true ecological interactions because of compositional effects, indirect associations, and environmental influences.[10,19-20] Recognizing these limitations is essential for responsible interpretation of microbial network analyses and for avoiding over-interpretation of correlation-based findings.

Improving the ecological interpretability of microbial networks will require integrating complementary approaches such as experimental validation, multi-omics integration, and dynamic modeling.[2,5,7] Together, these approaches provide functional and temporal context enabling more mechanistic inference. Moving beyond descriptive correlations toward mechanistic understanding of microbial interactions in soil ecosystems will require integrating time-series data, experimental validation and predictive modeling.[2,6] Future research will benefit from integrating statistical network inference with experimental and ecological approaches that more directly capture microbial interactions in natural environments. Although this review focuses on soil

ecosystems, the challenge of interpreting statistical associations as ecological interactions represents a broader issue in microbial ecology, extending to marine and host-associated microbiomes.[2,6-7] The scarcity of longitudinal soil microbial network studies remains a critical gap, as time-series approaches are needed to examine whether inferred associations are stable, transient, or context-dependent over time.[24-25]

Reference List

1. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol. 2017;15(10):579-590.

2. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J. 2016;10(11):2557-2568.

3. Bender SF, Wagg C, van der Heijden MGA. An underground revolution: biodiversity and soil ecological engineering for agricultural sustainability. Trends Ecol Evol. 2016;31(6):440-452. doi:10.1016/j.tree.2016.02.016.

4. Bach EM, Ramirez KS, Fraser TD, Wall DH. Soil biodiversity integrates solutions for a sustainable future. Sustainability. 2020;12(7):2662. doi:10.3390/su12072662.

5. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol. 2018;16(9):567-576.

6. Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol. 2012;10(8):538-550.

7. Röttjers L, Faust K. From hairballs to hypotheses—biological insights from microbial networks. FEMS Microbiol Rev. 2018;42(6):761-780.

8. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol. 2014;5:219. doi:10.3389/fmicb.2014.00219.

9. Aitchison J. The statistical analysis of compositional data. J R Stat Soc Ser B Methodol. 1982;44(2):139-177.

10. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224.

11. Barberán A, Bates ST, Casamayor EO, Fierer N. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J.

2012;6(2):343-351.

12. Goberna M, Verdú M. Cautionary notes on the use of co-occurrence networks in soil ecology. Soil Biol Biochem. 2022;166:108534. doi:10.1016/j.soilbio.2021.108534.

13. Lee KK, Kim H, Lee YH. Cross-kingdom co-occurrence networks in the plant microbiome: importance and ecological interpretations. Front Microbiol. 2022;13:953300. doi:10.3389/fmicb.2022.953300.

14. Wei Y, Chen L, Feng Q, Xi H, Zhang C, Gan K, et al. Structure and assembly mechanism of soil bacterial community under different soil salt intensities in arid and semiarid regions. Ecol Indic. 2024;158:111631. doi:10.1016/j.ecolind.2024.111631.

15. Wu D, Bai H, Zhao C, Peng M, Chi Q, Dai Y, et al. The characteristics of soil microbial co-occurrence networks across a high-latitude forested wetland ecotone in China. Front Microbiol. 2023;14:1160683. doi:10.3389/fmicb.2023.1160683.

16. Guseva K, Darcy S, Simon E, Alteio LV, Montesinos-Navarro A, Kaiser C. From diversity to complexity: microbial networks in soils. Soil Biol Biochem. 2022;169:108604. doi:10.1016/j.soilbio.2022.108604.

17. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2017;13(7):581-583.

18. Proulx SR, Promislow DEL, Phillips PC. Network thinking in ecology and evolution. Trends Ecol Evol. 2005;20(6):345-353.

19. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J. 2016;10(7):1669-1681.

20. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol. 2015;11(5):e1004226.

21. Layeghifard M, Hwang DM, Guttman DS. Disentangling interactions in the microbiome: a network perspective. Trends Microbiol. 2017;25(3):217-228.

22. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8(9):e1002687.

23. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol. 2012;8(7):e1002606.

24. Faust K, Lahti L, Gonze D, de Vos WM, Raes J. Metagenomics meets time series analysis: unraveling microbial community dynamics. Curr Opin Microbiol. 2015;25:56-66. doi:10.1016/j.mib.2015.04.004.

25. Matchado MS, Lauber M, Reitmeier S, Kacprowski T, Baumbach J, Haller D, et al. Network analysis methods for studying microbial communities: a mini review. Comput Struct Biotechnol J. 2021;19:2687-2698. doi:10.1016/j.csbj.2021.05.001.

26. Kim H, Jeon J, Lee KK, Lee YH. Longitudinal transmission of bacterial and fungal communities from seed to seed in rice. Commun Biol. 2022;5:772. doi:10.1038/s42003-022-03726-w.

27. Topçuoğ lu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. A framework for effective application of machine learning to microbiome-based classification problems. mBio. 2020;11(3):e00434-20.

Authors Bio

Yuxuan Liao is a Master's student in Business Analytics at the University of Bristol. Her research interests include data analytics, ecological systems, and interdisciplinary applications of computational methods. Her recent work focuses on microbial community stability, using network inference and ecological game dynamics to explore how complex microbial systems may respond to change. She is particularly interested in how quantitative modelling and data-driven approaches can support scientific interpretation and environmental research. Through this project, she hopes to further develop her skills in computational analysis and contribute to a better understanding of ecological and biological complexity.

Mathematical Models and Chemistry

Meet the Subject Editors

Alexandra Norris
Mathematical Modelling and Chemistry

Alexandra Norris is currently pursuing an integrated Master's degree in Chemistry with a year in Industry at the University of Bristol. Currently, she is undertaking a placement year in the Safety, Environmental, and Regulatory Sciences department of Unilever. She was excited to tackle the role of Chemistry Subject Editor for the Student Research Journal. Beyond her academic pursuits, Alexandra devotes her time to volunteering with young people and engaging in leadership roles of a student society. She hopes to further her career in scientific research laboratories within the field of analytical chemistry and make a positive mark on the world.

Hussain Almaainah
Mathematical Modelling and Chemistry

Hussain is graduating from the Bachelor's programme in Aerospace Engineering at the University of Bristol. Out of a passion for research and a desire for exposure to the realm of scientific research, he has taken on the role of Engineering Subject Editor for the Student Research Journal. His main research interests are experimental research in planetary exploration and space physics, so it was intriguing for him to read and give feedback on various related science and engineering topics. Having enjoyed participating in the journal, he hopes to contribute to more scientific journals in the future following his studies.

Hou Hin Ip (Izaac Ip)
Mathematical Modelling and Chemistry

Izaac Ip is a final-year BSc Data Science student at the University of Bristol. His research interests include machine learning, artificial intelligence, and data-driven modelling, with a particular focus on computer vision, LLMs, and NLP. He is interested in applying these technologies to address real world challenges and create positive social impact.

Numerical modelling of Fresnel and Fraunhofer diffraction using Gaussian quadrature and Monte Carlo integration methods.

Author: Chenhao Zhouhan

ORCID: 0009-0001-1307-5285

Programme: BScMathematics and Physics

Affiliation: School of Physics

Email: io24937@bristol.ac.uk

Abstract

This studyinvestigates the numerical modelling of Fresnel and Fraunhofer diffraction using Gaussian quadrature and Monte Carlo integration methods. This study simulates one-dimensional rectangular-aperture diffraction profiles and two- dimensional diffraction patterns for rectangular and circular apertures under ideal scalar diffraction assumptions. Gaussian quadrature is used as the deterministic integration method for the one-dimensional and two-dimensional cases, while Monte Carlo integration is tested as a stochastic sampling-based alternative for the circular aperture simulations. The resulting intensity distributions, numerical uncertainty maps, convergence behaviour and computational cost are compared to assess the accuracy- cost trade-off between the two approaches. The results show that Gaussian quadrature produces smoother and more stable diffraction patterns, particularly for two-dimensional simulations, but becomes computationally expensive for Fresnel diffraction and circular aperture geometries. Monte Carlo integration reproduces the main qualitative features with more predictable runtime but introduces sampling noise. Overall, the study demonstrates that the choice of numerical method depends on the required balance between precision, stability and computational efficiency.

Keywords: Fresnel diffraction, Fraunhofer diffraction, Gaussian quadrature, Monte Carlo integration, diffraction simulation.

Acknowledgement: The author would like to thank Dr Simon Hanna and Miss Yuyao Shi for helpful discussions on the numerical methods used in this computational physics project.

Introduction

Diffraction isa fundamental wave phenomenon that arises when light propagates through an aperture or around an obstacle, producing intensity patterns that depend on wavelength, aperture geometry and propagation distance [1]. In computational physics, numerical modelling provides a controlled way to study these patterns without requiring ideal experimental conditions [2]. This is particularly useful for comparing how different numerical integration methods reproduce oscillatory diffraction integral under well-defined assumptions.

Two important diffraction regimes are Fresnel (near-field) and Fraunhofer (far-field) diffraction. In Fresnel diffraction, the observation screen is at a finite propagation distance relative to the aperture size, meaning that the curvature of the wavefront remains significant. In contrast, Fraunhofer diffraction occurs when the observation screen is sufficiently distant relative to the aperture dimensions, allowing the diffracted wavefront to be approximated as locally planar. The transition between these regimes can be characterised using the Fresnel number, which is a dimensionless ratio that depends on the aperture size, wavelength and propagation distance.

The numerical evaluation of diffraction integrals presents a useful test case for comparing deterministic and stochastic integration methods. Gaussian quadrature provides a deterministic approach in which the integral is approximated using weighted function evaluations [3], often giving high accuracy for smooth integrands. In contrast, Monte Carlo integration estimates the integral statistically by sampling random points within the integration region. Although Monte Carlo methods typically converge more slowly, they can be easier to apply to high-dimensional integration problems [4].

This paper investigates the numerical modelling of Fresnel and Fraunhofer diffraction patterns through rectangular and circular apertures using Gaussian quadrature and Monte-Carlo integration. The study compares the resulting one-dimensional and two-dimensional intensity distributions, evaluates numerical uncertainty and convergence behaviour under different parameters settings and measures computational runtime. The aim is to assess the relative accuracy, stability and efficiency of the two integration approaches and to determine how the choice of numerical methods affects the simulation diffraction patterns in computational physics.

Methodology

Theoretical framework

Diffraction istheredistribution of a wave field after propagation through an aperture or around an obstacle, producing an intensity pattern that depends on wavelength, aperture geometry and propagation distance [1].

Fresnel diffraction refers to the phenomenon observed close to a two-dimensional object illuminated by plane parallel incident light. The amplitude of electric field E on any plane of observation at some distance beyond the object is represented by the following integral [5]:

$$E(x,y,z)= \frac{e^{ikz}}{i\lambda z}\int_{-\infty}^{\infty} E(x',y')e^{\frac{ik}{2z}[(x-x')^2+(y-y')^2]}dx'dy' \quad (1)$$

here λ is the wavelength, k is the wavenumber, z is the distance between the aperture and the screen and the integral is limited to the dimensions (x',y') of the aperture at which diffraction occurs.

While Fresnel diffraction happens at a finite distance to the screen (near-field), Fraunhofer diffraction happens at an infinite (large enough) distance to the screen (far- field). Therefore, Fresnel and Fraunhofer diffraction share the same physical nature, but operate under different regimes dictated by the combination of propagation distance and aperture dimensions.

Since the integrand in equation (1) is complex, the numerical calculation requires the real and imaginary parts to be evaluated separately and the equation to compute becomes:

$$E(x,y,z)= \frac{kE_0}{2\pi z}[Re(E)+iIm(E)] \quad (2)$$

where $Re(E)=cos(\frac{k}{2z}[(x-x)^{'2}+(y-y)^{'2}])$ is the real part and $Im(E) = sin(\frac{k}{2z}[(x-x)^{'2}+(y-y)^{'2}])$ is the imaginary part.

The observed diffraction pattern is represented by the intensity, which is proportional to the squared magnitude of the electric field.

$$I(x, y, z) = \epsilon_0 c\,[Re(E(x, y, z))^2 + Im(E(x, y, z))]^2 \quad (3)$$

3

.

with ϵ_0=$8.85\times10^{-12}A^2s^4kg^{-1}m^{-3}$ being the electric permittivity of free space and c=$3.00\times10^8ms^{-1}$ being the speed of light in free space.

The boundaries between Fresnel and Fraunhofer diffraction limits is determined by the Fresnel number [6],

$$F = \frac{r^2}{\lambda z} \tag{4}$$

where r is the characteristic aperture radius or half-width, λ is the wavelength and z is the aperture-screen distance. Fresnel diffraction corresponds to cases where the observation screen is sufficiently close, i.e. - $F \gg 1$ and $F \approx 1$. While Fraunhofer diffraction corresponds to cases where the observation screen is sufficiently large, i.e.- $F \ll 1$ [6].

Two numerical integration methods are compared. The first is the deterministic Gaussian quadrature method, which approximates an integral by evaluating the integrand at weighted points within the integration domain [7] [3].

$$\iint_A f(x,y)\,dx\,dy \approx \sum_{i=1}^{n}\sum_{j=1}^{m} w_i v_j f(x_i,y_j) \tag{5}$$

where $f(x,y)$ is the integrand evaluated at selected quadrature nodes (x_i,y_j) and w_i and v_j correspond to the corresponding weights in the x- and y-direction, respectively. In this study, Gaussian quadrature is used as a high-precision reference method for evaluating the diffraction integrals over rectangular and circular apertures.

The second method is Monte Carlo integration. For an aperture with area A , the integral of a function $f(x,y)$ over the aperture is estimated by randomly sampling N points within the integration region [8],

$$\int f(x,y)dA \approx A(\langle f\rangle \pm \sqrt{\frac{\langle f^2\rangle - \langle f\rangle^2}{N}}) \tag{6}$$

where $\langle f\rangle$ is the mean value of the integrand evaluated at the sampled points and $\langle f^2\rangle$ is the mean of the squared function, the square rooted term forms uncertainties of the estimation.

Simulation setup

Thesimulations were performed for both rectangular and circular apertures in one-dimensional and two-dimensional observation planes. For each geometry, the diffraction pattern was calculated in both the Fresnel and Fraunhofer regimes by choosing aperture dimensions, wavelength and screen distance values that produced the required Fresnel number range.

The physical constants and parameters used in the simulations are the free space permittivity, speed of light, wavelength, screen distance, rectangular aperture width, and the circular aperture radius. These values are kept constant throughout the simulations.

TABLE 1. Fundamental physical constants and parameters with their values used in the simulation. The values of the free space permittivity and the speed of light are taken from [9] & [10].

Magnitude	Symbol	Fresnel regime	Fraunhofer regime
Free space permittivity	ϵ^0	8.8542×10^{-12} $A^2s^4kg^{-1}m^{-3}$	
Speed of light	c	3.00×10^{8} ms^{-1}	
Wavelength	λ	0.7×10^{-6} m	
Screen distance	z	0.005 m	0.05 m
Rectangular width (x)	L^x	2.0×10^{-4} m	2.0×10^{-5} m
Rectangular width (y)	L^y	2.0×10^{-4} m	2.0×10^{-5} m
Circular radius	r	1.0×10^{-4} m	1.0×10^{-5} m

The simulations are produced under several simplifying physical assumptions to isolate the numerical behaviour of the integration methods. The incident wave was assumed to be monochromatic, the aperture edges were taken to be perfectly sharp, and the aperture plane and observation screen were assumed to be aligned perfectly parallel. Additionally, effects such as decoherence of the light source were neglected, assuming a perfectly coherent source. These assumptions allow the comparison to focus on numerical accuracy, stability and computational cost.

For Gaussian quadrature, the real and imaginary parts of the Fresnel integral were evaluated separately over the specified aperture limits. Rectangular apertures were integrated over fixed Cartesian bounds, while circular apertures required position-

dependent integration limits corresponding to the upper and lower semicircles. The absolute and relative tolerances were controlled through the numerical integration routine, with smaller tolerance values used to test convergence.

For the Monte Carlo method, random points were sampled uniformly within the circular aperture. The real and imaginary components of the integrand were evaluated at each sampled point, averaged over N samples, and then used to reconstruct the intensity function. The number of samples was varied to examine the convergence behaviour of the method.

The accuracy of each method was assessed using numerical error estimates and convergence tests. The convergence test qualifies how reliably a numerical method can reproduce accurate results when the numerical precision is varied [11]. A

numerical method F is called stable at some point x if:

$$\lim_{h \to tol} |F(x, t_h) - F(x, t_{ref})| = 0 \quad (7)$$

where t_h is the tolerance admitted for the computed results and t_{ref} is the reference tolerance. This test allows two approaches to be compared in terms of convergence behaviour, stability and computational cost. For reference, the precise values are computed with absolute tolerance $\epsilon_{abs}=10^{-14}$ and relative tolerance $\epsilon_{rel}=10^{-14}$.

Given that both methods can only calculate the errors in the real and imaginary part of the electric field, the errors in the intensity were calculated using the partial differentiation method on equation (3). Please see the Appendix for detailed procedures.

$$\Delta I = \sqrt{\epsilon_0 c}\left[\overline{(2Re \cdot \Delta Re)^2 + (2Im \cdot \Delta Im)^2}\right] \quad (8)$$

This quantity was used to generate the numerical error maps and indicates where the computed intensity is most sensitive to integration error. The numerical parameters used in the numerical methods for the convergence test are displayed as follows:

TABLE 2. Numerical control parameters used for convergence testing.

Method	Parameter	Values
Gaussian quadrature	Abs. tolerance	$10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$
	Rel. tolerance	$10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$

.

Monte Carlo integration	Number of samples N	$10^5, 10^4, 10^3, 10^2$

Finally, balancing numerical accuracy and computational cost, the figures presented in the study are all produced with $\epsilon_{abs}=\epsilon_{rel}=10^{-10}$ for the Gaussian quadrature and $N=10^5$.

The full Python implementation, including function definitions, plotting scripts, parameter definitions and figure generation code, is available in the accompanying GitHub repository [12].

Results

One-dimensionaldiffraction

The one-dimensional simulations provide an initial comparison between the Fresnel and Fraunhofer diffraction regimes for a rectangular aperture. In the Fresnel case, the intensity shows a more complex oscillatory structure across the observation screen, reflecting the contribution of curved wavefronts. In contrast, Fraunhofer diffraction is dominated by a central maximum with weaker secondary fringes.

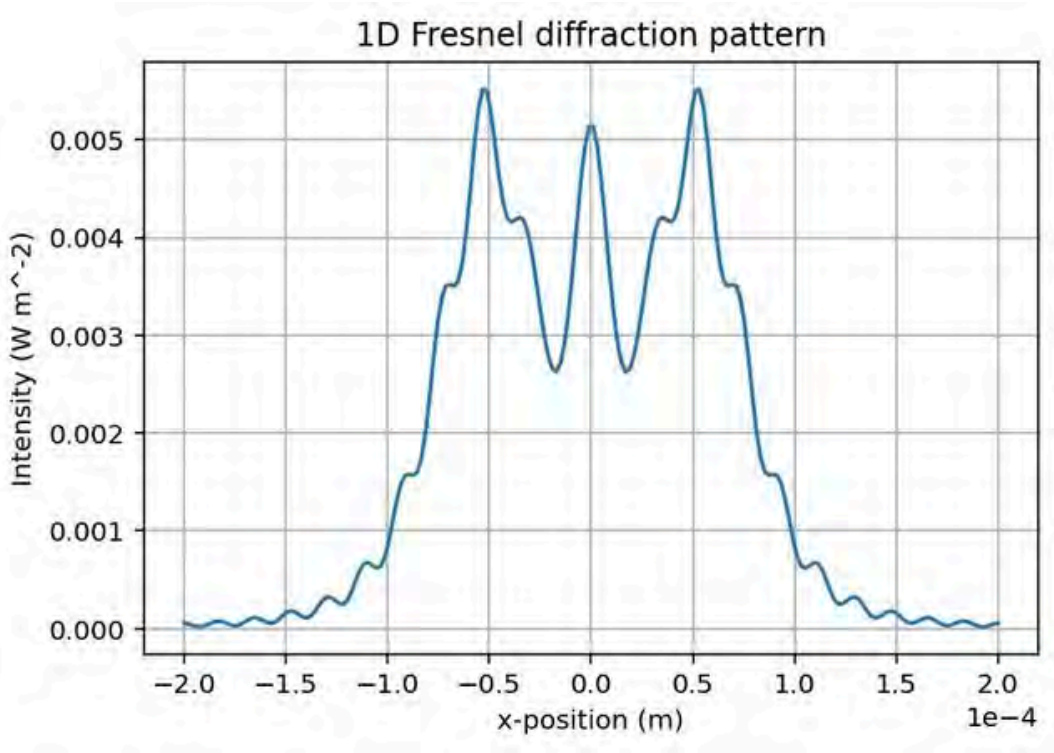


FIG 1. This figure illustrates the one-dimensional Fresnel diffraction pattern produced through a rectangularaperture. The intensity is evaluated along the x-axis by the Gaussian quadrature, showing characteristic near-fieldoscillations.

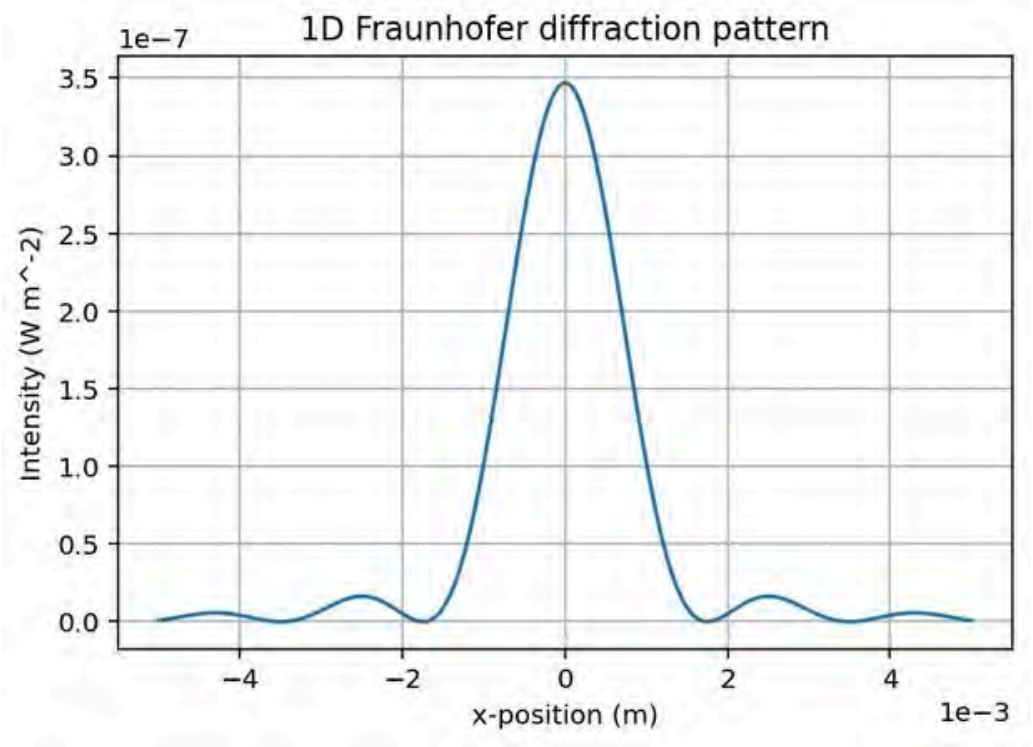


FIG 2. This figure illustrates the one-dimensional Fraunhofer diffraction pattern produced by a rectangularaperture. The intensity is evaluated along the x-axis by the Gaussian quadrature, showing characteristic far-field oscillations.

These results confirm that the chosen parameter values produce qualitatively distinct diffraction regimes before extending the calculation to two-dimensional aperture geometries. The one-dimensional results therefore serve as a baseline for interpreting the more computationally demanding two-dimensional simulations.

Two-dimensionaldiffractionwithGaussianquadrature

For the rectangular aperture, the Fresnel diffraction pattern shows a structured near-field distribution with multiple oscillatory regions around the centre. This reflects the stronger phase variation across the aperture when the screen is placed at a relatively short propagation distance.

In contrast, the Fraunhofer diffraction pattern is dominated by a central maximum with weaker surrounding fringes, showing the expected transition towards a far-field pattern.

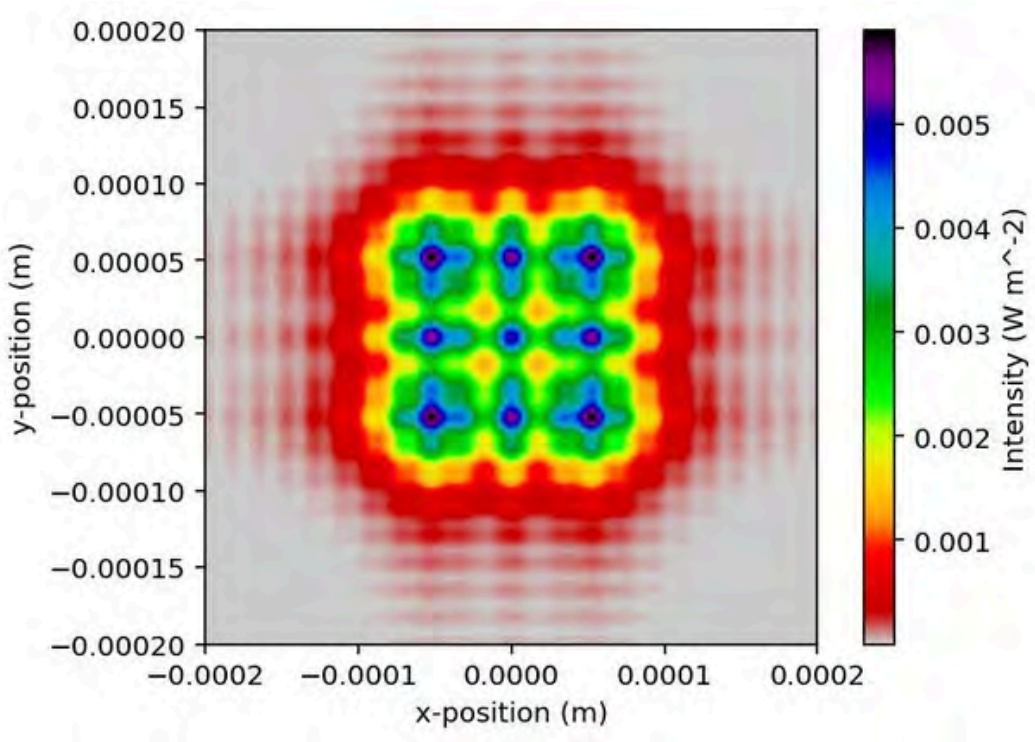


FIG 3. This figure illustrates the two-dimensional Fresnel diffraction pattern produced by a rectangularaperture evaluated with Gaussian quadrature. The intensity distribution shows the characteristic near-field oscillations.

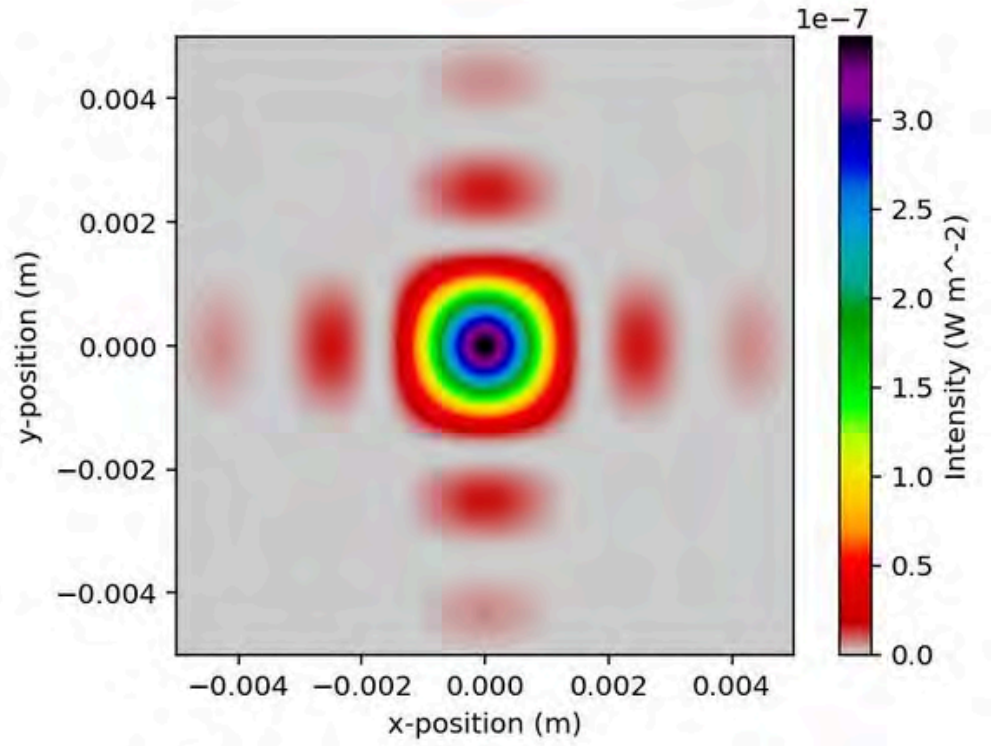


FIG 4. This figure illustrates the two-dimensional Fraunhofer diffraction pattern produced by a rectangularaperture evaluated with Gaussian quadrature. The intensity distribution shows the characteristic far-field oscillations.

The circular aperture simulations show the same distinction between the two regimes, but with radial symmetry imposed by the aperture geometry.

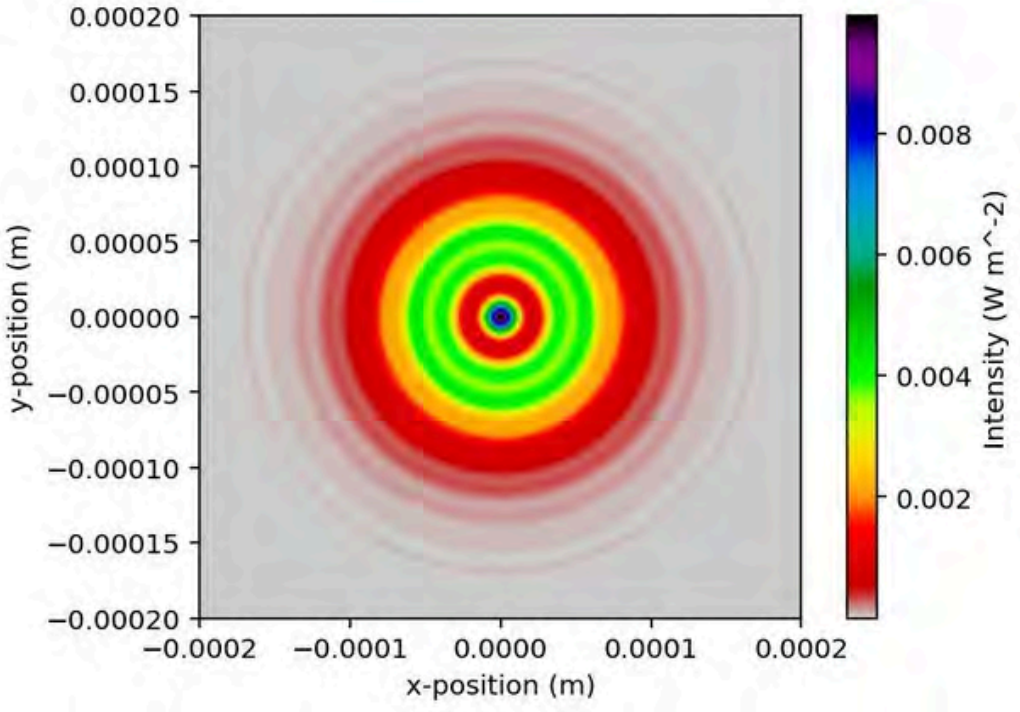


FIG 5. This figure illustrates the two-dimensional Fresnel diffraction pattern produced by a circular apertureevaluated

with Gaussian quadrature. The intensity distribution shows the characteristic near-field oscillations.

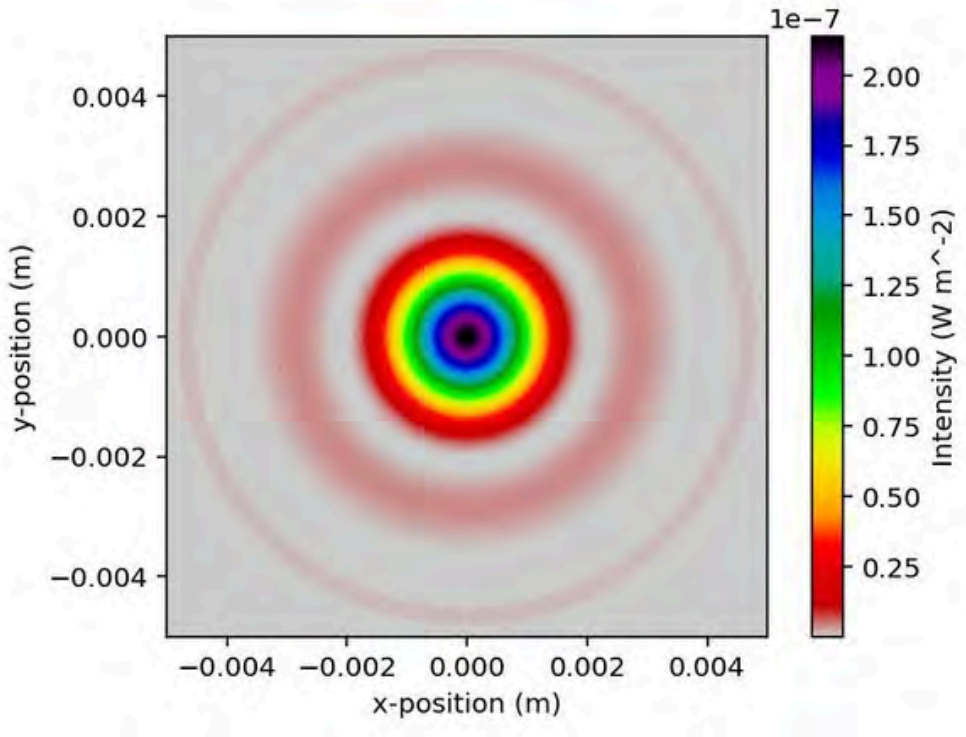


FIG 6. This figure illustrates the two-dimensional Fraunhofer diffraction pattern produced by a circularaperture evaluated with Gaussian quadrature. The intensity distribution shows the characteristic far-field oscillations.

The Gaussian quadrature results provide smooth and stable two-dimensional diffraction maps for both aperture geometries. These results support the use of Gaussian quadrature as a high-precision deterministic method for evaluating the diffraction integral, although the computational cost of this approach increases substantially for two-dimensional calculations.

Two-dimensionaldiffractionwithMonteCarlointegration

Monte Carlo integration was used to calculate two-dimensional diffraction patterns for the circular aperture. The method reproduces the main qualitative features observed in the Gaussian quadrature results, including radially symmetric structure imposed by the circular geometry and the distinction between the Fresnel and Fraunhofer regimes.

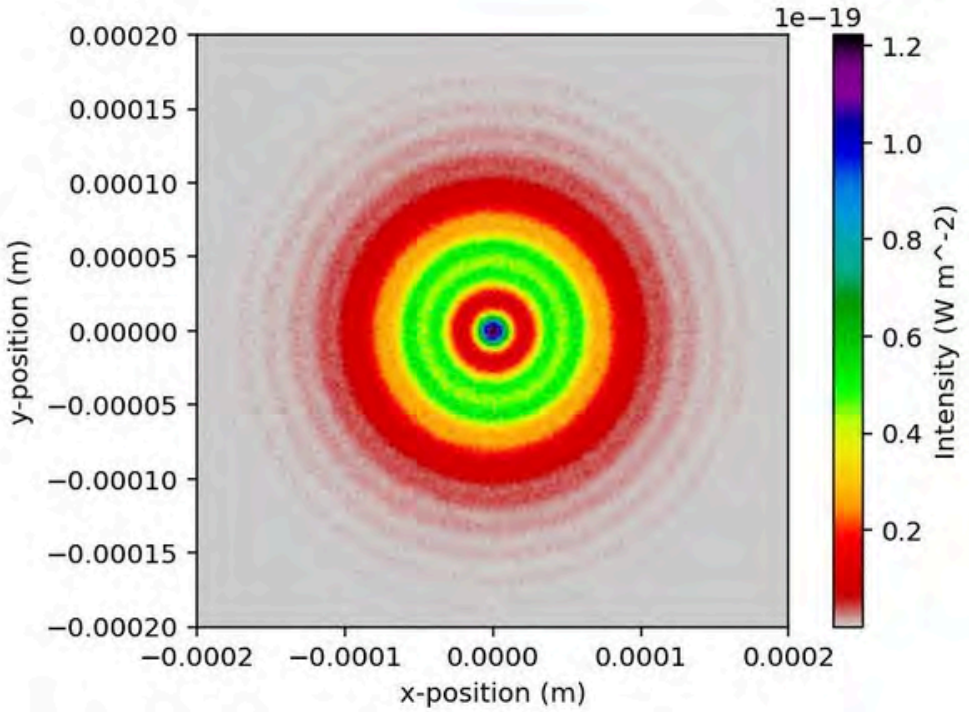


FIG 7. This figure illustrates the two-dimensional Fresnel diffraction pattern produced by a circular apertureevaluated with Monte Carlo integration. The intensity distribution shows the characteristic near-field oscillations.

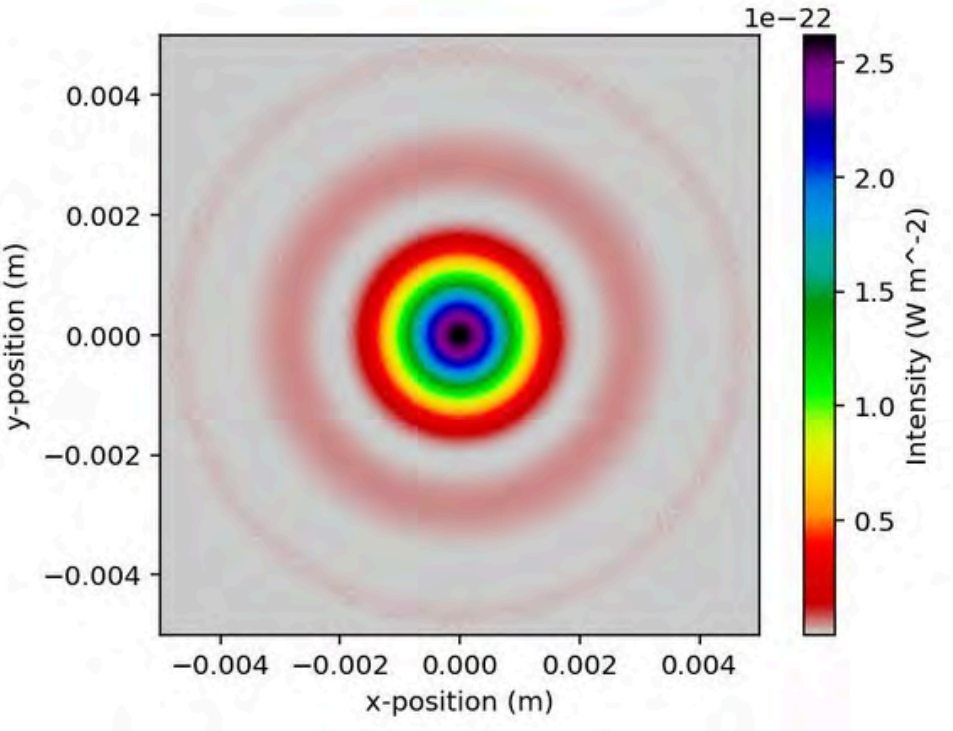


FIG 8. This figure illustrates the two-dimensional Fraunhofer diffraction pattern by a circular aperture evaluated with

.

Monte Carlo integration. The intensity distribution shows the characteristic far-field oscillations.

Although the main diffraction structure is recovered, the Monte Carlo intensity maps appear less smooth than the corresponding Gaussian quadrature results. This is expected because of the statistical nature of the Monte Carlo integration, which estimates the integral from a finite number of randomly sampled points. As a result, the calculated intensities contain sampling noise.

These results show that Monte Carlo integration can reproduce the overall form of the patterns, but with lower visual smoothness and numerical stability than Gaussian quadrature for the same physical configurations.

Numericaluncertaintyanderrormaps

The numerical uncertainty maps were produced using equation (8), calculated from the numerical uncertainties in the real and imaginary components of the electric field. The resulting maps indicate where the computed intensity is most sensitive to numerical integration error.

For Gaussian quadrature, the uncertainty maps are generally smooth and structured, reflecting the deterministic nature of the integration method.

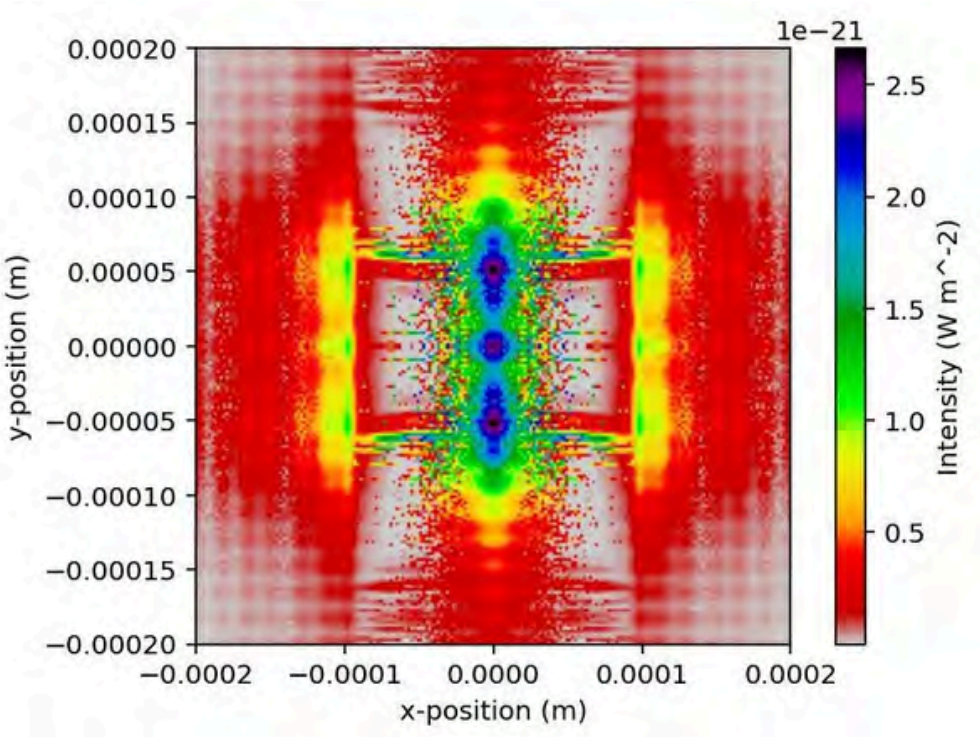


FIG 9. The figure illustrates the two-dimensional numerical error of Fresnel diffraction pattern through a rectangular aperture evaluated with Gaussian quadrature. The error mapping shows where the Gaussian quadrature is most sensitive.

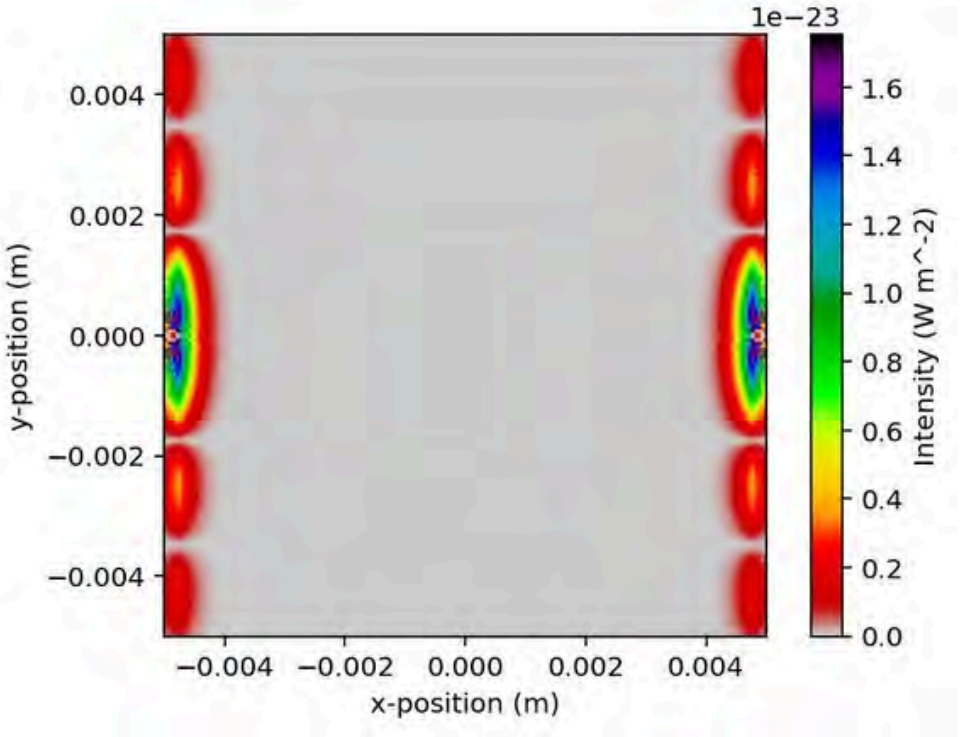


FIG 10. The figure illustrates the two-dimensional numerical error of Fraunhofer diffraction patternthrough a rectangular aperture evaluated with Gaussian quadrature. The error mapping shows where the Gaussianquadrature is most sensitive.

.

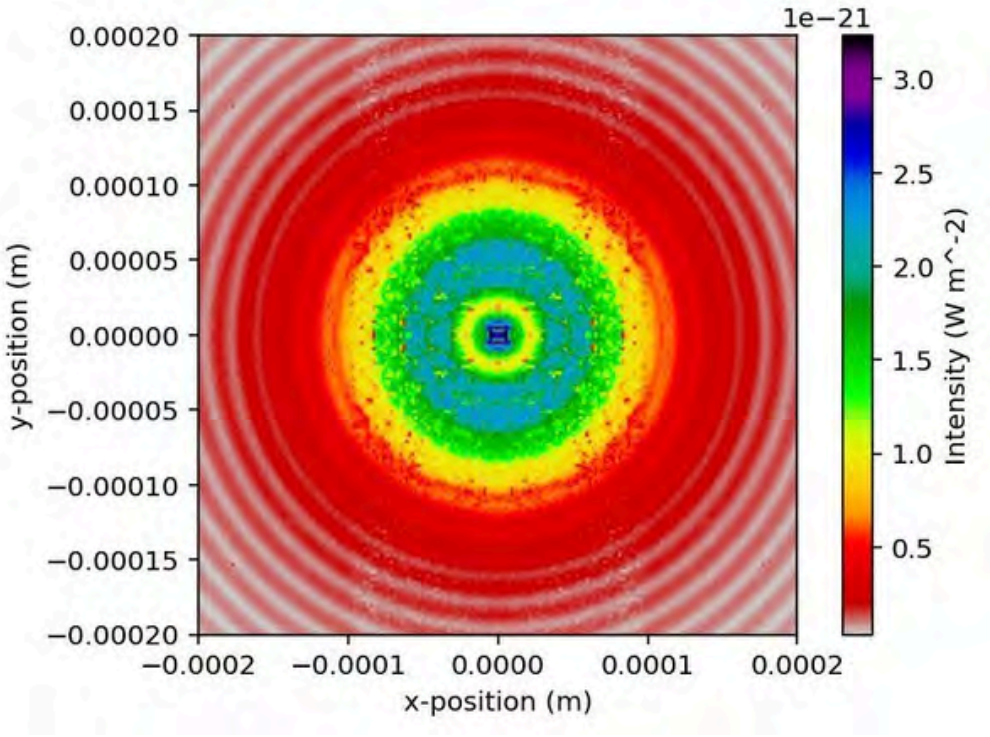


FIG 11. The figure illustrates the two-dimensional numerical error of Fresnel diffraction pattern throughacircular aperture evaluated with Gaussian quadrature. The error mapping shows where the Gaussian quadratureismost sensitive.

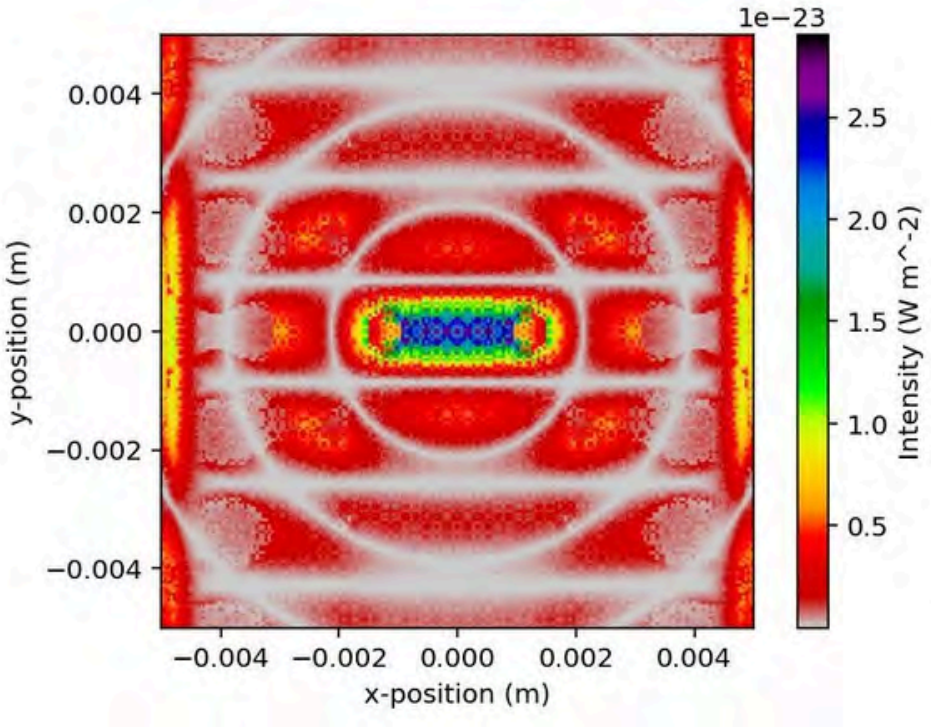


FIG 12. The figure illustrates the two-dimensional numerical error of Fraunhofer diffraction pattern throughacircular aperture evaluated with Gaussian quadrature. The error mapping shows where the Gaussian quadratureismost

.

sensitive.

For Monte Carlo integration, the uncertainty maps display a less regular structure because the method is based on random statistical sampling.

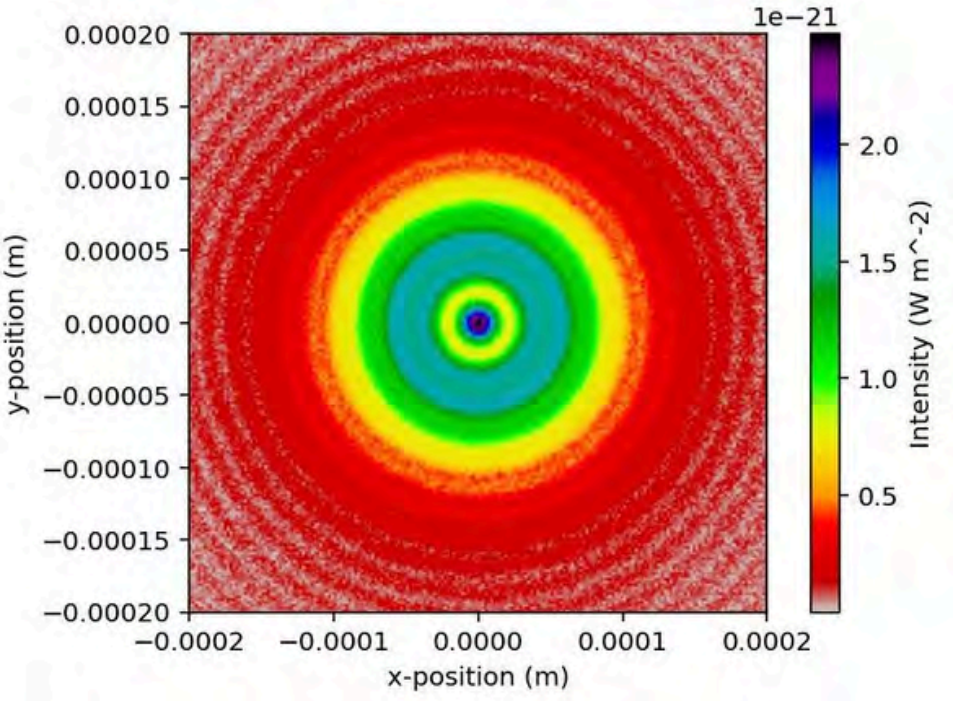


FIG 13. This figure illustrates the two-dimensional numerical error of Fresnel diffraction pattern throughacircular aperture evaluated with Monte Carlo integration. The error mapping shows where the Monte Carlo integrationismost sensitive.

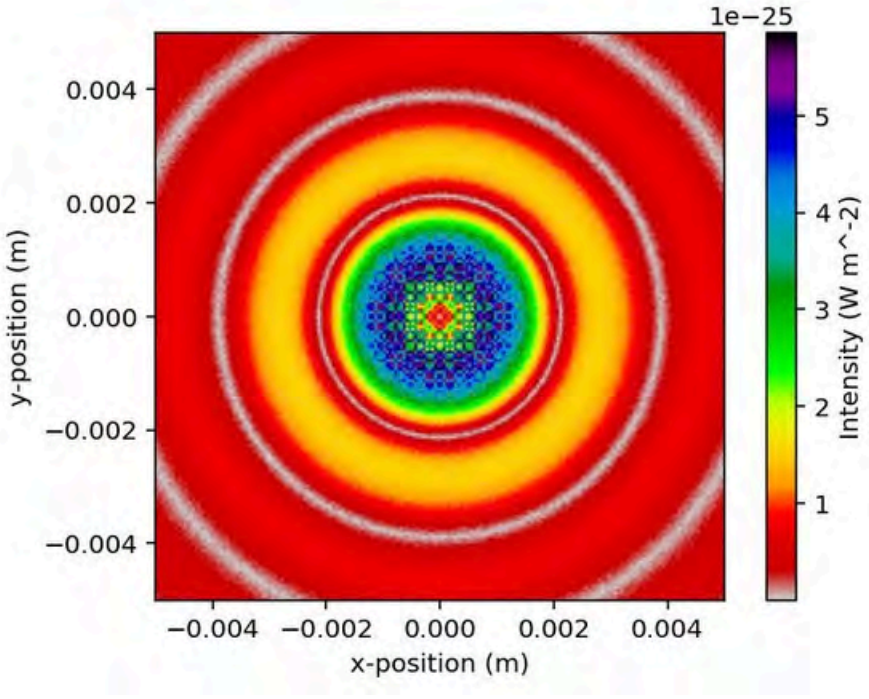


FIG 14. This figure illustrates the two-dimensional numerical error of Fraunhofer diffraction pattern through a circular

aperture evaluated with Monte Carlo integration. The error mapping shows where the Monte Carlo integration is most sensitive.

Convergence tests

Convergencetestswere performed to evaluate how the numerical results depend on the control parameters of each integration method. The absolute and relative tolerances for Gaussian quadrature and the number of random samples for Monte Carlo integration were varied as shown in TABLE 2.

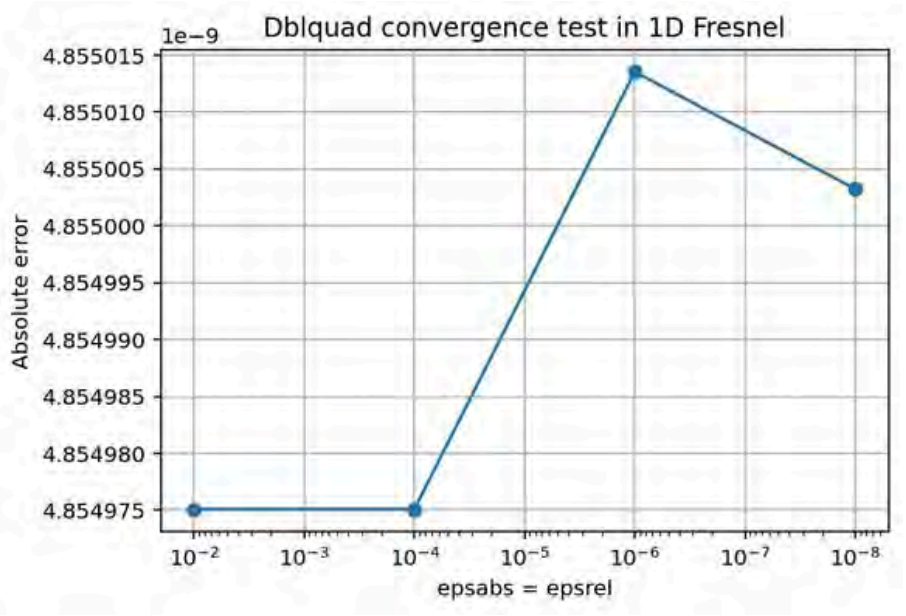


FIG 15. This figure illustrates the absolute error in one-dimensional Fresnel diffraction as a functionofthetolerances. The shape shows that Gaussian quadrature becomes precision-limited at this scale, producingnoimprovements with smaller tolerances.

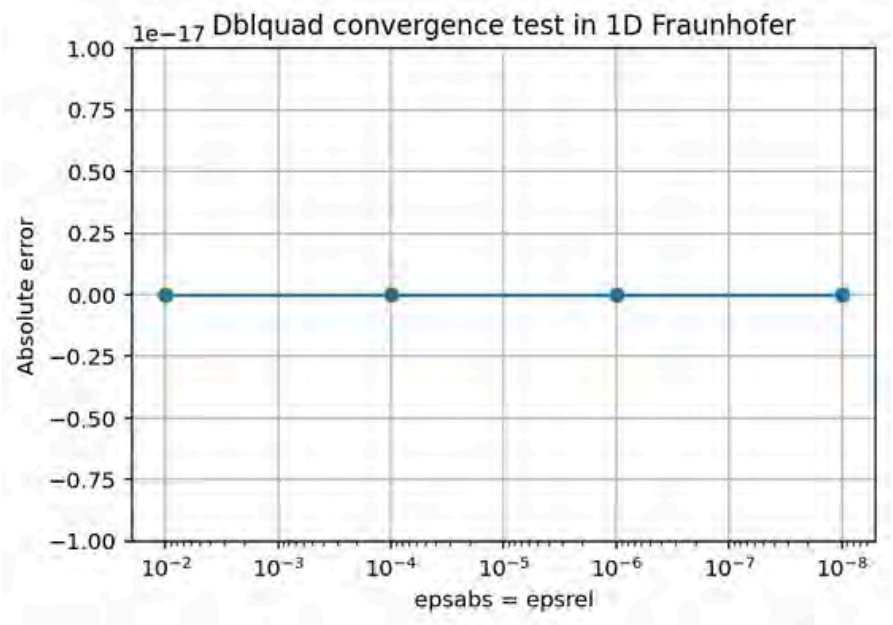


FIG 16. This figure illustrates the absolute error in one-dimensional Fraunhofer diffraction as a function of the tolerances. The shape shows that Gaussian quadrature becomes precision-limited at this scale, producing no improvements with smaller tolerances.

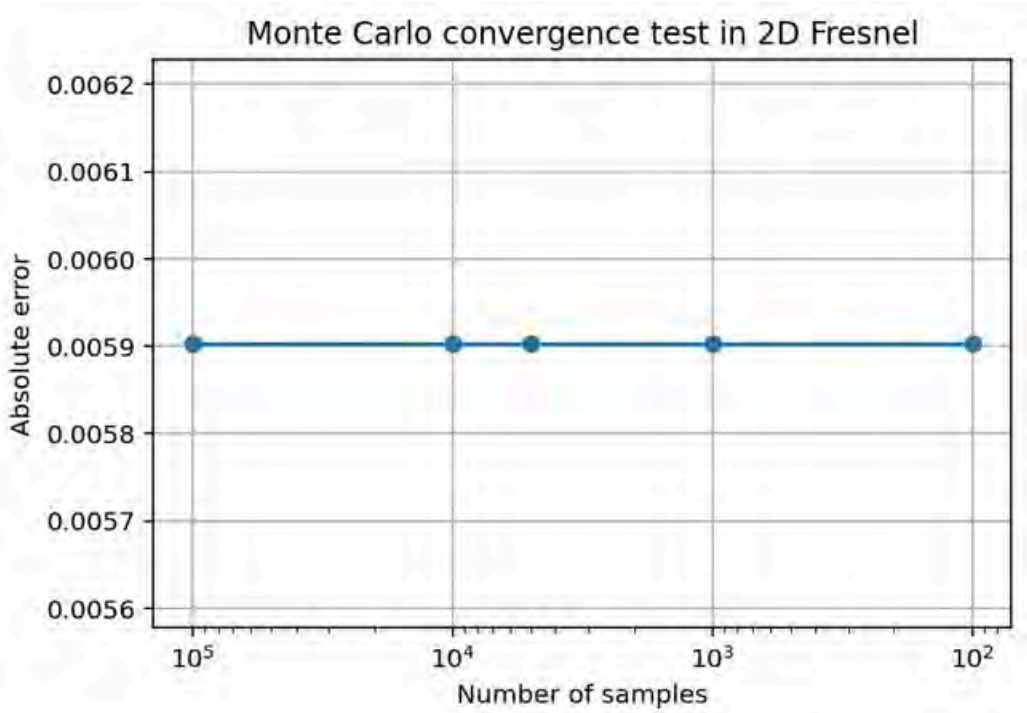


FIG 17. This figure illustrates the absolute error in two-dimensional Fresnel diffraction as a function of thenumber of samples. The shape shows that Monte Carlo integration becomes precision-limited at this scale, producing no improvements with greater number of samples.

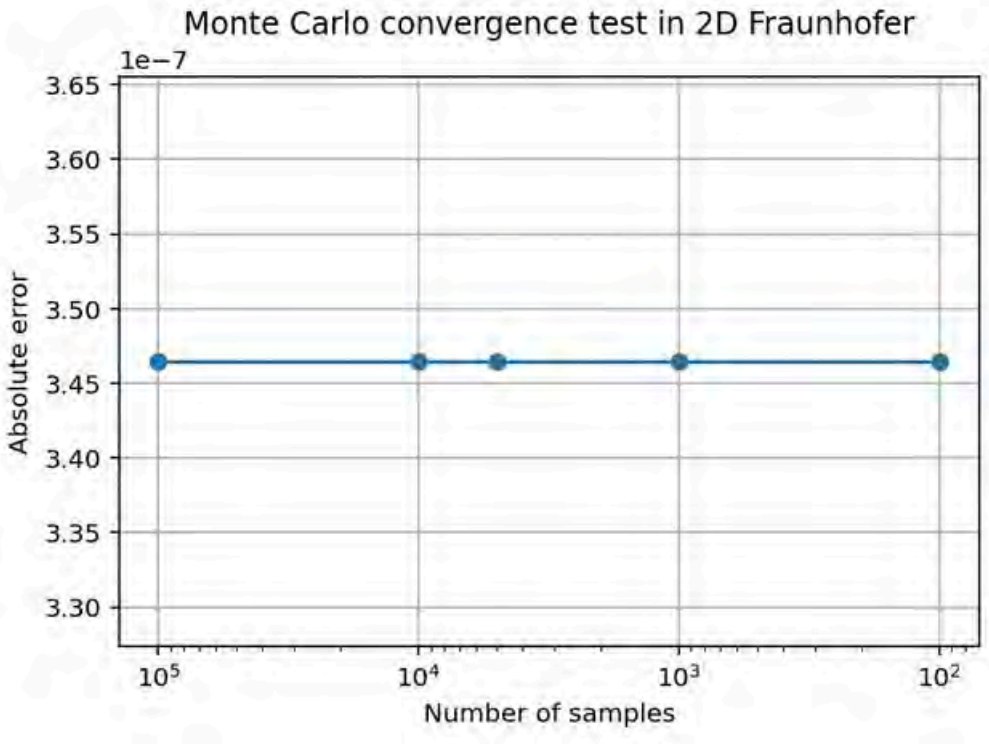


FIG 18. This figure illustrates the absolute error in two-dimensional Fraunhofer diffraction as a function of the number of samples. The shape shows that Monte Carlo integration becomes precision-limited at this scale, producing no improvements with greater number of samples.

Computational runtime

Thecomputationalruntime was measured to compare the efficiency of Gaussian quadrature and Monte Carlo integration for different aperture geometries and diffraction regimes. The results show that one-dimensional calculations were completed rapidly, while two-dimensional calculations required substantially longer runtimes, especially in the Fresnel regime.

TABLE 3. Time test for different methods on different aperture geometries. The values represent the time needed to compute 200 data points with and without @njit acceleration.

Method	Aperture	No @njit (s)	@njit (s)
Gaussian quadrature	Rectangular 1D Fresnel	3.12	1.06
	Rectangular 1D Fraunhofer	0.19 1112.54	0.08
	Rectangular 2D Fresnel	41.07	365.64
	Rectangular 2D Fraunhofer	1727.08	17.81
	Circular 2D Fresnel	53.94 188.07	646.73
	Circular 2D Fraunhofer	186.02	30.34
Monte Carlo	Circular 2D Fresnel		177.53
	Circular 2D Fraunhofer		179.13

The use of @njit acceleration reduced the runtime of some parts of the implementation, particularly where repeated numerical operations were performed.

Discussion

Thepresentwork is a computational modelling study of Fresnel and Fraunhofer diffraction patterns in one and two dimensions using Gaussian quadrature and Monte Carlo integration to estimate solutions to the Fresnel diffraction integral. The study combines scalar diffraction theory and numerical analysis to visualise diffraction patterns and assess the accuracy, stability and computational cost of two different integration approaches.

In the one-dimensional simulation, Fresnel and Fraunhofer diffraction patterns were obtained to provide a baseline comparison between the two diffraction regimes. The Fresnel number was used to distinguish between near-field and far-field conditions. The resulting plots show that the Fresnel regime produces a more oscillatory intensity profile, while the Fraunhofer regime has a clearer central maximum and weaker secondary fringes. This difference is expected because, in the near field, the curvature of the wavefront remains significant and produces rapidly varying phase differences across the observation screen [13]. In the far field, the wavefront can be approximated as locally planar, giving a more stable and regular diffraction structure [14].

The two-dimensional simulations show the same general distinction between the Fresnel and Fraunhofer regimes. The aperture geometry also has a clear effect on the final intensity distribution. Rectangular apertures produce structures aligned with the Cartesian axes, whereas circular apertures produce radially symmetric patterns. This confirms that the numerical model responds appropriately to both propagation regime and aperture geometry.

The two integration methods are based on different numerical principles. Gaussian quadrature is a deterministic method that approximates the integral using weighted evaluations of the integrand, while Monte Carlo integration estimates the same integral statistically from randomly sampled points. In the simulations, Gaussian quadrature generally produced smoother intensity maps and more stable numerical behaviour. This is expected because the aperture integrals are evaluated systematically over the integration domain. Monte Carlo integration reproduced the main qualitative diffraction structures, but resulting intensity maps were less smooth because a finite number of random samples introduces sampling noise.

The numerical uncertainty maps were obtained from equation (8), calculated from the uncertainties in the real and imaginary components of the electric field. These maps should be interpreted as numerical sensitivity maps rather than experimental error maps. They do not consider any experimental effects and factors. Instead, they show where the computed intensity is most sensitive to numerical integration uncertainty. In general, Gaussian quadrature produced smoother uncertainty maps as shown in FIG 9, FIG 10, FIG 11 & FIG 12, while Monte Carlo integration showed more irregular uncertainty distributions due to its sampling-based nature as seen in FIG 13 & FIG 14.

The convergence tests provide a more quantitative way to assess numerical reliability. For Gaussian quadrature, reducing the absolute and relative tolerances improves agreement with the reference calculation until the error reaches a plateau. Beyond this point, using smaller tolerance values does not produce a meaningful improvement in the plotted intensity distribution as shown in FIG 15 & FIG 16 and mainly increases computational cost. This suggests that the remaining differences are limited by factors such as numerical precision, grid resolution and the reference calculation itself rather than by the quadrature tolerance alone.

For Monte Carlo integration, the convergence behaviour is controlled by the number of random samples. Increasing the sample number reduces statistical fluctuations and improves the stability of the estimated intensity distribution. However, this improvement is gradual because the statistical uncertainty of Monte Carlo integration decreases slowly with sample size. As a result, a large increase in N is required to obtain a relatively modest improvement in smoothness and numerical precision. This explains why Monte Carlo integration can be computationally attractive but less accurate than Gaussian quadrature for the cases considered in this study.

The runtime measurements show an important trade-off between numerical accuracy and computational cost. Gaussian quadrature provides, high precision but it becomes computationally expensive for two-dimensional Fresnel diffraction, especially for circular apertures, as the rapidly oscillating phases make the integral more difficult to evaluate accurately. Additionally, circular apertures also require position-dependent integration limits, which increases the computational cost compared with rectangular apertures. In contrast, Monte Carlo integration has a more predictable runtime because the number of sampled points is fixed. This makes it faster for some computationally expensive two-dimensional cases, but the gain in speed comes at the cost of increased sampling noise and lower numerical smoothness.

The use of @njit acceleration reduced the runtime of some calculations, particularly for Gaussian quadrature which is based on repeated numerical operations. However, its effect was not uniform across all methods because external numerical routines such as dblquad are not fully accelerated by @njit. Therefore, @njit should be regarded as an implementation-level optimisation.

The central result remains the trade-off between deterministic quadrature and stochastic sampling: Gaussian quadrature is preferable when smooth and precise intensity maps are required, whereas Monte Carlo integration offers a faster but less precise alternative for computationally demanding two-dimensional simulations. Previous studies compared numerical methods for solving Fresnel-integral diffraction problems in aperiodic diffractive lenses and concluded that Gauss quadrature provided the best balance between accuracy and computation time [2]. The present study builds on this observation by using Gaussian and comparing it directly with Monte Carlo integration with different criteria. Therefore, the contribution of this work is to compare a high-performing deterministic method with a stochastic alternative under a simplified and reproducible diffraction framework.

The main limitations of this study come from the idealised physical assumptions and the numerical scope of the comparison. Due to the simplifying physical assumptions, the results describe an ideal scalar diffraction model rather than a complete experimental optical system. In addition, the Monte Carlo implementation was tested only for circular apertures, so a broader comparison involving more aperture geometries would be needed to fully assess its flexibility.

Future work could extend the study to adaptive quadrature, quasi-Monte Carlo methods, more complex aperture shapes and experimental validation against measured diffraction patterns. A further possible direction is to investigate whether numerical diffraction simulations can be combined with artificial intelligence and machine learning approaches [15]. Diffraction datasets generated from controlled numerical simulations could potentially be used to train or test models for pattern recognition or inverse design.

Conclusion

In conclusion, this study investigated the numerical modelling of Fresnel and Fraunhofer diffraction using Gaussian and Monte Carlo integration. The simulations reproduced one-dimensional and two-dimensional diffraction patterns for rectangular and circular apertures, allowing the influence of the diffraction regime, aperture

geometry and numerical method to be compared.

The results show that Gaussian quadrature provides smoother and more stable intensity distributions than Monte Carlo integration for the cases considered. This is shown in the two-dimensional diffraction maps and numerical uncertainty maps, where deterministic nature of Gaussian quadrature produces more regular results and Monte Carlo integration was able to reproduce the main qualitative diffraction structures with numerical fluctuations and uncertainties due to its stochastic nature.

The convergence and runtime test further demonstrate the importance of balancing numerical accuracy with computational cost. For Gaussian quadrature, reducing tolerance improved the result only up to a certain point, after which further reduction mainly increased runtime without producing a significant improvement in the plotted intensity distribution. For Monte Carlo integration, increasing the number of samples improved numerical stability, but convergence was slower and required substantially larger sample numbers to reduce sampling noise.

Overall, Gaussian quadrature is more suitable when high precision and smooth diffraction patterns are required, while Monte Carlo integration provides a more flexible and computationally predictable alternative for complex two-dimensional cases. The study therefore highlights the need to evaluate accuracy, convergence behaviour and runtime together when selecting numerical methods for diffraction modelling. Future work could extend this comparison to adaptive quadrature, quasi-Monte Carlo methods and potential applications to train machine-learning-assisted optimisation models.

References

[1] J. R. Wright, “Fundamentals of Diffraction,” *Journal of the Audio Engineering Society,* vol. 45, no. 5, pp. 347-356, 1997.

[2] A. Garmendía-Martínez, F. M. Muñoz-Pérez, W. D. Furlan, F. Giménez, J. C. Castro-Palacio, J. A. Monsoriu and V. Ferrando, “Comparative Study of Numerical Methods for Solving the Fresnel Integral in Aperiodic Diffractive Lenses,” *Mathematics,* vol. 11, no. 4, p. 946, 2023.

[3] N. Kovvali, Theory and Applications of Gaussian Quadrature Methods, Springer, 2022.

[4] J. Dick, F. Y. Kuo and I. H. Sloan, “High-dimensional integration: The quasi-Monte Carlo way,” *Acta Numerica,* vol. 22, pp. 133-288, 2013.

[5] J. Cowley, Diffraction Physics, Elsevier, 1995.

[6] S. Wang, E. Bernabei and J. Alda, “Unified and Generalized Fresnel Numbers,” *Optical and Quantum Electronics,* vol. 24, no. 12, pp. 1351-1358, 1992.

[7] J. Geweke, “Chapter 15 Monte Carlo Simulation and Numerical Integration,” *Handbook of Computational Economics,* vol. 1, pp. 731-800, 2023.

[8] C. Z. Mooney, Monte Carlo Simulation, SAGE Publications, 1997.

[9] NIST, “CODATA Value: vacuum electric permittivity,” NIST.gov, 2022. [Online]. Available: https://physics.nist.gov/cgi-bin/cuu/Value?ep0. [Accessed 31 05 2026].

[10] NIST, “CODATA Value: speed of light in vacuum,” NIST.gov, 2022. [Online]. Available: https://physics.nist.gov/cgi-bin/cuu/Value?c. [Accessed 31 05 2026].

[11] B. Chartres and R. Stepleman, “A General Theory of Convergence for Numerical Methods,” *SIAM Journal on Numerical Analysis,* vol. 9, no. 3, pp. 476-492, 1972.

[12] C. Zhouhan, “Fresnel-Fraunhofer-Diffraction-Numerical-Methods,” GitHub, 30 05 2026. [Online]. Available: https://github.com/zchen3141596-afk/Fresnel-Fraunhofer-Diffraction-Numerical-Methods . [Accessed 30 05 2026].

[13] C. Aime, É. Aristidi, Y. Rabbia, D. Mary , C. Theys and C. AIme, “The Fresnel Diffraction: a Story of Light and Darkness,” *EAS Publications Series,* vol. 59, pp. 37-58, 2013.

[14] Y. Liu, L. Wang, J. Dong, J. Xia, L. Yang and Y. Jin, “Theoretical Simulation on Fraunhofer Diffraction of Arbitrarily Shaped Aperture,” *Journal of Physics: Conference Series,* vol. 2313, no. 1, pp. 012026-012026, 2022.

[15] M. Puttegowda and S. Ballupete Nagaraju, “Artificial Intelligence and Machine Learning in Mechanical engineering: Current Trends and Future Prospects,” *Engineering Applications of Artificial Intelligence,* vol. 142, p. 109910, 2025.

Appendix

Stepstoevaluate the partial derivatives in equation (8).

$$\Delta I^2=\left(\frac{\partial I}{\partial Re(E)}\Delta Re(E)\right)^2+\left(\frac{\partial I}{\partial Im(E)}\Delta Im(E)\right)^2 \Rightarrow$$

$$\Delta I^2=\left(\frac{\partial \epsilon_0 c(Re^2+Im^2)}{\partial Re}\Delta Re\right)^2+\left(\frac{\partial \epsilon_0 c(Re^2+Im^2)}{\partial Im}\Delta Im\right)^2 \Rightarrow$$

$$\Delta I^2=(2\epsilon_0 cRe\cdot\Delta Re)^2+(2\epsilon_0 cIm\cdot\Delta Im)^2 \Rightarrow$$

$$\Delta I=\sqrt{\epsilon_0 c[(2Re\cdot\Delta Re)^2+(2Im\cdot\Delta Im)^2}]$$

Authors Bio:

Chenhao Zhouhan is a BSc Mathematics and Physics student at the University of Bristol. His academic interests include computational physics, numerical modelling and the application of mathematical methods to physical systems. Through his undergraduate studies, Chenhao has worked on projects involving numerical integration and physical modelling, with a particular focus on making abstract physical principles more accessible through simulation. He hopes to continue developing his skills in applied mathematics, computational physics and research-based scientific communication.

Temporal Kolmogorov-Arnold Networks (T-KAN) for High-Frequency Limit Order Book Forecasting: Efficiency, Interpretability, and Alpha-Decay

Author: Ahmad Makinde

ORCID: 0009-0002-9213-9204

Programme: Philosophy and Economics BSc

Affiliation: School of Economics

Email: Ahmad.makinde.2025@bristol.ac.uk

Abstract

High-Frequency trading (HFT) environments are characterised by large volumes of limit order book (LOB) data, which is notoriously noisy and non-linear. Alpha decay represents a significant challenge, with traditional models such as DeepLOB losing predictive power as the time horizon (k) increases. In this paper, using data from the FI-2010 dataset, we introduce Temporal Kolmogorov-Arnold Networks (T-KAN) to replace the fixed, linear weights of standard LSTMs with learnable B-spline activation functions. This allows the model to learn the 'shape' of market signals as opposed to just their magnitude. This resulted in a 19.1% relative improvement in the F1-score at the k = 100 horizon. The efficacy of T-KAN networks cannot be understated, producing a 132.48% return compared to the -82.76% DeepLOB drawdown under 1.0 bps transaction costs. In addition to this, the T-KAN model proves quite interpretable, with the 'dead-zones' being clearly visible in the splines. The T-KAN architecture is also uniquely optimised for low-latency FPGA implementation via High level Synthesis (HLS). The code for the experiments in this project can be found at https://github.com/AhmadMak/Temporal-Kolmogorov-Arnold-Networks-T-KAN-for-High-Frequency-Limit-Order-Book-Forecasting.

Keywords Limit Order Book (LOB), Alpha Decay, FI-2010 Dataset, Temporal Kolmogorov-Arnold Network (T-KAN), Interpretability

1. Introduction

The modelling and prediction of price dynamics in the Limit Order Book (LOB) are fundamental challenges in quantitative finance and market microstructure [1, 2]. Unlike low-frequency data, the LOB is a high-dimensional, discrete-event dynamic system where the latent state of supply and demand is shown via the placing and cancellation of orders across multiple price levels [3]. The LOB state at time t can be represented as a vector $\mathcal{L}_t = \{P_t^{(i)}, V_t^{(i)}\}^n_{i=-n}$' where P and V represent the price and volume at level i, where positive and negative indices denote ask and bid sides, respectively.

In this state space, the Auction Phase is important. The phase is characterised by intense price discovery and structural liquidity shifts, where changing from a closed-call auction to continuous trading causes high-volatility regimes [4]. In order to forecast accurately, this regime requires the model to be able to capture complex, "path-dependent" non-linearities, where a trade's price impact is a dynamic function of current book depth and historical order flow [5].

This study uses a 144-dimensional feature vector whose values were pre-normalised using z-score standardisation by the dataset provider. For a feature x, the normalised value $\hat{x}$ is defined as

$$\hat{x} = \frac{x - \mu}{\sigma}$$

where μ and σ are the mean and standard deviation calculated over a rolling window to maintain a stationary auction environment.

Traditional forecasting has shifted significantly from linear econometric models to deep recurrent architectures such as the Long Short-Term Memory (LSTM) network. Standard LSTMs are however reliant on fixed, point-wise activation functions within its gating mechanisms. A standard LSTM gate is defined by:

$$it = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (1)$$

$$ft = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$ot = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

where W represents static weight matrices. The architecture assumes that a linear transformation with a fixed non-linearity is adequate to map LOB features to price movements. We assume that this ’universal approximation’ approach is parameter-inefficient for capturing localised oscillations found in microstructure data.

This paper proposes Temporal Kolmogorov-Arnold Network (T-KAN) as a superior alternative to LOB forecasting. By replacing static matrices W with learnable univariate spline functions, the T-KAN allows "computation on the edges" [6]. This makes T-KANs a significant development within the field due to exciting properties - which will be discussed in detail. With a configuration using 532,675 parameters, our model provides a high-resolution manifold to capture aggressive price discovery in the Auction Z-score regime of the FI-2010 dataset.

2 Literature Review

2.1DeepLearningin MarketMicrostructure

The release of the FI-2010 benchmark dataset by Ntakaris et al. [4] in 2017 provided a large-scale standardised platform to evaluate machine learning models in LOB forecasting. Earlier studies used Convolutional Neural Networks (CNNs) to automate the extraction of spatial characteristics from 40-dimensional raw LOB data [7]. DeepLOB later integrated CNNs with LSTMs to model spatial and temporal dependencies [8].

The expressive capacity of these models relies on the Universal Approximation Theorem, which states that a network with fixed activations can approximate any continuous function. Although effective, the "curse of dimensionality" often impacts this approach when modelling functions with high-frequency components [9]. This has led researchers to look at architectures such as the Temporal Attention Augmented bilinear (TABL) network, which uses bilinear projections to compress LOB features whilst maintaining temporal relationships [10].

2.2 Existing LOB architectures

In addition to convolutional and bilinear architectures, recent work in LOB modelling has increasingly used Transformer-based models and self-attention mechanisms to capture long-range dependencies, for example: the Multi- Scale Axial Attention network (LOB-Attn). While self-attention architectures are good at parsing macro-temporal contexts, they scale quadratically with sequence length and heavily rely on stationary inputs, so often fail to smoothly adapt to rapid structural changes within volatile intraday

environments. In addition to this, Neural Ordinary Differential Equations (Neural ODEs) have been introduced to model LOB continuous-time dynamics. However, these models introduce substantial inference latency and face intense computational bottlenecks when handling sub-millisecond, high-frequency quote flows. As a result, modeling non-stationary price dynamics under ultra-low latency constraints is still an open challenge in high-frequency market microstructure literature.

2.3 Kolmogorov-ArnoldNetworks(KAN)andSplineTheory

Liu et al. (2024) introduced Kolmogorov-Arnold Networks (KANs) as a significant alternative to MultiLayer Perceptrons (MLPs) [6]. Based on the Kolmogorov-Arnold Representation Theorem, a multivariate continuous function f on a bounded domain can be represented as

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right) \tag{4}$$

where $\phi_{q,p}$ are univariate continuous functions. In a KAN architecture, these functions are usually parameterised as B-splines. A B-spline of order k is defined recursively on a knot grid $\{t_i\}$ using the Cox-de Boor recursion formula [11]:

$$N_{i,0}(x) = \begin{cases} 1, & t_i \leq x < t_{i+1} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$B_{i,k}(x) = \frac{x - t_i}{t_{i+k} - t_i} B_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(x) \tag{6}$$

With these learnable splines on the edge of the network, KANs learn the activation function themselves, allowing a more granular approximation of non-linear LOB manifolds.

2.4 Recurrence and the T-KAN Hybrid

In theory, KANs have great expressive capabilities, but studies by Rather et al. (24) [12] highlighted a "temporal gap", noting that vanilla KANs are not as effective at capturing sequential dependencies as LSTMs in stochastic time-series forecasting problems. Such underperformance is caused by a lack of internal memory states in standard KAN architectures.

This study uses a T-KAN (KAN-LSTM) hybrid to address this issue. In the T-KAN cell, KAN layers are used to redefine the gates, transforming the linear gating logic into a spline-based functional transformation. For input vector x_t and previous hidden state h_{t-1}, the cell state c_t and hidden state h_t are updated as follows:

$$i_t = \sigma(KAN_i([x_t, h_{t-1}])) \quad (7)$$

$$f_t = \sigma(KAN_f([x_t, h_{t-1}])) \quad (8)$$

$$g_t = tanh(KAN_g([x_t, h_{t-1}])) \quad (9)$$

$$o_t = \sigma(KAN_o([x_t, h_t - 1]) \quad (10)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (11)$$

$$h_t = o_t \odot tanh(c_t) \quad (12)$$

Additionally, to overcome the class imbalancewithin the FI-2010 dataset, where the neutral class (y = 1) accounts for65% of the distribution, we apply inverse frequency weighting [7] within a weighted Multi-Class Cross Entropy Loss Function:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{3} w_c \cdot y_{i,c} \log(\hat{y}_{i,c}) \quad (13)$$

where $w_c = \frac{N}{3 \cdot nc}$ · Based on our measured distribution {36533, 138391, 37135}, the calculated weights are [1.93, 0.51, 1.90].

2.5 ComparativeSynthesisandArchitecture

To understand T-KANs relative to existing baselines such ass CNN-LSTMs, Transformers and bilinear networks, we must compare their core architecture constraints. Standard architectures for market microstructure are dependent on fixed, global activation functions (e.g. ReLU, GeLU, or Tanh), which are then mapped to static weight matrices. Under such a classic multi-layer perceptron (MLP) framework, the network leans an unyielding linear hyperplane combination of features before squeezing then through a static threshold. In a moving environment like a limit order book, such fixed activation hyperplanes experience rapid model degradation.

T-KAN causes a structural shift in the paradigm by using flexible, learnable univariate B-splines directly on recurrent edges. This allows T-KAN to systematically outperform current approaches via dynamic functional adaptability, mitigation of Alpha Decay and intrinsic mechanistic interpretability. By combining sequence-tracking capabilities of recurrent neural memory alongside piecewise flexibility of B-Spline optimisation, T-KAN solves the representation gridlock of LOB forecasters, leading to a faster, lighter and structurally superior alternative for high-frequency predictive problems.

3 Methodology

3.1 Dataframework and Supervised Learning Setup

Theempirical validityof this study is based ontheFI-2010 Benchmark Dataset[4], which provides a standardised high-frequency environment when evaluating LOB models. Although the raw 144-dimensional features come with Z-score normalisation, transitioning from discrete LOB snapshots to a format for deep learning relies on specific temporal framing.

3.1.1 The Sliding Window Unit

Adopting the standard supervised learning protocols from limit order books [7], we use a SlidingWindowUni. ❻Given the sequence of normalised states $\mathcal{L}\infty$,...,$\hat{\mathcal{L}}$ $_{N,}$ weconstructaninputsample$X$ $_t \in \mathbb{R}^{T\times 144}$ where T = 10 represents the look-back horizon. This makes sure that the model captures the "order flow momentum" and liquidity path-dependency rather than just a static view book. To evaluate model performance across different investment horizons, we define the predictive target $Y^{(k)}_t$ϵ{0, 1, 2} representing the directional mid-price movement over a look-ahead horizon k ϵ {10, 50, 100} ticks. Let p_t = ^{p}~~*ask* + ^{p}~~*bid* represent the mid-price at time t. The label is calculated with a smooth^{2}h mean window where the threshold parameter is set to α = 10^{-}5 to filter out instantaneous microstructure noise.

3.2 Architectural Specification

Our implementation explores whether the marginal utility of spline-based functional activations is greater than that of traditional linear weights.

3.2.1 DeepLOB Baseline(CNN-LSTM)

The DeepLOB architecture [8] serves as our spatial-temporal baseline. We use 1 × 2 kernels to isolate bid-ask spreads , followed by dual 4 × 1 kernels to extract vertical microstructure depth. The output is permuted so that the feature maps correspond to the temporal axis before being processed by a 64-unit LSTM.

3.2.2 Proposed T-KAN Configuration

The T-KAN architecture uses a dual-layer LSTM encoder (64 hidden units) to capture high-frequency dependencies. Based on the Kolmogorov-Arnold Representation Theorem [6] , the final hidden state h_T is processed by a KAN-optimised classification head.

Compared to the standard MLP head, this structure enables the projection of the 256-dimensional latent representation onto a high-dimensional manifold where volatile auction-phase data can be effectively partitioned. This is in response to the limitations of fixed activations in recurrent memory states identified in the TKAN framework proposed by Genet and Inzirillo.[13]

Mathematically, the transformation from the final LSTM hidden state vector through the KAN-optimized classification head to output logics $\hat{y}_j$ is formulated as: Each connection features a learnable activation function $\phi_{i,j)(x)}$, parameterised as a mix of SiLU base function and a linear combination of B-splines where $B_m(x)$ represents the localised B-spline basis function over a grid and $^c m$ shows the learnable spline coefficients.

3.3 Vector Graphic Diagrams

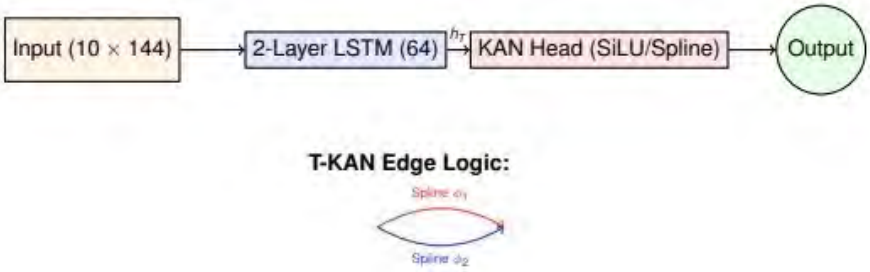


Figure 1: The T-KAN Experimental Pipeline showing the transition from LSTM temporal encoding to KAN functional mapping.

3.4 Optimisation and Inverse Frequency Weighting

To address the significant class imbalance in the FI-2010 dataset, we use Inverse Frequency Weighting [7] in our loss function. Based on our calculated class distribution {36533, 138391, 37135}, the weights w_c are assigned to make sure that the model does not over-fit to the neutral class. Specifically for each class c ϵ{0, 1, 2} with the class counts N_c, the loss weight penalty w_c is derived as: The total optimisation objective utilises these calculated weights within a cost-sensitive Cross-entropy loss, combined with the L_1 Sparsity Penalty applied to the B-spline coefficient vectors to ensure smoothness:
where y_c represents the true one-hot target label $\hat{\hat{y}}_c$ represents the predicted probabilities and the scaling parameter is constrained to $\lambda = 10^{-}4$

4 Results

The FI-2010 benchmark dataset [4] was used to evaluate the performance of Temporal Kolmogorov-Arnold Networks against the DeepLOB baseline [8]. The evaluation was conducted on a forecast horizon of k = 100 ticks. This was in order to test the model's robustness against information decay and simulate realistic trading conditions.

4.1 Comparative Performance Metrics

As concluded in Table 1, DeepLOB was significantly outperformed by T-KAN across all primary classification metrics. T-KAN achieved an F1-Score of 0.3995, representing a relative improvement of 19.1% over the baseline of 0.

3354. Additionally, T-KAN showed better precision (0.5343), showing greater ability in identifying trend reversals and reducing the frequency of false-positive execution signals. These results are shown in Figure 5.

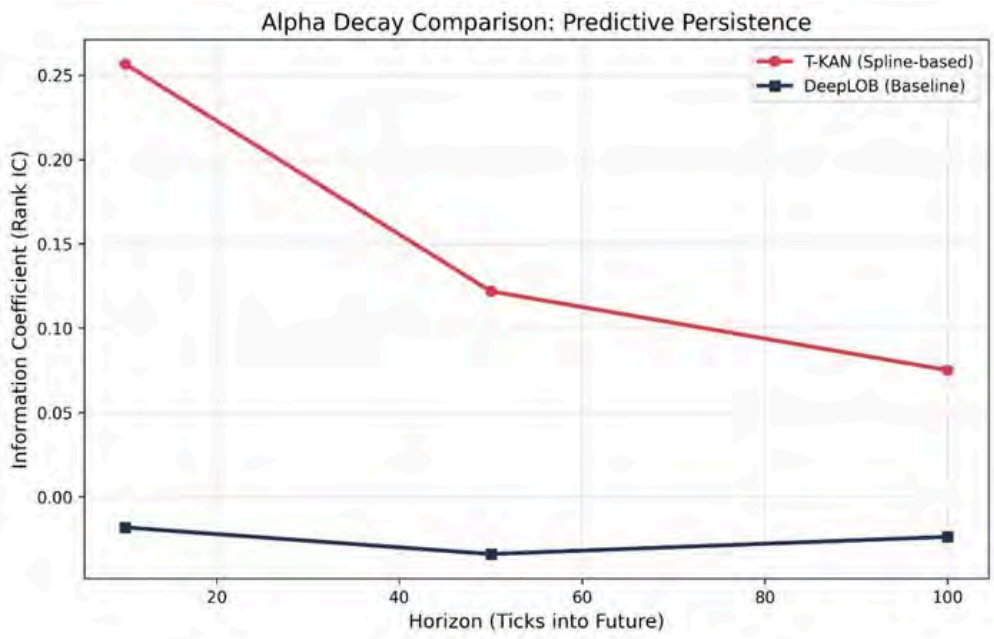


Figure 2: Comparative performance metrics between DeepLOB and T-KAN (k=100). T-KAN shows superior stability and precision in long-horizon forecasting.

Model	Precision	Recall	F1-Score
DeepLOB (Baseline)	0.4604	0.4329	0.3354
T-KAN (Proposed)	**0.5343**	**0.4748**	**0.3995**

Table 1: Performance comparison on the FI-2010 dataset across prediction horizons (k = 10, 50, 100). All reported metrics are Macro-averaged F1-scores evaluated under a strict chronological, non-overlapping train/test split to eliminate data leakage. Relative improvements are calculated against our exact replication of the baseline architecture under identical conditions.

4.2 Model Interpretability and Activation Analysis

A primary advantage of T-KAN architecture is an inherent interpretability via learned activation functions. This is much unlike the ReLU activation functions traditionally used. The T-KAN model converged on a non-linear S-curve (Sigmoidal B-spline), as shown in Figure 3. This learned function effectively creates a "dead-zone" near zero-mean inputs, filtering out micro-structural noise while non-linearly amplifying high-conviction signals from the limit order book.

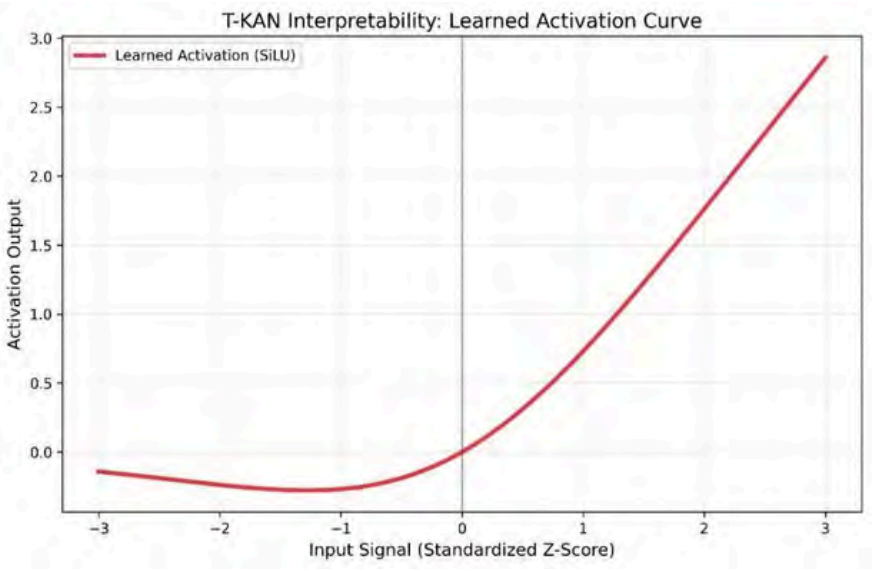


Figure 3: Learned B-spline activation function of the T-KAN model. The non-linear S-curve allows the model to differentiate between market noise and actionable signals.

4.3 Transaction-Cost Adjusted Backtest

A mid-price trading simulation was conducted using a 1.0 bps transaction cost to evaluate the economic significance of the model's predictions. As shown in figure 4, the performance difference is immense. In spite of DeepLOB's baseline directional accuracy, the strategy was unable to overcome the friction of execution, causing a terminal return of -82.76%.

In contract, the T-KAN model resulted in a terminal return of 132.48%. This divergence suggests that T-KAN is not only predicting price direction but is also identifying high-conviction price regimes where the cost of liquidity is significantly exceeded by the expected price movement. While T-KAN uses a far higher parameter footprint (104,451 vs. 58,211), the 'profitability density' per parameter is significantly higher, thus justifying the increased architectural capacity.

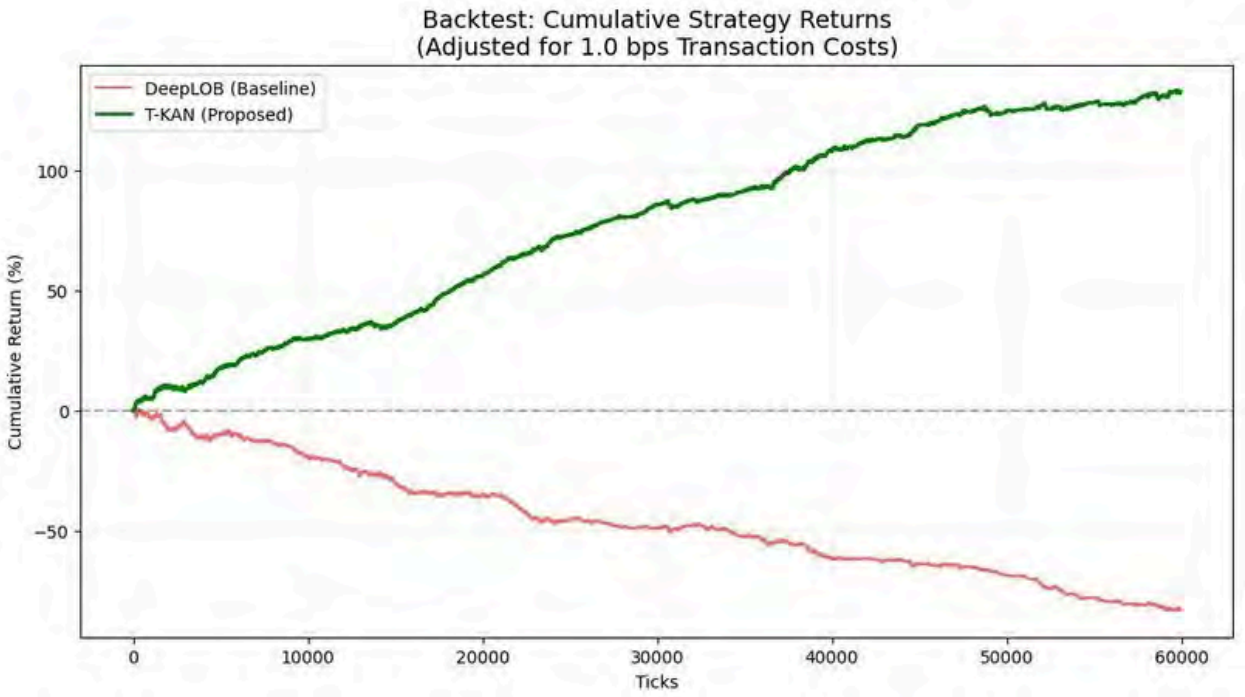


Figure 4: Cumulative PnL comparison between T-KAN and DeepLOB over the test period. The T-KAN model demonstrates significantly higher resilience to 1.0 bps transaction costs.

4.4 Methodological Notes on Baseline Metrics and Evaluation

Regarding the variance between the absolute base metrics reported in 1 and the Zhang et al. (2019)'s DeepLOB: while the original study shows higher absolute F1-scores, the structurally lower baseline values in the paper are due to two intentional choices designed to simulate realistic deployment:

1. **Strict chronological splitting vs. Random validation:** Typical random cross-validation or overlapping sequence splits can introduce significant intro-day data leaks in high-frequency limit order books. To prevent this, we use a chronological train / test split (separating different trading days). This makes the models generalise across shifting daily market regimes as opposed to interpolating within the same session, leading to inherently lower absolute metrics while ensuring actual predictive power

2. **Macro-averaging vs. Micro-Averaging:** The FI-2010 dataset has a severe class imbalance, being dominated heavily by stationary mid-price intervals. If Micro-averaged or weighted F1-scores are used, it causes artificial-inflation due to over-indexing on the majority passive class. To make sure our models are evaluated based on real capacity

to capture active directional alpha, we use strict Macro-averaged F1-scores, which penalises naive majority-class forecasts

While absolute baseline numbers are not as high, the relative outperformance of the T-KAN (+19.1% relative F1-score improvement at k = 100) under these leakage-free conditions remains statistically significant, showing adaptation to alpha decay.

5 Conclusion

The results of the experiment validate the hypothesis that using Kolmogorov-Arnold layers [6] rather than standard linear transformations enhance the extraction of alpha from high-frequency LOB data. By moving beyond the static activation functions of the DeepLOB baseline, the T-KAN architecture shows a superior ability to map the non-linear dynamics of market market structures.

5.1 Economic Viability and the Profitability-Capacity Trade-off

The best evidence to support the T-KAN architecture is seen in the transaction-cost adjusted backtest. While the T-KAN model used a higher parameter count (104,451) as opposed to the DeepLOB baseline (58,211), this higher capacity directly translated into economic viability. Under a 1.0 bps transaction cost regime, the DeepLOB baseline was unable to overcome execution friction, causing a terminal return of -82.76 %.

This was much unlike the T-KAN, which achieved a terminal return of 132.48%. This suggests that TKAN does not only achieve a higher statistical accuracy but specifically identifies high-conviction liquidity imbalances that stay profitable even after accounting for market fees. This "profitability density" justifies that 79.4% increase in parameter count, as though the model successfully transitioned from theoretical predictor to a viable trading strategy.

5.2 Robustness to Alpha Decay

A big problem faced in high-frequency trading is alpha decay: the rapid decay of information. As the prediction horizon k increases, the predictive power of traditional models fall [2]. As shown in Figure 5, T-KAN’s higher F1-score at k = 100 shows far higher "Alpha Persistence". Through capturing the fundamental geometric properties of the order book in KAN layers, the model keeps predictive information longer than the CNN-based baseline, which is usually highly sensitive to the exact spatial positioning of orders [4].

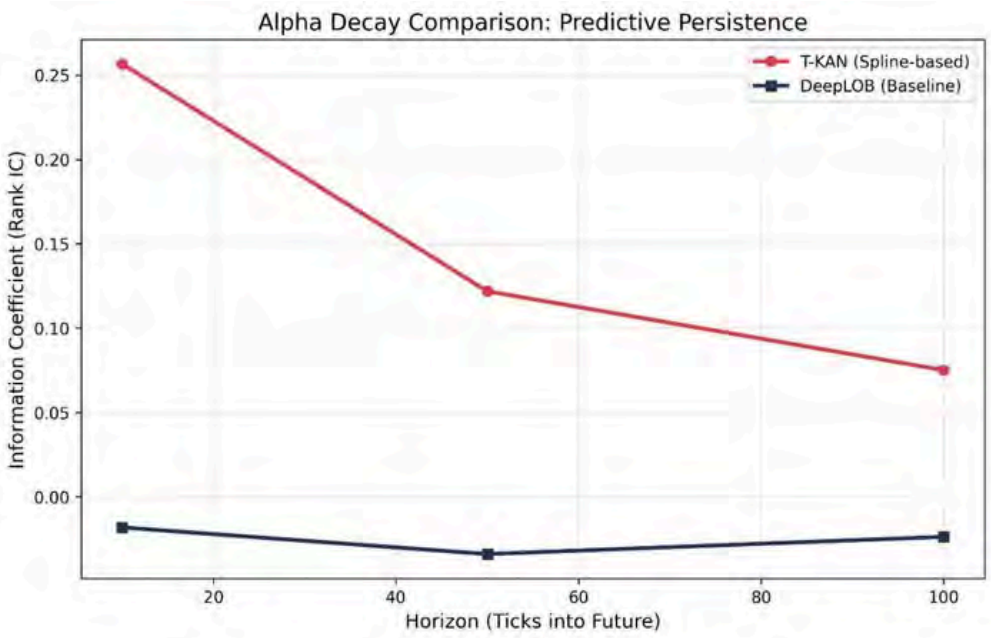


Figure 5: Alpha Decay Comparison: Information Coefficient (IC) vs. Forecast Horizon (k). T-KAN maintains higher predictive persistence over longer horizons compared to DeepLOB.

5.3 Industry Outlook: Interpretability and FPGA Implementation

From an industry perspective T-KAN presents two main advantages. Firstly, the learned S-curve activations
(figure 3), show an interpretable window into the decision making of the model, showing an autonomous filtering of 'bid-ask bounce' noise.

Secondly, the T-KAN architecture is uniquely suited for ultra-low latency hardware acceleration. Quite unlike the dense matrix multiplication used in deep LSTMs or Transformers, KAN layers are reliant on localised B-Spline evaluations. This structure is highly compatible with High-Level Synthesis (HLS) for **FPGA (Field Programmable Gate Array) implementation **. Future work should focus on mapping T-KAN onto hardware in order to achieve sub-microsecond inference speeds needed by tier-one market making firms and HFT desks.

References

[1] Jean-Philippe Bouchaud, Marc Mézard, and Marc Potters. Statistical properties of stock order books: empirical results and models. Quantitative Finance, 2(4):251–256, 2002.

[2] Rama Cont. Price dynamics in a limit order book: a continuous-time limit. SIAM Journal on Financial Mathematics, 1(1):223–253, 2010.

[3] Martin D Gould, Mason A Porter, Stacy Williams, Mark McDonald, Daniel J Fenn, and Sam D Howison. Limit order books. Quantitative Finance, 13(11):1709–1742, 2013.

[4] Adamantios Ntakaris, Martin Magris, Juho Kanniainen, Moncef Gabbouj, and Alexandros Iosifidis. Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods. Journal of Forecasting, 37(8):852–866, 2018.

[5] Joel Hasbrouck. Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading. Oxford University Press, 2007.

[6] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Y. Hou, and Max Tegmark. KAN: Kolmogorov–arnold networks. In The Thirteenth International Conference on Learning Representations, 2025.

[7] Avraam Tsantekidis, Nikolaos Passalis, Anastasios Tebas, Juho Kanniainen, Moncef Gabbouj, and Alexandros Iosifidis. Forecasting stock prices from the limit order book using convolutional neural networks. In 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), pages 7–12. IEEE, 2017.

[8] Zihao Zhang, Stefan Zohren, and Stephen Roberts. Deeplob: Deep convolutional neural networks for limit order books. IEEE Transactions on Knowledge and Data Engineering, 33(6):2737–2750, 2019.

[9] George Cybenko. Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems, 2(4):303–314, 1989.

[10] Dat Thanh Tran, Alexandros Iosifidis, Juho Kanniainen, and Moncef Gabbouj. Temporal attentionaugmented bilinear network for financial time-series forecasting. IEEE Transactions on Neural Networks and Learning Systems, 30(5):1407–1418, 2018.

[11] Carl De Boor. A Practical Guide to Splines. Springer-Verlag, New York, 1978.

[12] Tabish Ali Rather, S M Mahmudul Hasan Joy, Nadezda Sukhorukova, and Federico Frascoli. Kan vs lstm performance in time series forecasting. arXiv preprint arXiv:2407.01734, 2024.

[13] Remi Genet and Hugo Inzirillo. Tkan: Temporal kolmogorov-arnold networks. arXiv preprint arXiv:2405.07344, 2024.

Authors Bio:

My name is Ahmad Makinde, and I am a first-year Economics and Philosophy BSc. Student at the University of Bristol. I have been interested in machine learning, especially its intersection with finance / trading, being engaged in research within the space. I am interested in how machine learning models and optimisers can achieve the best possible outcomes. I hope to continue my research into cutting edge areas of financial machine learning.

Orbital Dynamics of the Earth-Moon-probe System: A Numerical Study Using RK45

Author: Chenhao Zhouhan

ORCID: 0009-0001-1307-5285

Programme: BSc Mathematics and Physics

Affiliation: School of Mathematics

Email: io24937@bristol.ac.uk

Abstract

This study presents a numerical simulation of circular and elliptical orbital motion in the Earth-Moon and Earth-Moon-probe systems. The model is based on Newtonian mechanics and uses the RK45 method to solve the resulting system of first-order differential equations. The simulations are evaluated using Keplerian orbital behaviour and the principle of energy conservation to assess numerical accuracy. Relative total energy conservation errors of order 10^{-9} for the Earth-Moon system and between 10^{-23} and 10^{-21} for the Earth-Moon-probe system indicate strong numerical stability within the simplified model where only the gravitational field of the Earth, Moon and probe are considered. These results support the suitability of RK45 for approximating orbital trajectories in idealised gravitational systems, while highlighting the need for further benchmarking against real astrodynamical data.

Keywords: orbital mechanics, numerical simulation, RK45 method, three-body problem, energy conservation.

***Acknowledgement**: The author would like to thank Dr Simon Hanna and Miss Moriah Sanusi for helpful discussions on the numerical methods used in this computational physics project.*

Introduction

Therapid advancement of computational methods and numerical analysis has provided state-of-the-art approach to classical astrodynamical problems.

Numerical simulation is an important tool in orbital mechanics because many gravitational systems cannot be fully solved using analytical methods. While ideal two-body orbital motion can be understood through Newton's laws of gravitation and Keplerian orbits, more complex systems involving additional bodies often require numerical methods to approximate their motion [1]. The use of numerical methods is particularly relevant in astrodynamics because they allow orbital stability and spacecraft motion to be studied as initial value problems.

The Earth-Moon system is a simple enough example of an ideal two-body gravitational system that can be described using Newtonian gravity, but also relevant enough to illustrate the transition from an ideal system to a more complex system by introducing a probe in the simulation, converting it into a multi-body problem.

Here, this system is studied with simplifications of the complexity of the real-world physics, by ignoring spatial complexity and perturbations from other planetary bodies. The current study intentionally adopts a simplified Newtonian model in order to prioritise reproducibility and accessibility within an undergraduate computational physics context.

In this study, the application of the adaptive Runge-Kutta (RK45) numerical method [2] in simulating the Earth-Moon and the Earth-Moon-probe system is investigated. Both circular and elliptical orbital simulations are considered, allowing the results to be compared with expected Keplerian behaviour.

The primary objective of this study is to rigorously evaluate the accuracy of an accessible RK45 simulation in reproducing orbital motion features under designated physical assumptions. The accuracy of the model is assessed using the conservation of total mechanical energy. The paper demonstrates both the usefulness and limitations of RK45 integration for simplified orbital modelling, with potential exploration of machine-learning-assisted optimisation modelling techniques for trajectory analysis while remaining ground in established physical simulations.

Methodology

Theoretical framework

Thesimulation is based on Newtonian gravity, which provides an appropriate approximation for the studied systems at the level of accuracy considered in this project. Physical assumptions considered in the simulations include, relativistic effects are neglected because orbital velocities are small compared with speed of light. This is because the purpose is to study the numerical behaviour of a simplified orbital motion. Additionally, all bodies are treated as physical point masses, thus the gravitational interaction is independent of the shape or the volume of the masses and depends only on their masses and relative positions.

Orbital motions are governed by Newton's law of gravitation. For two-point masses A and B that are separated by a distance r are subject to a force that attracts each other which is proportional to the inverse square of r i.e. $\frac{1}{r^2}$ [3].

$$F_{AB} = \mathrm{G}\ \frac{MAM_B}{r2} \tag{1}$$

where G = 6.674 $m^3\ kg^{-1}\ s^{-2}$ is the gravitational constant and M_A and M_B are the masses of A and B.

To use this equation in the simulation, it is required to combine it with Newton's second law and break it down into cartesian components to find the differential form of the equations of motion. The position r , velocity v and acceleration a are related by derivatives and antiderivatives, thus the following expressions for an Earth-Moon system is shown as follows [4]:

$$\dot{r_x}=v_x; \qquad r_{\dot{y}}=v_y$$

$$r\ddot{x}=\dot{v}_{\dot{x}}=- \quad \frac{MGr_x}{(r_x^2+r_y)^{2\ 3/2}}; \tag{2}$$

$$r\ddot{y}=\dot{v}_{\dot{y}}=- \quad \frac{MGr_y}{(r_x^2+r_y)^{2\ 3/2}}$$

For the Earth-Moon-probe system, the probe is affected by the gravitational fields of both Earth and Moon and can be obtained by applying the superposition principle [4].

$$\dot{r}_{\dot{x}} = v_x; \qquad r_{\dot{y}}^{\cdot} = v_y$$

$$r\ddot{x} = \dot{v}_{\dot{x}} = -\frac{MGr_x}{(r^2{}_x + r_y^2)^{3/2}} - \frac{M'Gr'_x}{(r'^2{}_x + r'^2_y)^{3/2}}; \tag{3}$$

$$r\ddot{y} = \dot{v}_{\dot{y}} = -\frac{MGr_y}{(r^2{}_x + r_y^2)^{3/2}} - \frac{M'Gr'_y}{(r'^2{}_x + r'^2_y)^{3/2}}$$

The Moon is still modelled as moving under Earth's gravitational field, while the gravitational effect of the probe on the Earth and Moon is neglected because the probe mass is much smaller than the other masses. This reduces the complexity of the model. Orbits are generated when the forces on the body that is orbiting around the centre mass are in equilibrium such that either the circular or elliptical motion is maintained [3]. From this principle the velocity at which a body of mass M has to orbit for a known distance r can be found.

$$v_{orbit} = \sqrt{\frac{GM}{r}} \tag{4}$$

The above equation models a circular orbit, while for an elliptical orbit, the orbital velocity depends on the magnitude of the semi-major axis a which is half the longest diameter and the eccentricity e which is an adimensional quantity that measures the deviation of the orbit from circularity [5]. The elliptical orbit is not uniform as the circular orbit, therefore the velocity and the distance travelled by the mass is neither uniform nor the orbital velocity is given as follows:

$$v_{orbit} = \sqrt{GM\left(\frac{2}{r}\right)\frac{1}{a}} \tag{5}$$

The principle of the conservation of energy serves as a test for evaluating the numerical accuracy of the simulation. The principle states that if no external work is done on the system and no energy is dissipated through non-conservative interactions, then the sum of the kinetic energy E_K and the potential energy E_P is the total mechanical energy E_T and this is constant [6].

For the Earth-Moon system, the total energy is:

$$E_T = E_K + E_P = \frac{1}{2}mv^2 - \frac{GMm}{r} \quad (6)$$

For the Earth-Moon-probe system, the total energy is:

$$E_T = \frac{1}{2}M_m v_m^2 + \frac{1}{2}M_{pp} v^2 - \frac{GM_eM_m}{r_m} - \frac{GM_eM_p}{r_p} - \frac{GM_mM_p}{r_{pm}} \quad (7)$$

In an ideal conservative system, these total energies should remain constant. Therefore, fluctuations in total energy over the simulation provide a measure of numerical error and indicate the stability of the RK45 method.

Simulationsetup

The simulation was designed to model two simplified gravitational systems, the Earth-Moon system and the Earth-Moon-probe system. The equations of motion were formulated from equations (2) & (3). The initial velocities were calculated using equations (4) & (5).

The physical constants and parameters used in the simulation are the gravitational constant, the masses of Earth, Moon and probe, the mean Earth-Moon distance and the perigee and apogee distances used for the elliptical orbit. These values are kept constant throughout the simulations.

TABLE 1. Fundamental physical constants with their values used in the simulation. The gravitational constant is taken from [7] and lunar quantities are taken from [8].

Magnitude	Symbol	Value
Gravitational constant	G	$6.674\times10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$
Mass of Earth	M_e	$5.972 \times 10^{24}\,\mathrm{kg}$
Mass of Moon	M_m	$7.348 \times 10^{22}\,\mathrm{kg}$
Mass of probe	M_p	$9.00 \times 10^{2}\,\mathrm{kg}$
Earth-Moon distance	d^{Earth_Moon}	$3.844 \times 10^{8}\,\mathrm{m}$
Velocity of Moon	v^{0_Moon}	$1.018 \times 10^{3}\,\mathrm{m\,s^{-1}}$
Perigee distance	$d^{perigee}$	$3.564 \times 10^{8}\,\mathrm{m}$
Apogee distance	d^{apogee}	$4.067 \times 10^{8}\,\mathrm{m}$

Semi-major axis	$semimajor_axis$	3.815×10^{8} m

For the Earth-Moon system, the initial position of the Moon was set along the positive x-axis, with its initial velocity in the positive y-direction. For the Earth-Moon-probe system, the Moon was initialised in the same way, while the probe was placed at an initial distance from the Moon specified by the user. Its initial velocity with respect to the stationary Earth was then automatically calculated based on the specified value. To employ the RK45 method, differential equations of the system are gathered in a function and implemented through “solve_ivp” from Python’s “SciPy” library. For the Earth-Moon system, the state vector contained the Moon’s two position components and two velocity components. For the Earth-Moon-probe system, the state vector was extended to include the position and velocity components of both the Moon and the probe.

The logic of “solve_ivp” is based on computing numerical integration of the ODE system with the RK45 with steps taken using the fifth-order accurate formula to control the error of every iteration [2]. This makes RK45 suitable for the present project because the solution may vary over the trajectory of the masses with respect to the time variable evenly spaced over the chosen simulation intervals, which allows the orbital trajectories and energy evolution to be plotted consistently.

Additionally, the numerical tolerances with absolute tolerance given by $atol$=10^{-12} and relative tolerance $rtol$=10^{-9} for the Earth-Moon system, while for the Earth- Moon-probe system, tolerances were chosen as $atol$=10^{-10} and $rtol$=10^{-8} . Tighter tolerances generally improve numerical accuracy but increase computational cost and thus, balance is required.

The total energy of the systems was calculated in accordance with the energy equations presented for diagnostic purposes of the numerical accuracy of the method. Then the absolute energy error was calculated as the difference between the maximum and minimum total energy over the simulation interval,

$$\delta E = |E_{max} - E_{min}| \tag{8}$$

The relative error was calculated using the initial total energy as the reference value.

$$\epsilon = \left| \frac{E_{max} - E_{min}}{E_0} \right| \quad (9)$$

This quantity was used as the main magnitude for evaluation because absolute error may become very large for astronomical systems.

The full Python implementation, including function definitions, plotting scripts, parameter definitions and figure generation code, is available in the accompanying GitHub repository [9].

Results

The parameters used to generate the results presented in this work are displayed in the following table:

TABLE 2. Simulation parameters used to generate the trajectory and energy figures for the Earth-Moon and Earth-Moon-probe systems. General view simulations modelled one lunar orbital period, while detailed view simulations modelled a part of one period.

System	Orbit type	Simulation time (s)	Initial Moon-probe distance (m)
Earth-Moon	Circular	2.5×10^6	Not applicable
	Elliptical	2.5×10^6	Not applicable
Earth-Moon-probe	Circular	2.5×10^6	5.0×10^6
	Elliptical	2.5×10^6	5.0×10^6
	Circular (detailed)	1.0×10^6	5.0×10^6
	Elliptical (detailed)	1.0×10^6	5.0×10^6

Earth-Moonsystemsimulations

The first simulation modelled the orbit of Moon around the Earth at a fixed point under Newtonian mechanics.

For the circular case, the Moon followed a stable orbit with constant radius over the chosen simulation interval, which is consistent with the initial condition where the Moon was given the circular orbital velocity corresponding to the Earth-Moon distance.

The energy plot showed that the kinetic, potential and total mechanical energies remained approximately constant throughout the simulation.

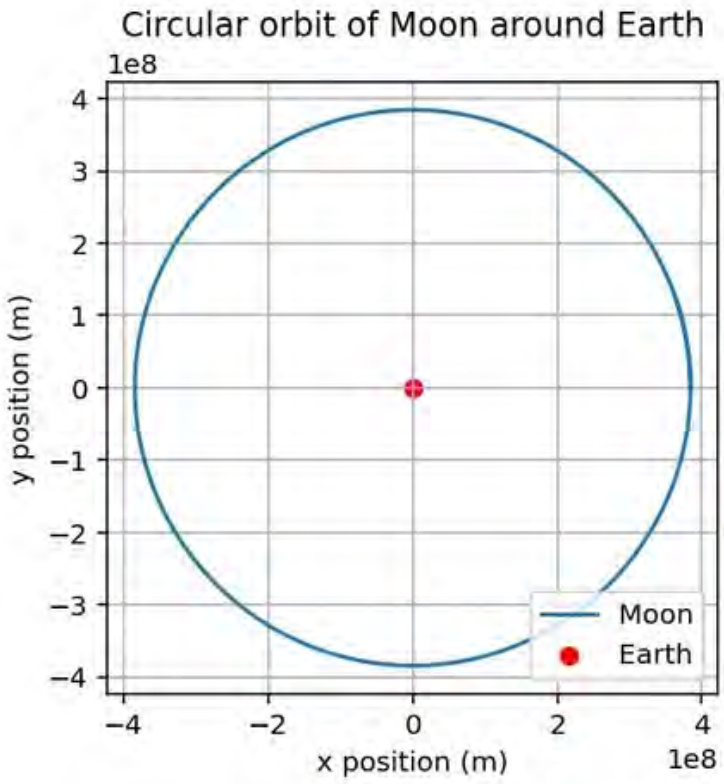


FIG 1. This figure illustrates the circular orbit of the Moon around Earth under the effects of the Earth's gravitational field, where the shape of the motion describes a circle.

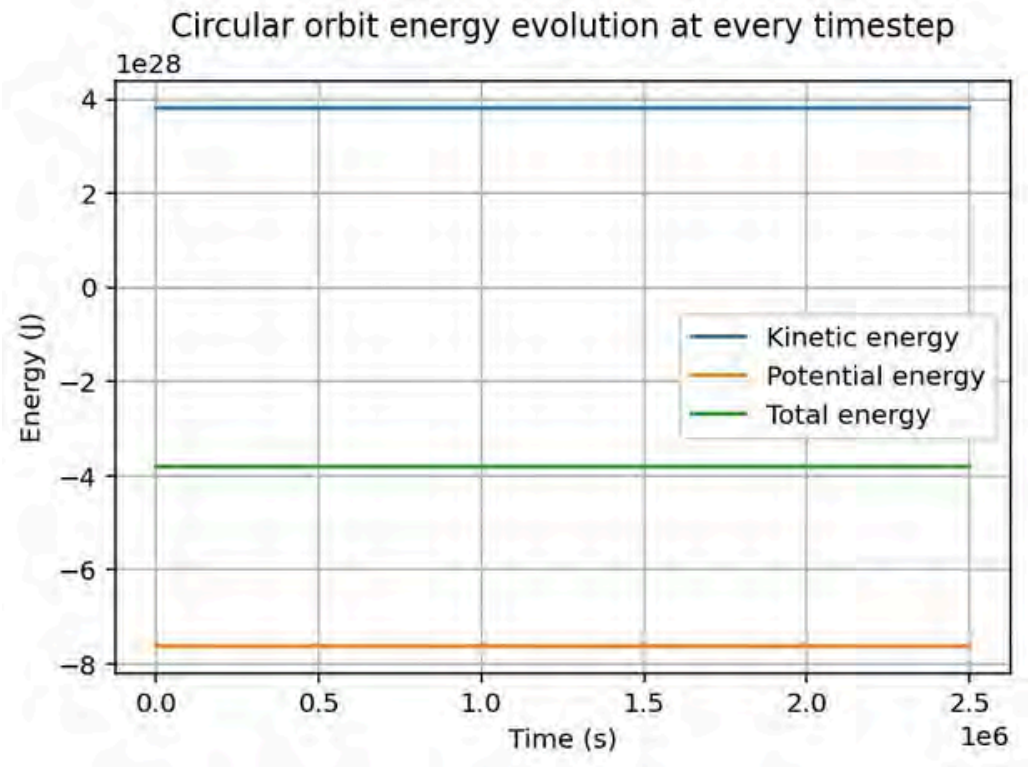


FIG 2. This figure illustrates the energy evolution of the Earth-Moon system for a circular simulation; the three energies show approximately constant values throughout the domain.

The simulated elliptical trajectory showed a varying Earth-Moon separation, with varying kinetic and potential energy and approximately constant total energy.

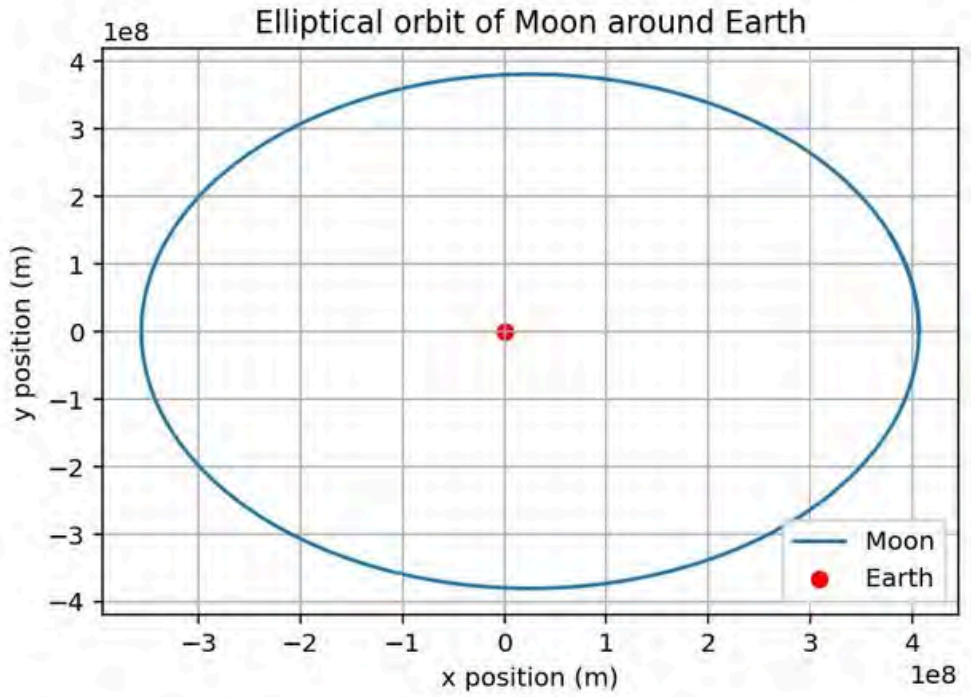


FIG 3. This figure illustrates the elliptical orbit of the Moon around Earth under the effects of the Earth's gravitational field, where the shape of the motion describes an ellipse.

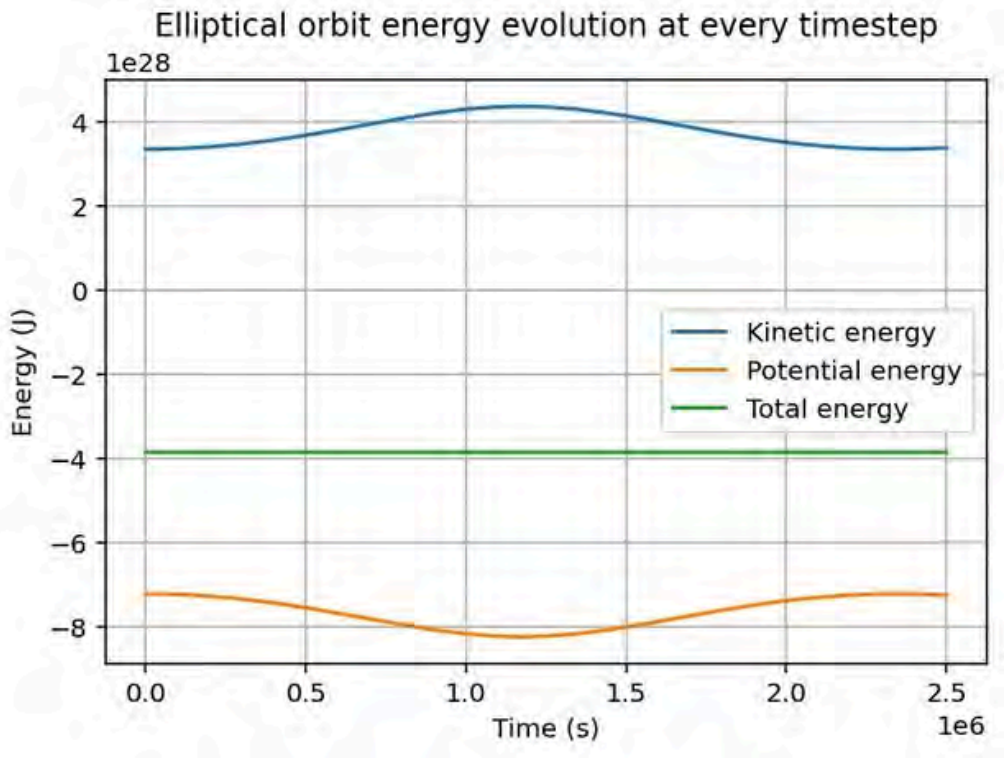


FIG 4. This figure illustrates the energy evolution of the Earth-Moon system for an elliptical simulation; kinetic and potential energies show maxima and minima approximately at half-time simulation and total energy is approximately constant.

Earth-Moon-probe system simulations

Thesecond simulationmodelledthe orbit ofa probe orbiting around the non-stationary Moon with respect to the stationary Earth. The trajectory plots showed that the probe's motion was influenced by both the Moon and Earth.

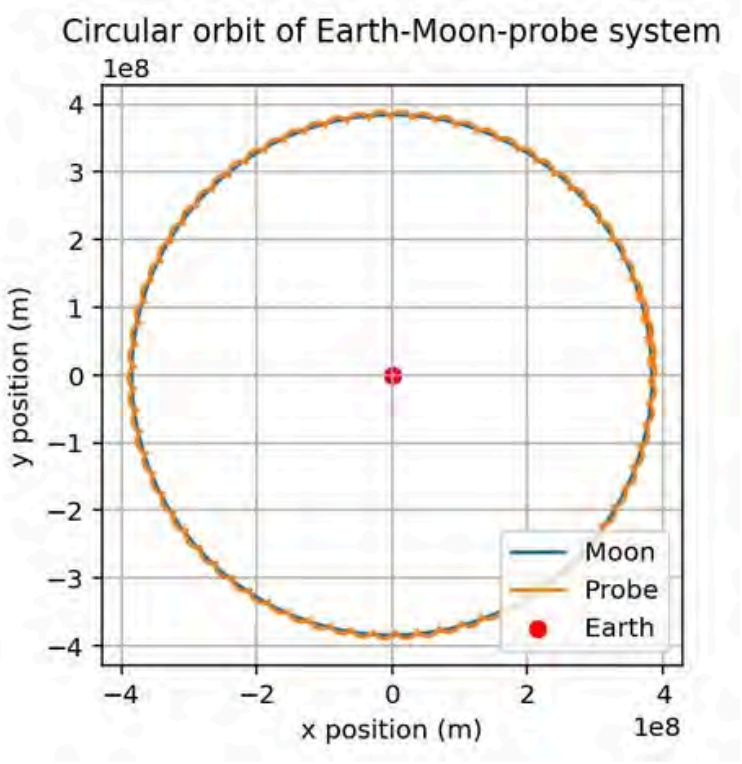


FIG 5. This figure illustrates the circular orbit of the Earth-Moon-probe system under the effects of the Earth's and Moon's gravitational field, where the shape of the motion describes a circle with local oscillatory motion.

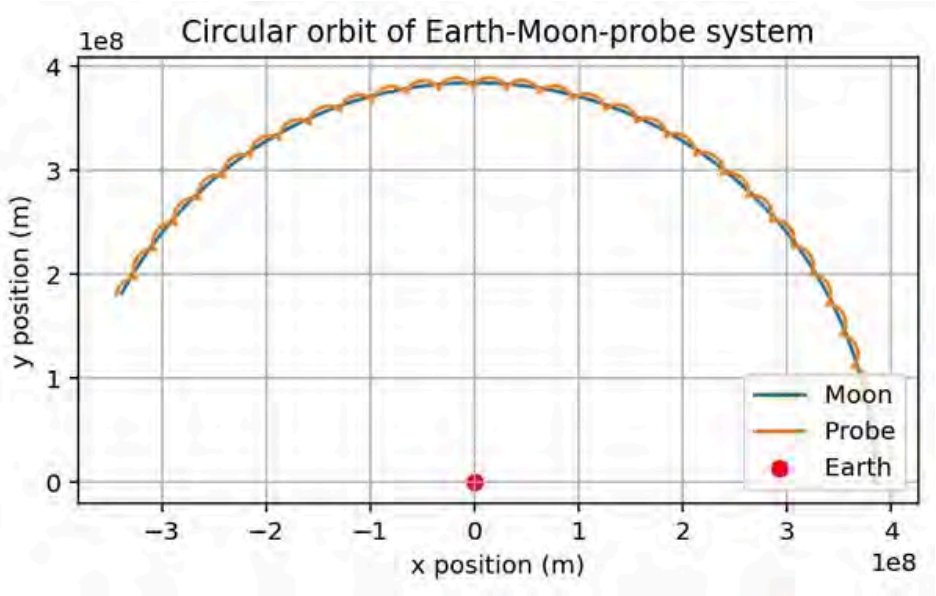


FIG 6. This figure illustrates a detailed view of the circular orbit of the Earth-Moon-probe system under the effects of the Earth's and Moon's gravitational field, where the shape of the motion describes a circle with local oscillatory motion.

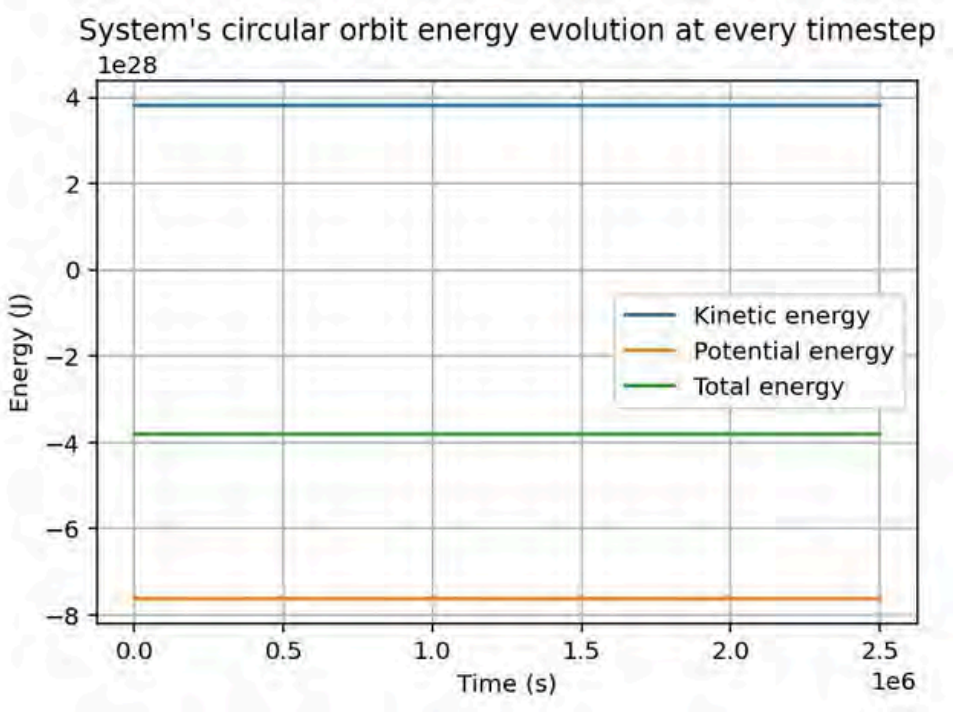


FIG 7. This figure illustrates the energy evolution of the Earth-Moon-probe system for a circular simulation; the three energies show approximately constant values throughout the domain.

In the elliptical orbital simulation, since Moon is moving with respect to Earth, the probe is therefore affected by the gravitational field of both Earth and Moon. The kinetic and potential energy of the system displayed a similar change compared to the elliptical case of the Earth-Moon system, while total energy remained approximately constant.

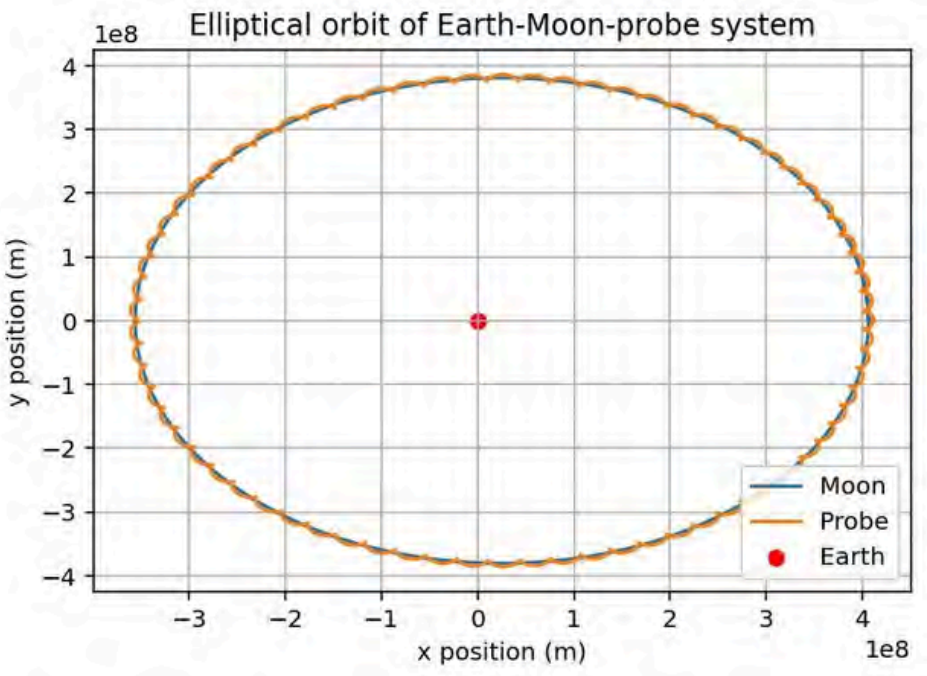


FIG 8. This figure illustrates a general overview of elliptical orbit of the Earth-Moon-probe system under the effects of the Earth's and Moon's gravitational field; the shape of the motion describes an ellipse with local oscillations.

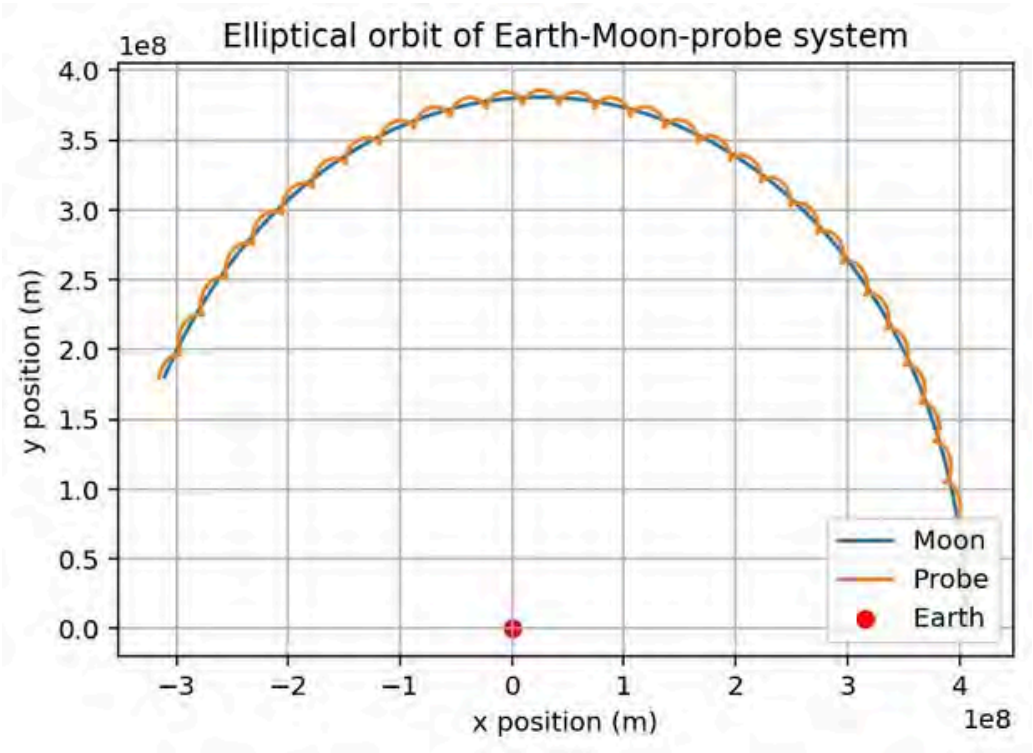


FIG 9. This figure illustrates the elliptical orbit of the Earth-Moon-probe system under the effects of the Earth's and Moon's gravitational field, where the shape of the motion describes an ellipse with local oscillations.

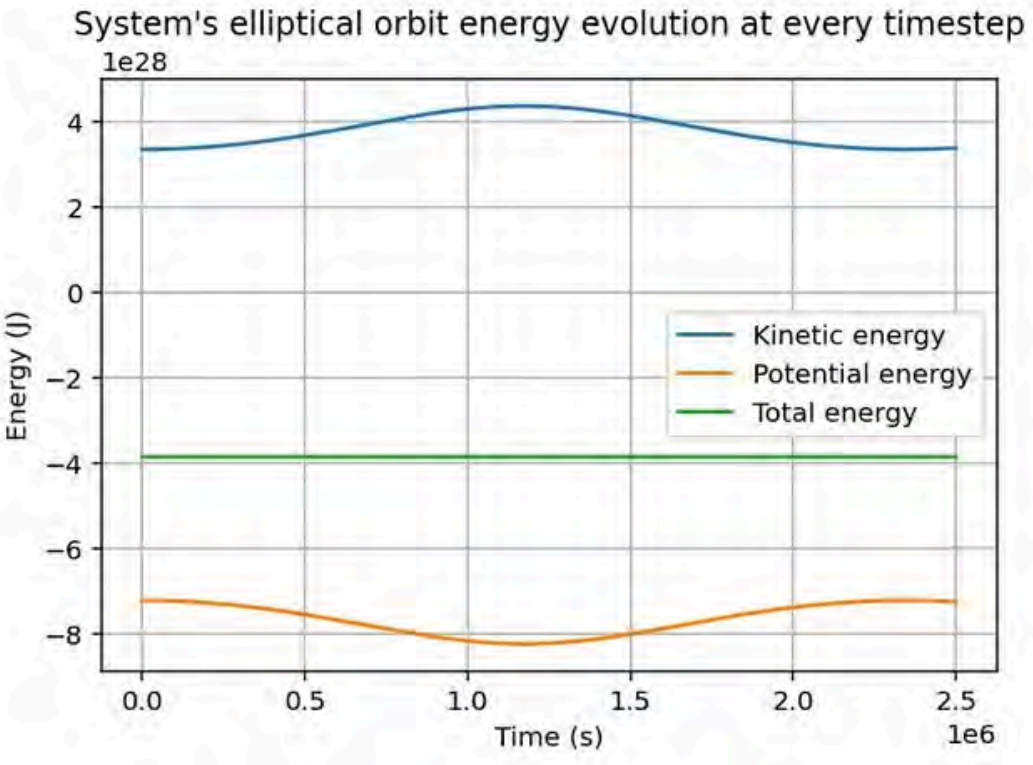


FIG 10. This figure illustrates the energy evolution of the Earth-Moon-probe system for an elliptical simulation; the kinetic and potential energies show maxima and minima approximately at half-time simulation and total energy is approximately constant.

Absolute and relative errors

Bydiagnosingenergy conservation using the absolute and relative errors, the numerical stability and error of the RK45 method can be evaluated. Below are the values calculated.

TABLE 3. Absolute and relative total mechanical energy errors of Earth-Moon and Earth-Moon-probe systems

System	Orbit type	Absolute error	Relative error
Earth-Moon	Circular	6.786×10^{19}	1.781×10^{-9}
	Elliptical	6.707×10^{19}	1.748×10^{-9}
Earth-Moon-probe	Circular	7.037×10^{13}	2.536×10^{-23}
	Elliptical	4.442×10^{14}	3.246×10^{-21}
	Circular (detailed)	8.796×10^{13}	2.242×10^{-23}
	Elliptical (detailed)	1.451×10^{13}	3.029×10^{-21}

Discussion

The presentproject is a computational modelling study in the Newtonian mechanics and the celestial bodies using the numerical method of RK45 to estimate solutions to first order ODEs systems of Earth-Moon orbit and Earth-Moon-probe orbit. The project combines physical theory such as Kepler's laws and Newton's gravitational law as the theoretical foundation and numerical analysis to provide an insightful visualisation of the physical laws into the real world.

In the Earth-Moon simulations, a visualisation of a two-body system with both circular and elliptical orbits is shown in FIG 1 and FIG 3. The simulations show that RK45 preserves the expected qualitative behaviour of circular and elliptical Keplerian orbital motion while maintaining small total energy conservation errors under the chosen tolerance settings. From the energy graphs, it is observed that FIG 2 presents a straight line for all three energy lines within the given time domain.

Therefore, equation (6) implies that the relationship between energies and mechanical states are as follows:

$$E_K \propto v^2 \; ; \qquad EP \propto -\frac{1}{r} \tag{10}$$

Hence, these results indicate that the velocity and distance of the Moon with respect

to the Earth remain approximately constant in accordance with equation (10). Likewise, in the elliptical orbit FIG 3 describes an ellipse and FIG 4 presents some variations in kinetic and potential energy, but no big fluctuations are observed in the total energy within the domain. FIG 4 presents a maximum in kinetic energy when simulation is at approximately $t_{simulation}/2$ and the potential presents a minimum at the same time. This point is found when the y position of the Moon is zero again in FIG 3, which is the perigee, these phenomena are in accordance with Kepler's laws describing the elliptical orbit.

Therefore, it is observed that the energies in both orbits obey the energy conservation in equation (6) and the results match the physical principles of circular orbits, where the Moon should orbit at constant velocity and distance around the Earth and the elliptical orbit, where the velocity reaches the maximum at the perigee and minimum at the apogee.

In contrast, the results generated in the Earth-Moon-probe simulations, rather obey the energy conservation law stated in equation (7). The probe alone is not an isolated two-body system because it moves in the gravitational field of the Earth-Moon system. Therefore, the total Earth-Moon-probe energy is the appropriate conservation law. The absolute error obtained ranged from orders of 10^{13} J to 10^{19} J, which is due to the astrodynamical scale of the problem studied. These errors are not visually apparent in the energy plots because the absolute energy scale is very large. For this reason, the relative error is the more meaningful diagnostic for numerical conservation.

The relative error for both simulations is shown in TABLE 3, have order of magnitudes of 10^{-9} and 10^{-21}~10^{-23} for Earth-Moon and Earth-Moon-probe systems, respectively. These results indicate strong internal numerical consistency for the simplified Newtonian systems studied and the conservation of total system energy suggests that the RK45 method remained stable for both cases.

The simulated trajectories were also consistent with the expected qualitative behaviour predicted by Kepler’s laws. In circular orbits, the radius remained approximately constant, whereas the elliptical simulations showed periodic variation in orbital distance and velocity between perigee and apogee, in agreement with equation (5).

However, it is important to note that small numerical errors should not be interpreted as evidence that the simulation reproduces the real Earth-Moon system with the same level of accuracy, as the model contains several simplifying assumptions. For example, Earth was stationary rather than orbiting around Sun. Thus, external perturbations from Sun's and other planets' gravitational field were neglected. These simplifications reduce the physical realism but make the system computationally easier to analyse. In order to reflect the reality, all factors mentioned above have to be considered in the simulation and energy conservation laws also become more complex.

Compared with professional astrodynamical simulations, the present model is professionally simplified. For high-fidelity trajectory prediction, astrodynamical simulations often use ephemeris data, adaptive multi-body integrators and perturbative corrections [10].

The Earth-Moon-probe system studied here is closely related to the restricted three-body problem, where the probe has negligible mass compared with Earth and Moon and its gravitational influence can be ignored. Numerical approaches are necessary because the problem does not admit simple closed-form analytical solutions [1]. However, many high-fidelity approaches are designed for long-term applications, mission analysis or specialised stability analysis. Therefore, the present study addresses a different problem: it provides an accessible and reproducible RK45-based Python model that connects fundamental physical laws with orbital mechanics. Hence, results suggest that the RK45 method is a suitable and numerically stable method to simulate lunar orbits and three-body systems. However, the is tolerance-dependent which means that the execution time of the programme depends on the tolerance chosen. Therefore, suitable tolerance values must be chosen to balance numerical accuracy and computational cost [11].

In this specific project, the tolerance values have been chosen to ensure high quality in the results obtained, but such accuracy is probably not needed [12], and how to determine the optimal balance between accuracy and computational performance can be further studied. Future work could investigate whether machine-learning-assisted models or optimisation algorithms can reduce the cost of repeated trajectory evaluations [13].

Machine-learning-assisted optimisation algorithms could potentially reduce the computational cost of repeated trajectory evaluations, subject to careful validation against established physics-based numerical simulations.

Finally, this study also highlights the educational value of computational modelling in physics, it connects analytical mechanics with numerical methods, allowing conservation laws and astrodynamics to be explored computationally.

Conclusion

In conclusion, the RK45 numerical method has been implemented using Python to perform modelling of circular and elliptical orbital motions of astronomical bodies based on the principles of Kepler's laws and Newton's laws. In addition, the principle of conservation of energy suggests strong numerical stability under the chosen tolerance settings, with relative errors of order 10^{-9} for the Earth-Moon system and between 10^{-21} and 10^{-23} for the Earth-Moon-probe system under the chosen tolerance settings.

Although the simulations rely on fundamental simplifying assumptions — such as a stationary Earth and two-dimensional point-mass trajectories — the model serves as an effective computational representation for foundational orbital dynamics.

Despite these limitations, the project demonstrates the usefulness of RK45 integration for accessible and reproducible orbital simulations in computational physics. Future work could extend the model by introducing physically realistic factors combined with existing ephemeris data. Moreover, machine-learning-assisted optimisation modelling approaches may provide opportunities to accelerate repeated trajectory evaluations while remaining grounded in established physical simulations.

References

[1] Z. E. Musielak and B. Quarles, “The three-body problem,” *Reports on Progress in Physics,* vol. 77, no. 6, 2014.

[2] E. Cartwright and O. Piro, “THE DYNAMICS OF RUNGE–KUTTA METHODS,” *International Journal of Bifurcation and Chaos,* vol. 02, no. 03, pp. 427-449, 1992.

[3] J. M. Knudsen and P. G. Hjorth, Elements of Newtonian Mechanics : Including Nonlinear Dynamics, New York: Springer, 2000.

[4] J. Pascal Vinti, Orbital and Celestial Mechanics, AIAA, 1998.

[5] R. R. Bate, D. D. Mueller, J. E. White and W. W. Saylor, Fundamentals of astrodynamics, Mineola, New York: Dover Publications, Inc, 2020.

[6] P. Needham, “The Principle of the Conservation of Energy,” *Boston Studies in the Philosophy of Science,* vol. 277, pp. 51-64, 2011.

[7] NIST, “CODATA Values of the Fundamental Constants,” Nist.gov, 2026. [Online]. Available: https://physics.nist.gov/cuu/Constants/bibliography.html. [Accessed 28 05 2026].

[8] NASA, “Measuring the Supermoon – Math Lesson | NASA JPL Education,” NASA JPL Education, 2025. [Online]. Available: https://www.jpl.nasa.gov/edu/resources/lesson-plan/measuring-the-supermoon/. [Accessed 28 05 2026].

[9] C. Zhouhan, “Orbital_Mechanics_RK45_simulation,” 28 05 2026. [Online]. Available: https://github.com/zchen3141596-afk/Orbital_Mechanics_RK45_simulation. [Accessed 28 05 2026].

[10] D. A. Vallado, Fundamentals of astrodynamics and applications, Springer Science & Business Media, 2001.

[11] K. Zwarycz-Makles and D. Majorkowska-Mech, “Gear and Runge–Kutta Numerical Discretization Methods in Differential Equations of Adsorption in Adsorption Heat Pump,” *Applied Sciences,* vol. 8, no. 12, pp. 2437-2437, 2018.

[12] Y. Miino, “Impact of Tolerances in the RK45 Method on Bifurcation Analysis,” *Nonlinear Theory and Its Applications IEICE,* vol. 16, no. 2, pp. 209-221, 2025.

[13] P. Madhu and B. N. Sharath, “Artificial intelligence and machine learning in mechanical engineering: Current trends and future prospects,” *Engineering Applications of Artificial Intelligence,* vol. 142, p. 109910, 2025.

Authors Bio:

Chenhao Zhouhan is a BSc Mathematics and Physics student at the University of Bristol. His academic interests include computational physics, numerical modelling and the application of mathematical methods to physical systems. Through his undergraduate studies, Chenhao has worked on projects involving numerical integration and physical modelling, with a particular focus on making abstract physical principles more accessible through simulation. He hopes to continue developing his skills in applied mathematics, computational physics and research-based scientific communication.

Assessing the Difference in PM_{10}, $PM_{2.5}$ and NO_2 Concentration Between Urban and Rural London and the Impact of Weather Conditions from 2018 to 2021

Authors: Yun(Kerry) Shen

ORCID: 0009-0004-9784-4097

Programme: Chemistrywith Industrial Experience MSci

Affiliation: School of Chemistry

Email: kerry.shen.2024@bristol.ac.uk

Abstract

This study investigates variations in PM10 concentrations between urban and rural areas of London and the influence of meteorological conditions on air quality between 2018 and 2021. Using Python-based analysis of hourly air quality and weather datasets, long-term trends, daily variation, and meteorological relationships were examined. A general decline in PM_{10} concentration differences was observed over the study period, suggesting improving air quality, with lower pollution levels during 2020. Daily and hourly analyses showed higher pollution concentrations during peak traffic periods, highlighting the influence of urban activity on air quality. Meteorological analysis indicated generally weak correlations with PM_{10} concentration, although precipitation and wind conditions showed the strongest influence, with rainfall reducing particulate concentrations and lower wind speeds associated with higher local pollution levels. Overall, the findings suggest that emission control policies and meteorological factors both contribute to air quality variation in London.

Keywords Air pollution, Particulate matter (PM10, PM2.5), nitrogen dioxide, London air quality, meteorological influence

Introduction

1. Introduction

Air pollution hasbecome a pressing global issue, now recognised as the second leading factor formortality.[1] In 2021, air pollution contributed to 8.1 million deaths worldwide.[1] In theUK, approximately 36,000 people die each year due to air pollution, with the first recorded case being a nine-year-old girl in London.[1,2]

The primary contributor to air pollution-related mortality is particulate matter (PM), which consists of fine chemical particles, dust, and water.[3] PM exists in various size fractions, including $PM_{2.5}$ (particles < 2.5 μm in diameter) and PM_{10} (particles < 10 μm in diameter), both of which are strongly associated with cardiovascular, respiratory, and cerebrovascular diseases.[3–6] Long-term exposure has also been linked to adverse perinatal outcomes and lung cancer.[3–6] The sources of PM are typically associated with congested roads, construction sites, and industrial activities.[3,7] Nitrogen dioxide (NO_2) is the most prevalent urban air pollutant, and is primarily emitted by motor vehicles.[8,9] Exposure to NO_2 concentrations above 200 μg m^{-3} can cause airway inflammation, increase susceptibility to respiratory infections, and exacerbate symptoms in individuals with pre-existing lung or heart conditions.[10]

To mitigate air pollution, London has implemented several emission control policies. In 2008, the Low Emission Zone (LEZ) was introduced to reduce the number of non-compliant vehicles and decrease toxic air pollutants.[11] In 2009, the Congestion Charge Zone (CCZ) was established to limit traffic flow during peak hours.[12] More recently, in 2019, the Ultra Low Emission Zone (ULEZ) was implemented, covering the same area as the CCZ (Figure 1).[13]

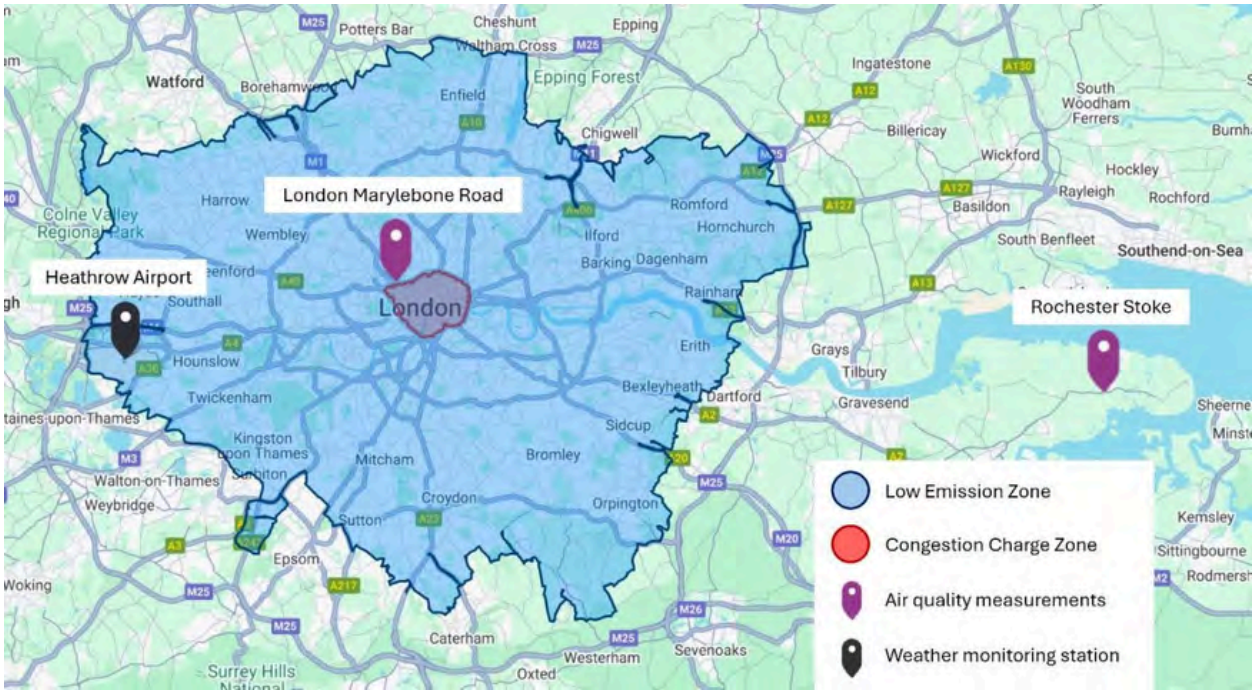


Figure 1: *Map of London showing the areas covered by the Low Emission Zone, the Congestion Charge Zone (coinciding with the Ultra Low Emission Zone introduced in 2019), and the locations of air quality measurement sites at Marylebone Road (urban) and Rochester Stoke (rural). Weather condition data was collected from Heathrow Airport.* [13]

This article[13] hypothesised that pollutant concentrations, particularly PM_{10}, $PM_{2.5}$ and NO_2, would decrease following the introduction of the ULEZ. Therefore, this investigation aims to evaluate post-ULEZ air pollution trends by comparing pollutant concentrations between a central London site and a more rural site, while also assessing the influence of meteorological conditions on pollutant behaviour. Using Python to analyse large air quality and weather datasets, relationships between pollution concentration and temperature, pressure, precipitation, wind speed, and wind direction were examined to better understand how weather conditions influence pollution concentration and transport.

2. Methodology

2.1 Data Sources

Hourly air quality measurements for Marylebone Road and Rochester Stoke were obtained from the Department for Environment, Food and Rural Affairs (Defra) UK Air

Data Archive.[14] Marylebone Road is in an urban traffic region, near heavy traffic, and just outside the ULEZ.[15,16] In contrast, Rochester Stoke is a rural background site situated 20 km from major urban areas.[15,17] Meteorological data, including temperature, pressure, precipitation, wind speed, and wind direction, were collected from Meteostat at the London Heathrow Airport site.[18] The geographical relationship between the three locations is shown in **Figure 1**.

2.2 Python Libraries

Python 3.14.5 libraries including *NumPy*,*Matplotlib* , and*Pandas* were used for data cleaning, temporal alignment, statistical calculations, regression analysis, and graphical visualisation.[19–21]

2.2.1 PM_{10} Long-Term Trends

PM_{10} concentrations at Marylebone Road and Rochester Stoke were compared to investigate long-term pollution trends between 2018 and 2021. Periods of missing data in the original dataset were excluded from analysis. An overview comparison of the full dataset was first conducted, followed by a direct comparison of annual trends for 2018, 2019, and 2020, plotted on the same graph to examine changes in pollution over time. Data from 2021 was excluded due to limited availability.

2.2.2 Daily Variation and Pollution Levels

To examine daily pollution patterns, concentrations measured at 6 AM and 6 PM were compared for PM_{10}, $PM_{2.5}$, and NO_2, allowing direct comparison between morning and evening pollution levels across pollutants. For PM_{10}, a 24-hour daily average profile was also generated by averaging concentrations at each hour of the day to identify periods of elevated pollution.

2.2.3 Meteorological Influence on Pollution

The relationship between PM_{10} concentration and meteorological conditions was investigated using linear correlation and polar graph analysis. Temperature, pressure, and precipitation were analysed using linear regression, with R^2 values and standard deviation (SD) included to assess correlation strength and variability. For precipitation, zero-rainfall measurements were excluded, and an additional regression excluding an outlier (around 20 mm) was performed.

Wind speed and direction were analysed using polar graphs for both PM_{10} and $PM_{2.5}$, where wind direction was represented by angle, wind speed by radial distance, and pollutant concentration by colour intensity. A consistent colour scale (0-25 μg m^{-3}) was applied to allow direct comparison between pollutants.

3. Results and Discussion

3.1 Long-term PM_{10} Trends

Figure2 illustrate the variation in PM_{10} concentrations between urban and rural London from January 2018 to February 2021. The data reveals a general decline in PM_{10} concentration differences, with an average difference of 8.31 μg m^{-3}. The observed trend suggests an overall improvement in air quality over time, aligning with government reports indicating a long-term decrease in PM_{10} levels at urban background sites.[22] Despite some missing data during the winter of 2020, the year displayed consistently lower PM_{10} concentrations, reflecting the positive impact of environmental measures like the LEZ in London, which likely contributed to reduced urban pollution. Notable fluctuations occurred in 2019, where PM_{10} concentration differences between urban and rural areas were more pronounced, likely due to seasonal factors, such as heating activities and varying weather conditions.[23] The introduction of the ULEZ likely influenced the trends observed in early 2019, as traffic may have shifted outside the zone, leading to changes in local pollution levels, particularly along streets like Marylebone Street, located just outside the zone.

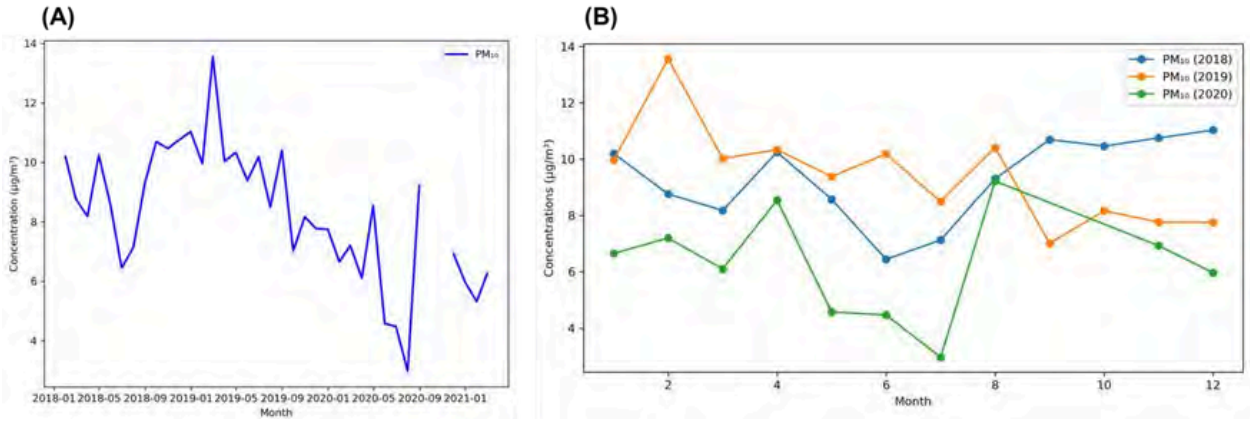


Figure 2: *Comparison of PM_{10} concentration differences between urban and rural London. (A) Long-term trend in PM_{10} concentration difference (μg m^{-3}) between January 2018 and February 2021, showing missing data between September 2020 and January 2021 due to unavailable measurements. (B) Monthly variation in PM_{10} concentration difference (μg m^{-3}) for 2018 (blue), 2019 (orange), and 2020 (green).*

3.2 Daily Variation of Pollution Levels

Figure 3A highlights the diurnal variation in pollutant concentrations, showing that PM_{10}, $PM_{2.5}$, and NO_2 levels are consistently higher at 6 PM compared to 6 AM. For PM_{10}, the average difference is 2.73 µg m-3, with elevated evening concentrations likely linked to increased traffic during peak commuting hours. Similar trends are observed for $PM_{2.5}$ and NO_2, with NO_2 showing the greatest difference (28.57 µg m-3), reflecting its strong association with vehicle emissions. These findings suggest that traffic and urban activity contribute substantially to higher pollution levels later in the day.

This trend is further supported by the hourly PM_{10} profile shown in **Figure 3B**, which reveals a clear daily pattern in pollution concentration. PM_{10} levels peak between 9-10 AM, corresponding to the morning rush hour, and remain relatively elevated during the afternoon and evening period (approximately 5-7 PM), consistent with increased traffic activity. In contrast, the lowest concentrations occur around 4 AM, when traffic volume and human activity are minimal. Together, **Figure 3** demonstrates that pollution levels generally increase during periods of higher urban mobility, with evening concentrations remaining consistently higher than those observed in the morning.

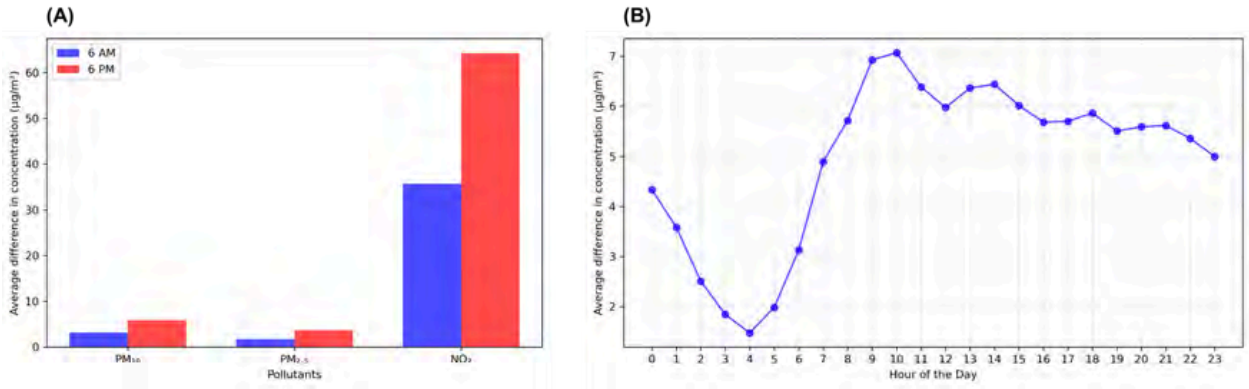


Figure 3: *Daily variation in pollution concentrations. (A) Comparison of average concentrations of PM_{10}, $PM_{2.5}$, and NO at 6 AM (blue) and 6 PM (red). (B) A $^{-3}{}_{2}$veragehourlyvariationin PM_{10}concentration(µgm) across a 24-hour period.*

3.3 Meteorological Influence

Figure 4 (A-C) examines the relationship between PM_{10} concentration and meteorological conditions. Overall, the low R^2 values (0.00-0.04) indicate that pressure, temperature, and precipitation individually show weak correlations with PM_{10} concentration, suggesting that meteorological variables alone cannot fully explain variations in particulate pollution. The relatively large standard deviations (SD = 11.63) also indicate substantial variability in PM_{10} concentration, likely reflecting the influence of additional factors such as traffic density, local emissions, and atmospheric mixing.[24] Among the variables investigated, precipitation showed the clearest negative relationship with PM_{10} concentration (y = -1.17x +16.32), supporting the expected washout effect whereby rainfall removes suspended particulate matter from the atmosphere. However, an extreme precipitation event (~20 mm) strongly influenced the regression trend. After excluding this outlier (gold regression line, y = -1.88x + 16.70), a steeper negative relationship was observed, suggesting that moderate rainfall may have a greater influence on reducing PM_{10} concentrations than implied by the full dataset. Despite this, the low R^2 value indicates that precipitation remains only aweak predictor of PM_{10} concentration when considered independently.

The influence of wind is further illustrated by the polar plots (**Figure 4 D-E**). PM_{10} concentrations generally appear higher than $PM_{2.5}$, as indicated by the greater prevalence of warmer colours in **Figure 4D**, supporting the higher PM_{10} levels observed previously in **Figure 3A**. Pollution was more spatially dispersed towards the south-west, consistent with the dominant south-westerly winds commonly experienced in the UK. [24,25] However, higher pollutant concentrations were more frequently associated with easterly wind directions, suggesting pollutants are transported in from more urbanised regions east of London or possibly from continental Europe. Similar directional patterns have been reported in previous London air quality studies, where easterly airflows were linked to elevated particulate matter concentrations.[26] Lower wind speeds were also associated with higher PM concentrations in literature, suggesting weaker winds limit pollutant dispersion and promote local accumulation.[27]

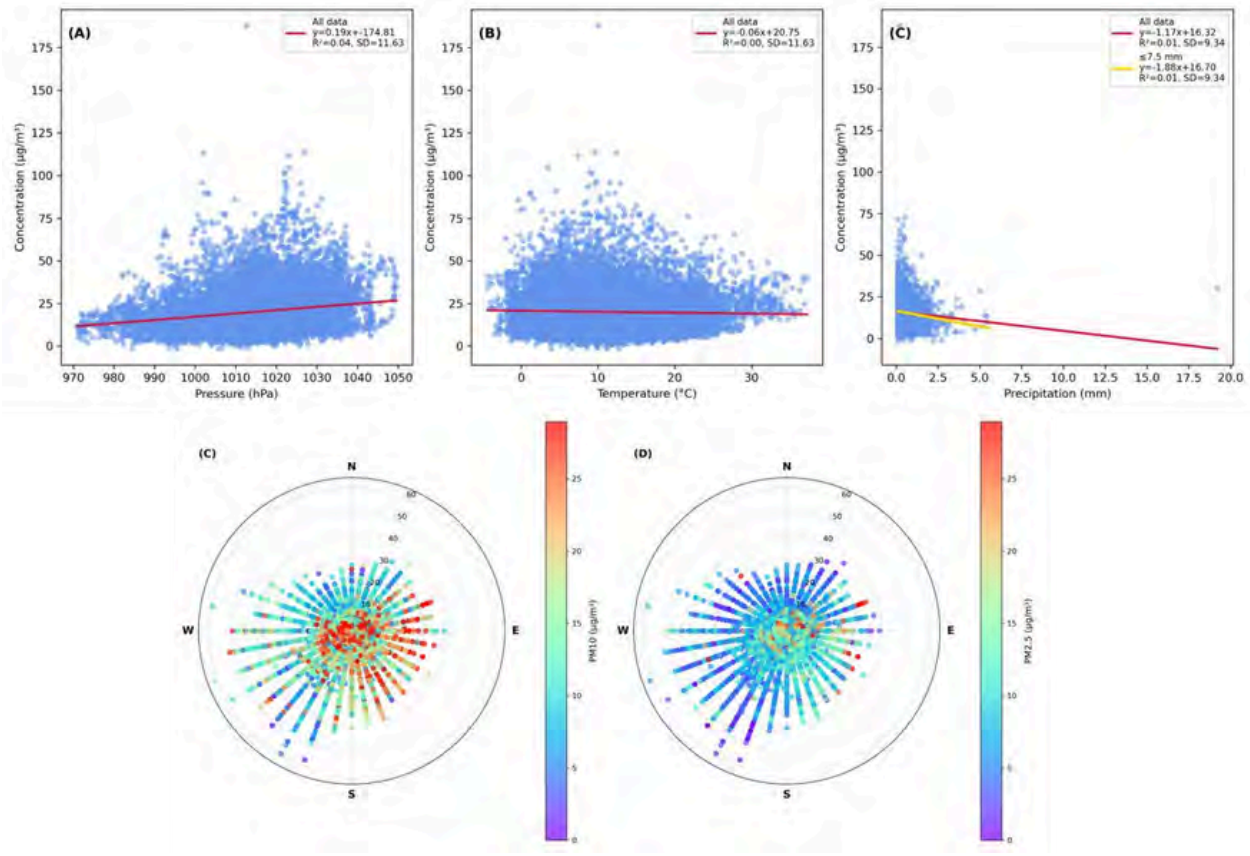


Figure 4: *Relationship between PM_{10} concentration and meteorological conditions. (A-C) Scatter plots showing the relationship between PM_{10} concentration (µgm^{-3}) and pressure (hPa), temperature (°C), and precipitation (mm), respectively. Blue points represent measured data. Red regression lines represent the full dataset: (A) y = 0.19x - 174.81 (R^2 = 0.04, SD = 11.63 µg m^{-3}); (B) y = -0.06x + 20.75 (R^2 = 0.00, SD = 11.63 µg m^{-3}); (C) y = -1.17x + 16.32 (R^2 = 0.01, SD = 9.34 µg m^{-3}). In (C), the gold regression line represents precipitation data after excluding the extreme rainfall outlier (≤ 7.5 mm), described by y = -1.88x + 16.70 (R^2 = 0.01, SD = 9.34 µg m-3). (D-E) Polar plots illustrating the distribution of PM_{10} and $PM_{2.5}$ concentrations (µg m^{-3}) in relation to wind direction and wind speed, using a shared colour scale of 0-25 µg m^{-3}, where warmer colours indicate higher pollutant concentrations.*

4. Conclusion

This study identified an overall decline in PM_{10} concentration differences between urban and rural London between 2018 and 2021, suggesting improving air quality during the post-ULEZ period. A more pronounced reduction was observed during 2020, which may partly reflect reduced traffic and urban activity during COVID-19 lockdowns, consistent with reported decreases in traffic-related pollutants across London during this period.[26] Daily and hourly analyses showed higher pollution levels during peak traffic periods, with elevated concentrations observed during morning (9-10 AM) and evening (5-7 PM) commuting hours, highlighting the contribution of traffic emissions to urban pollution. Meteorological analysis showed generally weak

correlations between PM_{10} concentration and individual weather variables, although precipitation and wind conditions appeared to exert the strongest influence. Rainfall showed a weak negative relationship with PM_{10}, consistent with the washout effect, while higher pollutant concentrations associated with easterly wind directions suggest transport from more urbanised areas or long-range transport from continental Europe.[24,25,27]

This study was limited by missing data between September 2020 and January 2021, which may have underestimated winter pollution peaks and affected long-term trend interpretation. Future work could improve the robustness of the analysis through the inclusion of longer time-series datasets, additional monitoring sites, and broader pollutant coverage to better assess seasonal and spatial variation. More advanced approaches, including multivariable regression, time-series analysis, and machine learning models such as random forest regression, could also be applied to better understand the combined effects of meteorological and traffic-related factors on urban air pollution.[28–30]

References

1 Air pollution accounted for 8.1 million deaths globally in 2021, becoming the second leading risk factor for death, including for children under five years, https://www.unicef.org/press-releases/air-pollution-accounted-81-million-deaths-globally-2021-becoming-second-leading-risk, (accessed 20 October 2024).

2 Rosamund Kissi-Debrah: Woman whose daughter died from air pollution welcomes ULEZ expansion, https://news.sky.com/story/rosamund-kissi-debrah-woman-whose-daughter-died-from-air-pollution-welcomes-ulez-expansion-12950036, (accessed 20 October 2024).

3 Air quality, energy and health, https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants, (accessed 19 October 2024).

4 Air pollution sources, https://www.eea.europa.eu/themes/air/air-pollution-sources-1, (accessed 21 October 2024).

5 Health matters: air pollution, https://www.gov.uk/government/publications/health-matters-air-pollution/health-matters-air-pollution, (accessed 20 October 2024).

6 P. Orellano, J. Reynoso, N. Quaranta, A. Bardach and A. Ciapponi, *Environ. Int.*, 2020, **142**, 105876.

7 Air Pollution and Your Health, https://www.niehs.nih.gov/health/topics/agents/air-pollution, (accessed 21 October 2024).

8 K. Katsouyanni, *Br. Med. Bull.*, 2003, **68**, 143–156.

9 *UK Air Quality Standards Regulations*, 2010.

10 Review of evidence on health aspects of air pollution: REVIHAAP project: technical report, https://www.who.int/europe/publications/i/item/WHO-EURO-2013-4101-43860-61757, (accessed 28 October 2024).

11 London’s poor air quality tackled with launch of Low Emission Zone, https://tfl.gov.uk/info-for/media/press-releases/2008/february/londons-poor-air-quality-tackled-with-launch-of-low-emission-zone, (accessed 31 March 2025).

12 Congestion Charge, https://tfl.gov.uk/modes/driving/congestion-charge, (accessed 31 March 2025).

13 Ultra Low Emission Zone, https://tfl.gov.uk/modes/driving/ultra-low-emission-zone, (accessed 31 March 2025).

14 Data Selector, https://uk-air.defra.gov.uk/data/data_selector_service#mid, (accessed 31 March 2025).

15 Site environment types, https://uk-air.defra.gov.uk/networks/site-types, (accessed 31 March 2025).

16 Site Information for London Marylebone Road (UKA00315), https://uk-air.defra.gov.uk/networks/site-info?uka_id=UKA00315, (accessed 31 March 2025).

17 Site Information for Rochester Stoke (UKA00251), https://uk-air.defra.gov.uk/networks/site-info?uka_id=UKA00251, (accessed 31 March 2025).

18 London Heathrow Airport, https://meteostat.net/en/station/03772?t=2025-03-16/2025-03-23, (accessed 31 March 2025).

19 W. McKinney, 2010, 56–61.

20 J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 90–95.

21 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.

22 Particulate matter (PM10/PM2.5), https://www.gov.uk/government/statistics/air-quality-statistics/concentrations-of-particulate-matter-pm10-and-pm25, (accessed 31 March 2025).

23 R. Cichowicz, G. Wielgosiński and W. Fetter, *Environ. Monit. Assess.*, 2017, **189**, 605.

24 T. M. Habeebullah, S. Munir, A. H. A. A. Awad, E. A. Morsy, A. R. Seroji and A. M. F. Mohammed, *Int. J. Environ. Sci. Dev.*, 2015, **6**, 576–580.

25 P. Sekuła, A. Bokwa, J. Bartyzel, B. Bochenek, Ł. Chmura, M. Gałkowski and M. Zimnoch, *Atmos. Chem. Phys.*, 2021, **21**, 12113–12139.

26 C. Tudor, *Heliyon*, 2022, **8**, e11384.

27 R. Cichowicz, G. Wielgosiński and W. Fetter, *J Atmos Chem*, 2020, **77**, 35–48.

28 S. Subramaniam, N. Raju, A. Ganesan, N. Rajavel, M. Chenniappan, C. Prakash, A. Pramanik, A. K. Basak and S. Dixit, *Sustainability*, 2022, **14**, 9951.

29 R.Liu, L. Pang, Y. Yang, Y. Gao, B. Gao, F. Liu and L. Wang, *Sustainability*, 2023, **15**,4531.

30 Y.Liu, P. Wang, Y. Li, L. Wen and X. Deng, *Sci Rep*, 2022, **12**, 9253.

Authors Bio: Kerry Shen is a second-year Chemistry MSci student at the University of Bristol with interests in analytical chemistry, environmental science, and data-driven approaches to scientific research. Through both academic study and independent projects, she has developed an interest in understanding how chemical and environmental systems interact, particularly in relation to air quality and sustainability. She enjoys applying computational tools such as Python to analyse large scientific datasets and explore real-world environmental challenges. Beyond academics, Kerry is passionate about science communication and youth engagement in STEM and climate-related issues.

Molecular Biology andTranslational Medicine

Meet the Subject Editors

Adrian-Mihai Costea
Molecular Biology and Translational Medicine

Adrian is currently pursuing a PhD at the University of Bristol, where his research focuses on tumour-immune interactions in the context of solid malignancies, interested in how inflammatory and tumour-derived cues rewire immune cell function. Alongside his research, he has a strong passion for science communication and pedagogical research, believing that accessible and rigorous dissemination of research is the motor driving scientific advancement. Working as a Subject Editor has provided invaluable insights into the standards and craft of an impactful manuscript, and this experience has shaped his development as a scientist and contributor to the wider scholarly community.

Shakirah Rashid
Molecular Biology and Translational Medicine

Shakirah Rashid is a final year Biochemistry student, currently completing a year in industry at Pfizer. Through her work supporting early-phase oncology trials and coordinating study communications, she has developed a strong interest in clear and engaging scientific communication. As a Life Sciences Subject Editor for the Student Research Journal, she enjoyed reading a wide range of research and helping authors strengthen their work. Her academic interests focus on molecular biology and gene regulation, and she hopes to continue into postgraduate study and pursue a career in scientific research.

Beyond Transcription: The Mechanistic Functions of RNA in Double-Strand Break Repair

Author: Jack Kalib Callum Magson

ORCID: 0009-0002-2969-2875

Programme: Biochemistry (MSci)

Affiliation: School of Biochemistry and Cellular and Molecular Medicine

Email: np23892@bristol.ac.uk

Abstract

Double-strand breaks represent a severe form of DNA damage, with the potential to threaten genomic integrity and cell survival. Traditional models of double-strand break repair focus on protein-centric mechanisms, however, increasing evidence suggests that various RNA species directly influence the double-strand break repair response through complex scaffolding, localisation, and regulatory modulation of repair factors. This review synthesises the mechanistic roles which RNAs are shown to take in the DNA damage response. Evidence is presented showing long non-coding RNAs acting as molecular scaffolds, facilitating the recruitment and stabilisation of repair complexes. Small non-coding RNAs enhance repair efficiency by enabling sequence-specific localisation and modulation of repair factors, particularly through RNA-guided protein complexes at double-strand break-proximal chromatin. The dual roles of DNA-RNA hybrids are also discussed, with them contributing to both R-loop resolution and end-joining pathway choice, while in some contexts also obstructing repair complex assembly and access to DNA ends. Despite these advances, mechanistic questions remain regarding the precise molecular interactions, 3D-structures, and *in vivo* dynamics of RNA-mediated double-strand break repair complexes. This highlights the need for advanced structural and cellular analyses, with potential implications for understanding genome maintenance and developing novel therapeutics.

Keywords: DNA, RNA, Molecular Genetics, DNA repair, Molecular Biology

Acknowledgements - Many thanks to my supervisor, Professor Nigel Savery, for his guidance with this project, and to my girlfriend Millie, who has supported me relentlessly throughout this term.

Introduction

Double-strand breaks (DSBs) are one of the most severe forms of DNA damage, and can be caused by radiation damage, DNA replication issues and oxidative stress [1,2]. Unrepaired DSBs can lead to chromosomal rearrangement, cancers, and cell death[3]. Therefore, accurate and efficient DSB repair is essential for maintaining genomic integrity.

There are two major, conserved, DSB repair pathways: non-homologous end joining (NHEJ) and homologous recombination (HR)[4]. Both pathways result in DSB repair, however, the mechanisms by which they operate are fundamentally different (Figure 1). NHEJ does not use a DNA template to assist in repair, instead, DNA ends are recognised and aligned by the repair complex with any incompatible overhangs being trimmed into blunt ends, which are ligated[5]. Due to the lack of a template, NHEJ is error-prone and can produce small insertions and deletions, called indels. Alternative end-joining pathways may also be employed, one being microhomology-mediated end joining (MMEJ), which relies on short regions of homology to align DNA ends[6]. Conversely, HR requires a template, usually a sister chromatid formed during the S and G2 stages of the cell cycle[7]. Due to the template, the repaired DNA is an exact copy of the original and is mostly error free.

Historically, RNA was shown to be a passive intermediary, assisting in protein synthesis as a template (mRNA) or as part of the translation machinery (ribosomal RNA and tRNA)[8]. However, recent studies show RNA playing a prominent role in the DNA damage response (DDR). In the context of DSB repair, RNA molecules have been shown to act as scaffolds for repair proteins, form DNA-RNA hybrids at break sites and guide repair protein recruitment[9–11].

These new findings both complement and challenge the classical, protein-centric, models for DSB repair, and have wide-ranging consequences including cancer therapeutics [12]. This review adopts a mechanistic framework to synthesise current evidence for RNA-mediated DSB repair, characterising the multiple ways in which RNA molecules interact with both DNA and protein complexes to influence pathway choice and repair efficiency.

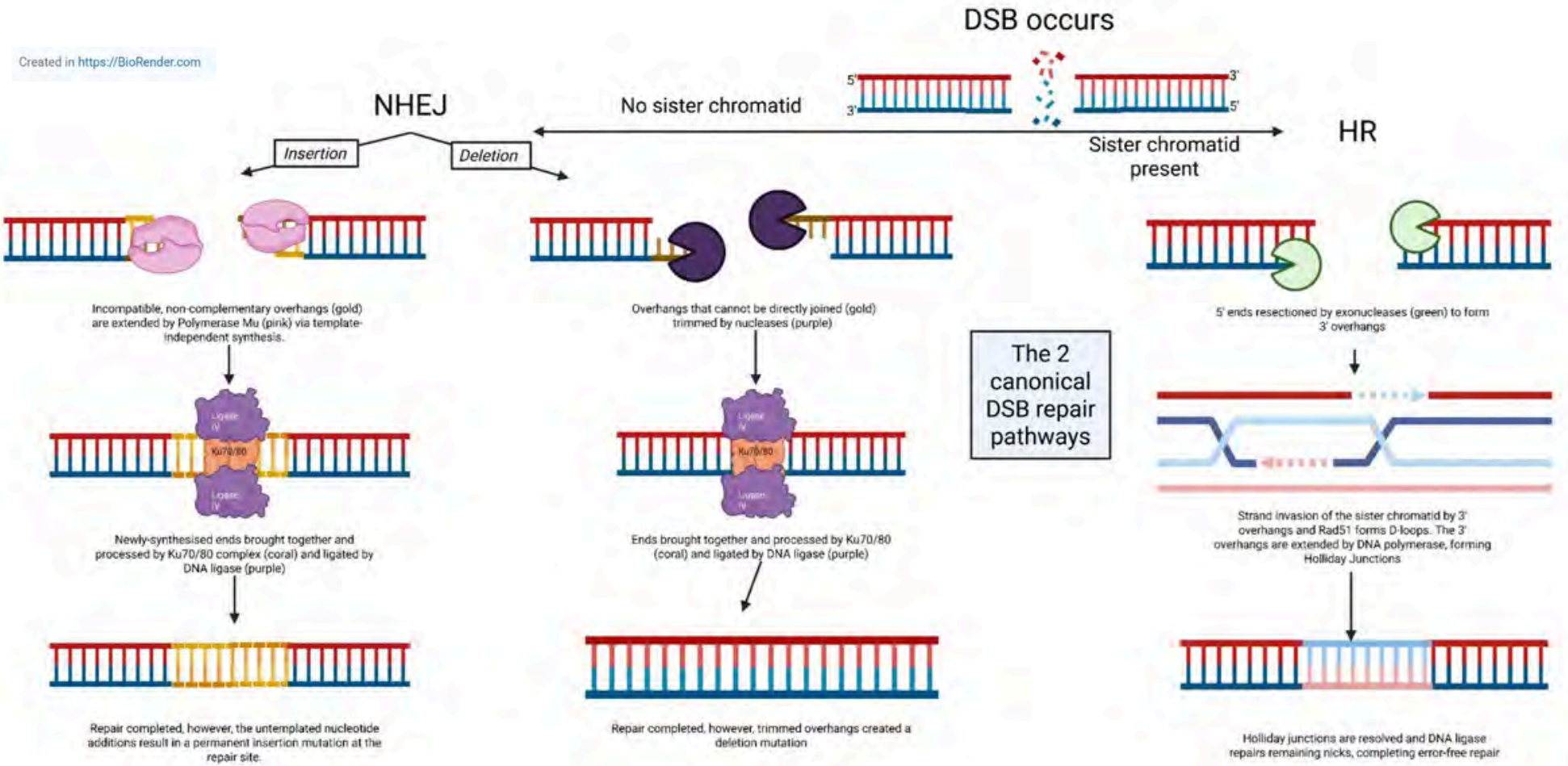


Figure 1. Overview of canonical double-strand break repair pathways (NHEJ and HR), the NHEJ diagram showsboththe formation of insertions and deletion mutations. HR repairs DSBs using a homologous DNA sequence, resulting in high-fidelity repair and an accurate restoration of the original sequence.

Long non-coding RNAs (lncRNAs) in DSB repair

A class of RNA that has received increasing attentioninrecent years are lncRNAs - non-protein coding RNAs over 200 bp in length[13]. They servevarious functions within the cell, from influencing gene expression to epigenetic regulation.Notably, lncRNAs also function as molecular scaffolds for protein complexes, includingDSB repair complexes, aiding in the recruitment and positioning of protein subunits in bothHR and NHEJ scenarios. One example of this is HOTAIRM1[9], which plays prominent rolesin chromatin organisation and cell differentiation[14,15]. HOTAIRM1 also interacts with several RNA-binding proteins, acting as a molecular scaffold to promote repair complex assembly.

Upon induction of DSBs via laser microirradiation, HOTAIRM1 was observed through fluorescence microscopy to accumulate at break sites, colocalising with canonical DNA damage markers, including γH2AX (a phosphorylated histone variant) and p53-binding protein 1 (53BP1)[16,17]. Upon colocalisation, HOTAIRM1 facilitates the recruitment of exon-junction complex (EJC) protein Y14[18], forming an RNA-protein bridge which promotes the continued assembly of repair factors, such as KU70/80, to the break site. This scaffolding activity facilitates effective end alignment and ligation, increasing repair efficiency. Knockdown of HOTAIRM1 by antisense oligonucleotides conferred delayed recruitment of NHEJ repair factors, confirming HOTAIRM1's significant role in repair complex stabilisation[9].

In addition to facilitating repair complex formation, HOTAIRM1 also recruits RNA surveillance factors during NHEJ. Double-strand breaks form damage-induced long non-coding RNAs (dilncRNAs) which aid in early NHEJ, acting as both RNA-protein bridges and DNA-RNA hybrids to help recruit repair factors[19]. While dilncRNAs play a significant role in initiating DSB repair, their accumulation must be tightly regulated as persistent dilncRNAs can interfere with repair. HOTAIRM1 mediates the recruitment of surveillance factors UPF1 and SMG6 to DSB break sites, which target persistent dilncRNAs for degradation and clear the break site to allow for further complex recruitment. siRNA knockdown of UPF1, SMG6 and HOTAIRM1 reduced NHEJ efficiency[9], supporting a second role for HOTAIRM1 at break sites via surveillance factor recruitment and demonstrating its dual role in NHEJ (Figure 2).

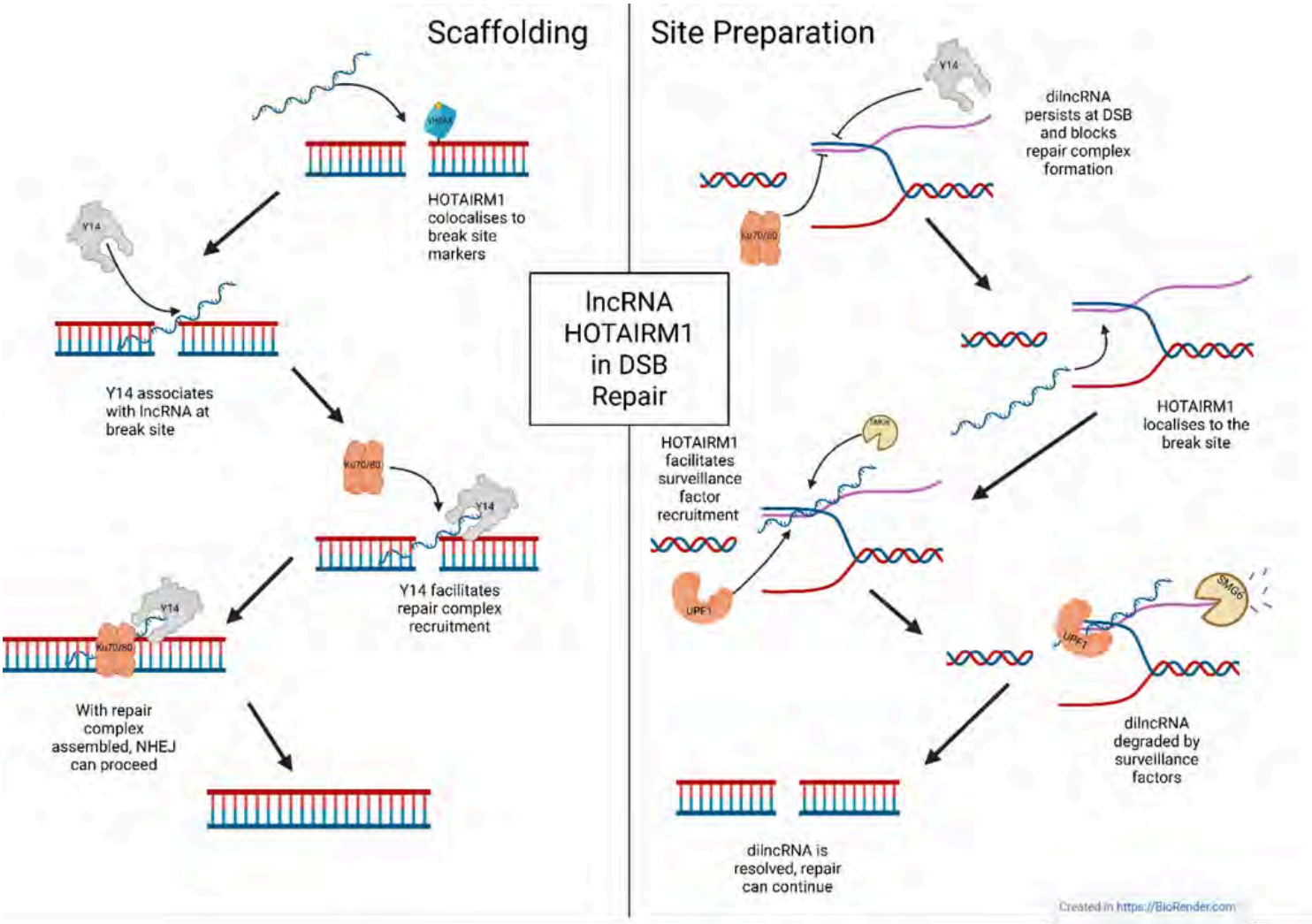


Figure 1 - The mechanistic roles of HOTAIRM1 in scaffolding and repair site preparation, showing specific targeting to DSB sites and active repair complex and dilncRNA resolvase recruitment.

NHEJ is not the only repair pathway influenced by lncRNA expression. Increasing evidence shows that lncRNAs also play a role in homologous recombination, with one being BGL3[20]. Following DSB induction, PARP1 recruits BGL3 through its DNA binding domain, localising BGL3 to the break site. BGL3 then associates with the BRCA1/BARD1 complex, forming an RNA-protein bridge which stabilises the repair factors at the break site. This association promotes the stable recruitment of downstream effectors, such as HP1γ and Rad51, enabling efficient end resection and repair. siRNA knockdown of BGL3 led to an increase in γH2AX retention, indicating impaired repair functionality[20]. Although HOTAIRM1 and BGL3 function in different repair contexts, the conservation of lncRNA scaffolding between NHEJ and HR highlights the broader importance of RNA in DSB repair.

While enhanced repair efficiency is generally beneficial for the maintenance of genomic integrity, hyperactive repair can be problematic. For example, the lncRNA-EME1 acts as a molecular

scaffold, facilitating the recruitment of BRCA1 to the break site and enhancing HR efficiency[21]. Upon lncRNA-EME1 being overexpressed in cervical cancer cells, the increase in repair factor recruitment and HR efficiency led to the cells developing radioresistance against cancer therapies. This highlights the need for tight regulation of DSB-related lncRNA expression in the cell, as excessive repair activity can promote therapeutic resistance rather than directly improving cellular outcomes.

Collectively, these studies demonstrate that lncRNAs scaffold and coordinate both NHEJ and HR repair complexes, highlighting their broad mechanistic diversity[9,20,21]. This highlights a unifying mechanistic principle in which structurally flexible lncRNA molecules serve as dynamic platforms for complex assembly that increase the local concentration and stability of repair factors. However, much of this evidence is derived from laser microirradiation systems, immortalised cell lines, and artificial overexpression models. While localisation and activity are supported in such systems, with necessity often being inferred from knockdowns, the sufficiency of these lncRNAs for initiating repair within physiological contexts *in vitro* remains less clearly established. Whether there is cross talk between these processes, or if they act independently, also remains unclear.

Overall, lncRNAs act primarily as structural scaffolds that organise and stabilise repair protein assemblies, functioning at the recruitment and early coordination stages of both NHEJ and HR to enhance repair complex localisation and retention at DSB sites.

Small non-coding RNAs and RNA Fragments

Alongside lncRNAs, several classes of small non-coding RNAs (sncRNAs) participate in the DSB damage response. Defined as non-protein-coding transcripts under 200 nucleotides, sncRNAs hold both regulatory and structural functions by associating with RNA-binding proteins (RBPs) to regulate gene expression, chromatin dynamics, and DNA repair pathway activity[22]. With regards to the DSB repair response, sncRNAs are shown to function as targeting guides and modulators for larger protein complexes. Among these, fragments derived from Y RNAs have been shown to promote DSB repair[23].

Y RNAs are 80-120nt long sncRNAs originally identified through their association with the Ro60 antigen in autoimmune patients. They adopt a stem-loop secondary structure containing a double-stranded stem, a large internal loop and a single stranded tail, allowing them to specifically bind multiple RBPs[24,25]. Under stressful conditions, Y RNAs are processed by the 3’-

5’ exosome complex and cleaved at the loop region by stress-activated endonucleases such as RNase L[26]. This generates ~30 nucleotides long fragments known as ysRNAs.

DSBs were induced through ionising radiation, which led to an increase in cleavage of the Y3 isoform of Y RNA. This resulted in the accumulation of Y3-ysRNA in both the cytoplasm and the nucleus[23]. These fragments then underwent cytosine methylation by the RNA methyltransferase NSUN2, forming a 5-methylcytosine (m^5C) residue. Following their formation, the ysRNAs were confirmed, through RIP-seq (a technique which immunoprecipitates RNA–protein complexes to identify RNAs bound to a specific protein by sequencing), to associate with the DDR-associated factor YBX1, which was shown by fluorescence microscopy to localise to γH2AX enriched sites. The cytosine modification was reported to be essential for YBX1 binding, with NSUN knockdowns reducing ysRNA-YBX1 binding affinity and diminishing ysRNA localisation to DSBs.

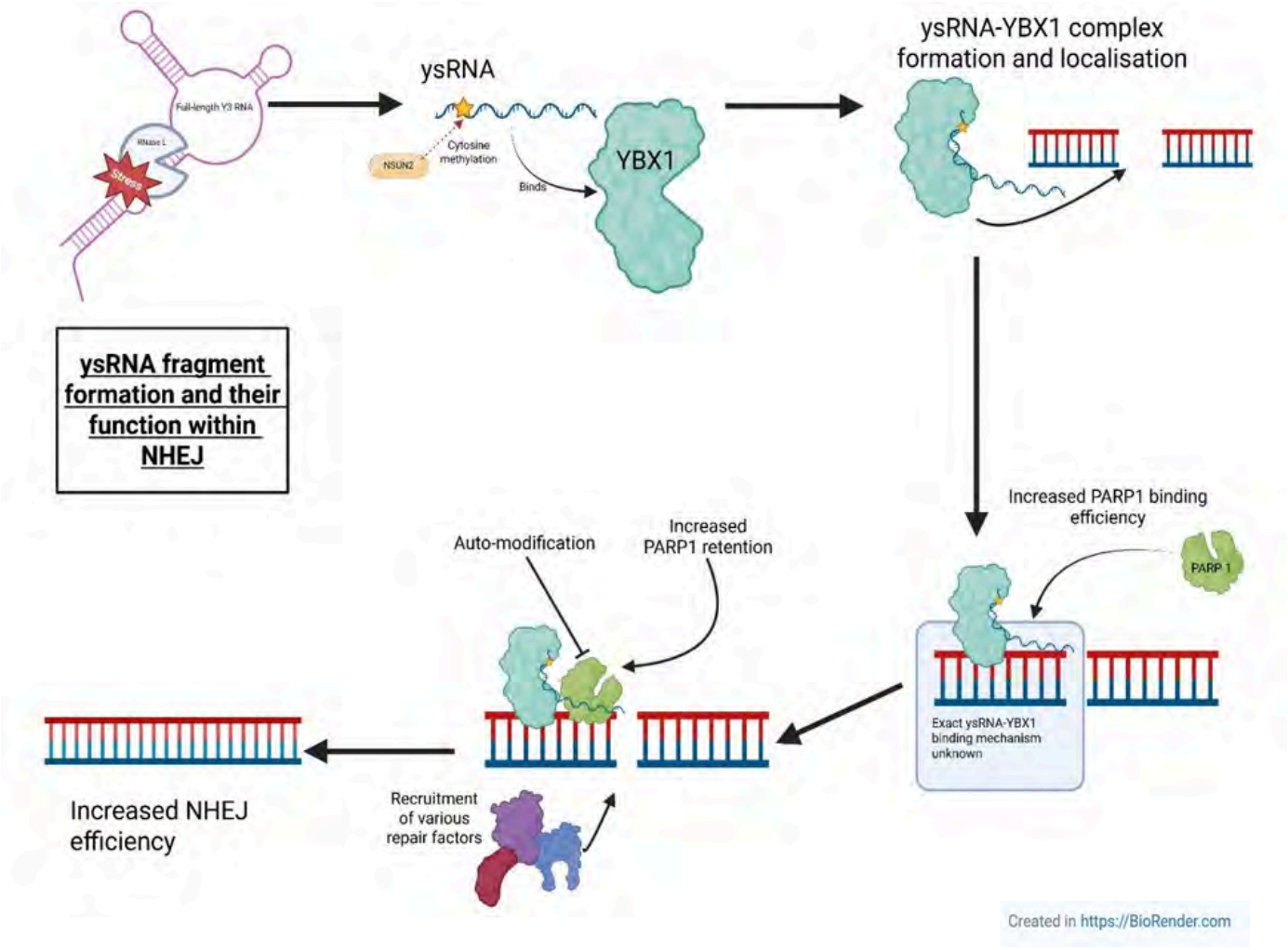


Figure 3 - An overview of the role of Y3 RNA fragments (shown in blue) in NHEJ, higlighting the stress-dependant activation and localisation of the Y RNA fragment to break sites, and the resulting retention of NHEJ factors at break sites.

At the break site, the Y3-ysRNA-YBX1 complex enhanced PARP1 recruitment and activity on chromatin while reducing PARP1 auto-modification. This prolonged PARP1 chromatin association and correlated with enhanced downstream recruitment of repair factors, increasing repair efficiency. DSB repair efficiency was reduced when either Y3 RNA or YBX1 were knocked down. These findings, alongside PARP1's established role in NHEJ[5], highlight the critical role of Y RNA fragments in DSB repair and suggest a mechanistic principle in which Y RNA fragments function through protein activity modulation rather than structural scaffolding, distinguishing them from lncRNA-mediated assembly processes.[25] (Figure 3).

Another group of small RNAs which influence DSB repair are damage induced small RNAs (diRNAs). These are double-stranded small RNA fragments which are processed from dilncRNAs by various RNases (such as Drosha and Dicer) and often terminate at the end of a gene near the DSB site[27]. diRNAs are sequence-specific to the DNA flanking the break site and can therefore act as markers.

In the context of HR, the RNAi factor Argonaute 2 (Ago2) has been shown to localise at DSB sites in a diRNA dependent manner. While this has been directly observed in plant cells[27], the sequence specific recruitment of Ago2 to DSB sites in mammals is also supported by experimental evidence[28].

Ago2 interacts with Rad51 (the recombinase required for strand invasion) via its PAZ and MID RNA-binding domains, forming an Ago2-Rad51 complex. Following the formation of a DSB, the complex binds the diRNAs complementary to damaged DNA, forming a Ago2-Rad51-diRNA complex which assists in localising Rad51 to DSBs [28], although a direct interaction of this complex with the 3' overhangs has yet to be demonstrated. This localisation promotes Rad51 loading and filament nucleation, which is required for strand invasion [7].

Knockdown of Dicer and Drosha reduced diRNA formation, which in turn reduced the association of the Ago2-Rad51 complex to the break site, although complex assembly was shown to be unaffected. The reduced association to the break site prevented Rad51 filament assembly, reducing repair efficiency. Upon reintroduction of diRNAs produced in non-knockout cells, it was recorded that HR activity was recovered[28], showing that diRNAs formed around DSBs are essential for the HR initiation, serving as guides to enable the correct placement of DSB repair factors.

Overall, sncRNAs function as sequence-dependent targeting and regulatory guides that act primarily at the initiation and recruitment stages of homologous recombination and end-joining pathways by directing repair factors such as Ago2 and Rad51 to DSB-proximal chromatin. This is supported by strong localisation evidence in both plant and mammalian models, as well as knockdown studies confirming the necessity of Dicer/Drosha activity. However, much of the mechanistic model is derived from laser-induced damage systems and immortalised cell lines, and while diRNAs appear sufficient to enhance recruitment in rescue experiments, their sufficiency for initiating physiological DSB repair in endogenous chromatin contexts remains less clearly established. In addition, direct biochemical and structural characterisation of these RNA- protein complexes remain limited, leaving uncertainty over the precise binding architecture underlying Ago2 and Rad51-mediated targeting. This contrasts with lncRNAs, which primarily act as structural scaffolds that organise and stabilise repair complex assembly.

R-loops: RNA-DNA Hybrids in DSB Repair

Some dilncRNAs, as well as being processed into diRNAs, can directly associate with complementary DNA flanking the break. This forms structures known as RNA-DNA hybrids[29], including *in cis* R-loops and *in trans* DR-loops. Functionally, these hybrids regulate repair by controlling repair-factor accessibility, stabilising repair intermediates, and influencing pathway choice between HR and end-joining mechanisms. These are, therefore, critical regulators of the DSB response, acting both as aids and obstacles to efficient repair.

It was demonstrated that accumulation of R-loops at DSB sites can be detrimental to repair efficiency, as their persistence after repair initiation impedes the recruitment of HR machinery[30]. The resolution of these R-loops was shown to be mediated by E3 ubiquitin-protein ligase (RNF8). Upon the detection of an R-loop, RNF8 ubiquitylates the exoribonuclease XRN2, which then localises to R-loop prone locations on the genome. XRN2 facilitates R-loop resolution, making way for repair factors to access the break site. In cancer cells expressing an inactive RNF8 mutant, XRN2 recruitment to R-loop sites was impaired, leading to an accumulation of R-loops around break sites, as observed by DRIP-qPCR. This resulted in defective HR and reduced genomic stability, highlighting the potentially detrimental role of RNA-DNA hybrids in DSB repair. Conversely, DNA-RNA hybrid formation can promote DSB repair through DR-Loops, a mechanistically separate class of DNA-RNA hybrids which form *in trans* through hybrid formation between RNA transcripts and homologous DNA sequences on a different molecule from which

the RNA was transcribed (Figure 4a)[31]. When dilncRNAs associated with homologous sequences on the sister chromatid, their proximity and alignment stabilised the ssDNA at the break. This process depended on the protein RAD51AP1, which promoted the invasion of dilncRNA into the sister chromatid[31]. DR-loops acted as scaffolds to promote efficient Rad51

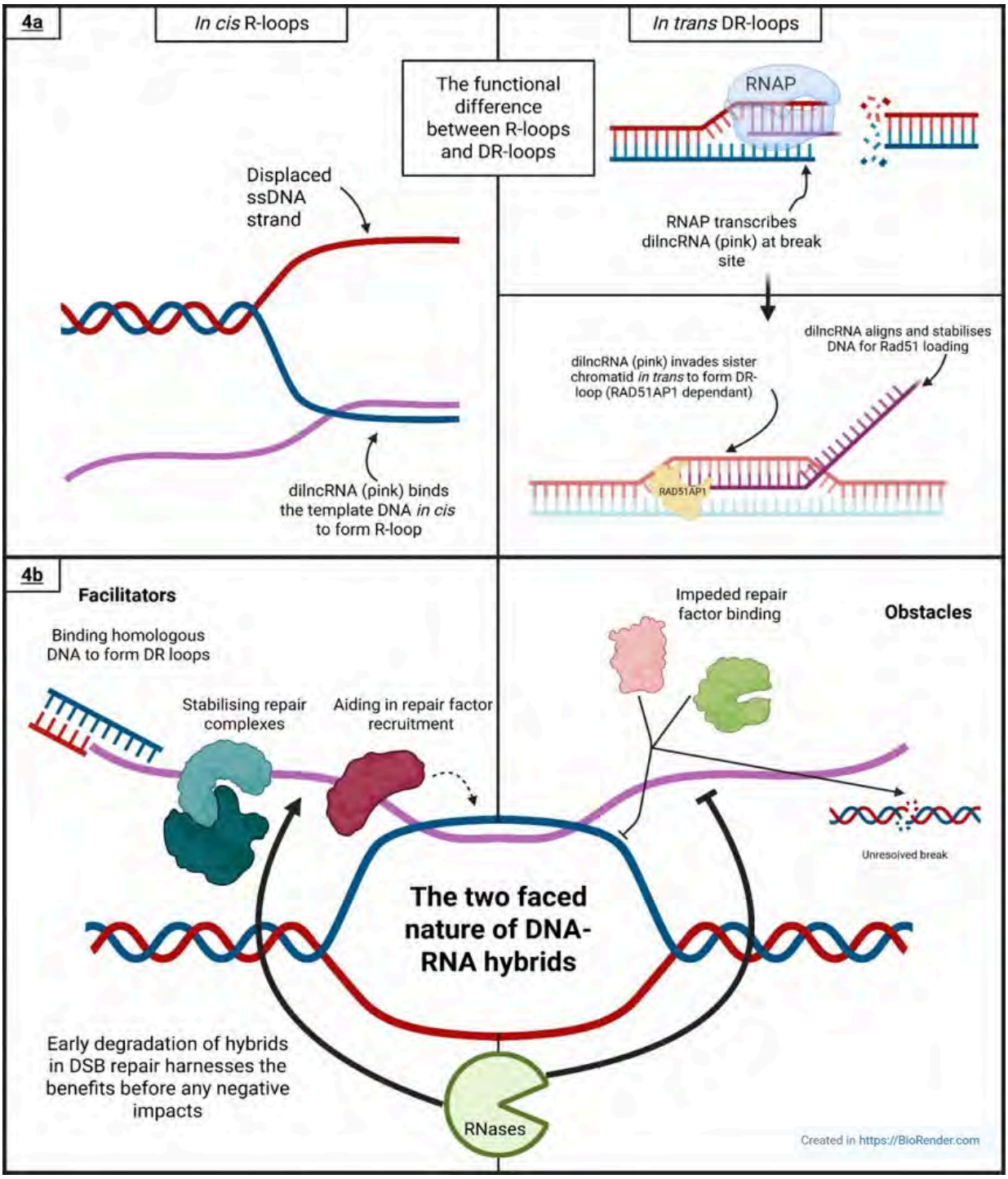


Figure 4 – (a) A diagram showing the differences between R-loops and DR-loops. (b) Various examples demonstrating the two-faced role DNA-RNA hybrids play in DSB repair, with repair facilitators being shown on the left and obstacles on the right.

loading and nucleation, as confirmed by DRIP-seq and RNase-H sensitive assays. Degradation of DR-loops by RNase-H reduced RNA-DNA hybrid levels, suppressing the RNA-stimulated enhancement of HR observed in these experiments. When RNase-H was overexpressed prior to Rad51 recruitment, HR efficiency dropped. These results highlight the context-dependent roles of RNA-DNA hybrids, aptly described as “Janus-faced” by Marnef and Legube[10], as *trans*- acting hybrids such as DR-loops promote repair initiation, while persistent *cis*-acting R-loops impede successful repair (Figure 4b).

There is increasing evidence showing that RNA-DNA hybrids have a distinct role in repair pathway selection[11] (e.g. HR vs NHEJ/MMEJ). diRNAs generated at break sites can anneal with the opposite DNA termini, forming RNA-DNA hybrids that act as a molecular splint to align and stabilise the DNA ends. By stabilising the ends and recruiting end-joining proteins, RNA-DNA hybrids may bias repair towards NHEJ or MMEJ.

It was demonstrated that in yeast and human cells, the processing state of diRNAs influenced hybrid stability and repair pathway choice[11]. Using Cas9-induced DSBs within reporter constructs, spliced and non-spliced RNA transcripts capable of hybridising to sequences flanking the break were compared. Non-spliced transcripts showed a preference for NHEJ, with longer regions of homology providing stronger hybridisation to the DNA ends. In contrast, spliced transcripts, lacking the long homology regions present in the non-spliced variants, preferentially promoted MMEJ. The presence of RNA-DNA hybrids at these sites was confirmed by DRIP- qPCR and their sensitivity to RNase-H digestion, validating that the pathway bias was caused by the hybrids themselves. In both cases, HR was disfavoured, as hybrids likely impede end- resection and block Rad51 recruitment. This suggests that, at breaks where such hybrids form, end-joining mechanisms are favoured under conditions which would usually favour HR. Together, these studies support a multifaceted role for DNA-RNA hybrids in DSB repair as context-dependent physical intermediates that influence repair pathway choice by modulating DNA end accessibility. Their functional outcome is determined by hybrid stability, genomic origin, and processing state. While biochemical and genetic experiments provide robust evidence for their functional involvement in repair outcome determination, much of the causal framework is derived from engineered reporter systems and defined break induction models, leaving their relative contribution and structural organisation at *in vivo* chromatin breaks less clearly resolved. The ability to scaffold directly mirrors that of lncRNAs, however, the two-sided nature of these constructs highlights the necessity of strict regulation to prevent aberrant repair.

Overall, RNA-DNA hybrids act as context-dependent regulators of DSB repair. They function as dynamic intermediates that either facilitate repair factor recruitment and pathway progression or impede repair and destabilise genome integrity depending on their stability, localisation and processing state.

Discussion and Conclusions

The discovery that RNA molecules participate directly in the DSB response has expanded the scope of RNA function beyond just gene expression. Experimental evidence now supports the view that various distinct categories of RNA participate in DSB repair through structural, catalytic, and sequence-guiding roles, expanding upon the diverse role RNA plays in the cell. This review categorises RNA involvement in DSB repair into scaffolding, guiding, and hybrid-mediated regulatory mechanisms.

Mechanistically, the different RNAs take on distinct roles. lncRNAs, such as HOTAIRM1, act as scaffolds to aid in the structural assembly of repair complexes, while also preparing the site for repair by removing blockages[9]. sncRNAs function in a more direct way, whether it be diRNAs guiding repair factors such as Rad51 to the break sites, or Y RNA fragments enhancing NHEJ efficiency by modulating PARP1 activity[23,28]. Conversely, DNA-RNA hybrids formed early in the DDR are shown to facilitate repair factor recruitment and end-resection during HR, while persistent hybrid formation was shown to impede repair and induce genomic instability, demonstrating the two-faced role of DNA-RNA hybrids[29,30]. However, it remains unclear whether there is functional overlap between RNA classes in different repair contexts. Addressing these uncertainties is essential for a complete mechanistic framework of RNA-driven DSB repair and exposes an important conceptual gap in the current understanding.

Despite significant advances, several mechanistic gaps remain unresolved. The interactions that RNAs form are often inferred and not directly visualised, with direct biochemical evidence of binding being lacking. Determining the 3D structure and binding surfaces of these complexes and interactions is essential to understanding the nature of these mechanisms. Such information opens the door to determining the dynamics of these interactions and is necessary to define canonical RNA-mediated DSB repair pathways. Modern technologies, such as AlphaFold and 3D-NMR spectroscopy simplify this process, with AlphaFold3 being able to predict protein- substrate interactions[32], and 3D-NMR allowing RNA-Protein complexes to be modelled in solution[33], enabling conformational dynamics to be measured and quantified.

Alarge proportion of these mechanisms were characterised *in vitro.* While this allows for the characterisation of molecular interactions under controlled conditions, the use of immortalised cell lines (such as HeLa) and cell-free systems may not accurately represent the conditions of native cells [34]. Future studies should aim to visualise these mechanisms *in vivo* in model organisms, such as zebrafish, whose translucent epidermis facilitates fluorescent imaging [35], or Drosophila, due to high homology with the human genome and well characterised DSB repair pathways.

Looking ahead, determining the molecular interactions between RNA and classical DSB repair mechanisms may reveal potential drug targets. These studies showed that aberrant RNA processing, or overexpression of specific RNAs post-break, led to genomic instability, and in some cases radioresistant cancers [20,21]. Identifying the RNAs and repair complexes responsible allows for the development of next-generation targeted therapeutics aimed at selectively modulating the DSB response, whether it be knocking down hyperactive repair or by serving as a preventative measure in cells predisposed to oncogenesis.

Overall, emerging evidence clearly establishes RNA as a central regulator of DSB repair through scaffolding, guiding and modulation. Despite gaps in understanding precise molecular interactions *in vivo*, it is increasingly clear that RNA is not a passive bystander but actively shapes repair pathways and genome stability. Continued investigation of structural biology and *in vivo* characterisation is essential to fully understand the range of influence RNA has on genome integrity, and its potential exploitation in therapeutics and cancer research.

References

1. Jackson SP, Bartek J. The DNA-damage response in human biology and disease. *Nature*. 2009;461(7267):1071-1078. doi:10.1038/nature08467

2. Tang J, Li Z, Wu Q, Irfan M, Li W, Liu X. Role of Paralogue of XRCC4 and XLF in DNA Damage Repair and Cancer Development. *Front Immunol*. 2022;13:852453. doi:10.3389/fimmu.2022.852453

3. Jeggo PA, Löbrich M. DNA double-strand breaks: their cellular and clinical impact? *Oncogene*. 2007;26(56):7717-7719. doi:10.1038/sj.onc.1210868

4. Wyman C, Kanaar R. DNA Double-Strand Break Repair: All's Well that Ends Well. *Annu Rev Genet*. 2006;40(1):363-383. doi:10.1146/annurev.genet.40.110405.090451

5. Davis AJ, Chen DJ. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res*. 2013;2(3):130-143. doi:10.3978/j.issn.2218-676X.2013.04.02

6. Sfeir A, Symington LS. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem Sci*. 2015;40(11):701-714. doi:10.1016/j.tibs.2015.08.006

7. Wright WD, Shah SS, Heyer WD. Homologous recombination and the repair of DNA double-strand breaks. *Journal of Biological Chemistry*. 2018;293(27):10524-10535. doi:10.1074/jbc.TM118.000372

8. Moore PB, Steitz TA. The Roles of RNA in the Synthesis of Protein. *Cold Spring Harbor Perspectives in Biology*. 2011;3(11):a003780-a003780. doi:10.1101/cshperspect.a003780

9. Chuang TW. LncRNA HOTAIRM1 functions in DNA double-strand break repair via its association with DNA repair and mRNA surveillance factors. *Nucleic Acids Research*. 2023;51(7):3166-3184. doi:10.1093/nar/gkad143

10. Marnef A, Legube G. R-loops as Janus-faced modulators of DNA double-strand break repair. *Nature Cell Biology*. 2021;23(4):305-313. doi:10.1038/s41556-021-00663-4

11. Jeon Y, Lu Y, Ferrari MM, et al. RNA-mediated double-strand break repair by end-joining mechanisms. *Nat Commun*. 2024;15(1):7935. doi:10.1038/s41467-024-51457-9

12. Zhang Y. R-loops' m6A modification and its roles in cancers. *Molecular Cancer*. 2024;23(1). doi:10.1186/s12943-024-02148-y

13. Mattick JS, Amaral PP, Carninci P, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol*. 2023;24(6):430-447. doi:10.1038/s41580-022-00566-8

14. Wang XQD, Dostie J. Reciprocal regulation of chromatin state and architecture by HOTAIRM1 contributes to temporal collinear HOXA gene activation. *Nucleic Acids Res*. 2017;45(3):1091-1104. doi:10.1093/nar/gkw966

15. Rea J, Menci V, Tollis P, et al. HOTAIRM1 regulates neuronal differentiation by modulating NEUROGENIN 2 and the downstream neurogenic cascade. *Cell Death Dis*. 2020;11(7):527. doi:10.1038/s41419-020-02738-w

16. Mah LJ, El-Osta A, Karagiannis TC. γH2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia*. 2010;24(4):679-686. doi:10.1038/leu.2010.6

17. Mirman Z, De Lange T. 53BP1: a DSB escort. *Genes Dev*. 2020;34(1-2):7-23. doi:10.1101/gad.333237.119

18. Gehring NH, Lamprinaki S, Hentze MW, Kulozik AE. The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol*. 2009;7(5):e1000120. doi:10.1371/journal.pbio.1000120

19. Böttcher R, Schmidts I, Nitschko V, Duric P, Förstemann K. RNA polymerase II is recruited to DNA double-strand breaks for dilncRNA transcription in *Drosophila*. *RNA Biology*. 2022;19(1):68-77. doi:10.1080/15476286.2021.2014694

20. Hu Z, Mi S, Zhao T, et al. BGL3 lncRNA mediates retention of the BRCA1/BARD1 complex at DNA damage sites. *The EMBO Journal*. 2020;39(12):e104133. doi:10.15252/embj.2019104133

21. Li Q, Zhou Z, Li X, Lan Q. LncRNA-EME1 enhances BRCA1 recruitment and alters repair of DNA damage in cervical cancer radioresistance. *Non-coding RNA Research*. 2026;16:57-69. doi:10.1016/j.ncrna.2025.09.006

22. Xiong Q, Zhang Y. Small RNA modifications: regulatory molecules and potential applications. *J Hematol Oncol*. 2023;16(1):64. doi:10.1186/s13045-023-01466-w

23. Shaw A. Non-coding Y RNA fragments in a complex with YBX1 modulate PARP1 residency at DNA double strand breaks. *Nucleic Acids Research*. Published online 2025. doi:10.1093/nar/gkaf517

24. Wolin SL, Steitz JA. The Ro small cytoplasmic ribonucleoproteins: identification of the antigenic protein and its binding site on the Ro RNAs. *Proc Natl Acad Sci USA*. 1984;81(7):1996-2000. doi:10.1073/pnas.81.7.1996

25. Lima Neto QAD, Junior FFD, Bueno PSA, et al. Structural and functional analysis of four non-coding Y RNAs from Chinese hamster cells: identification, molecular dynamics simulations and DNA replication initiation assays. *BMC Molecular Biol*. 2016;17(1):1. doi:10.1186/s12867-015-0053-5

26. Donovan J, Rath S, Kolet-Mandrikov D, Korennykh A. Rapid RNase L–driven arrest of protein synthesis in the dsRNA response without degradation of translation machinery. *RNA*. 2017;23(11):1660-1671. doi:10.1261/rna.062000.117

27. Kawaguchi K, Satoh S, Obokata J. Transcription of damage-induced RNA in *Arabidopsis* was frequently initiated from DSB loci within the genic regions. *Genes to Cells*. 2024;29(8):681-689. doi:10.1111/gtc.13133

28. Gao M, Wei W, Li MM, et al. Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Res*. 2014;24(5):532-541. doi:10.1038/cr.2014.36

29. Bader AS, Bushell M. DNA:RNA hybrids form at DNA double-strand breaks in transcriptionally active loci. *Cell Death Dis*. 2020;11(4):280. doi:10.1038/s41419-020-2464-6

30. Krishnan R, Lapierre M, Gautreau B, et al. RNF8 ubiquitylation of XRN2 facilitates R-loop resolution and restrains genomic instability in BRCA1 mutant cells. *Nucleic Acids Research*. 2023;51(19):10484-10505. doi:10.1093/nar/gkad733

31. Ouyang J, Yadav T, Zhang JM, et al. RNA transcripts stimulate homologous recombination by forming DR-loops. *Nature*. 2021;594(7862):283-288. doi:10.1038/s41586-021-03538-8

32. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024;630(8016):493-500. doi:10.1038/s41586-024-07487-w

33. Dey A, York D, Smalls-Mantey A, Summers MF. Composition and Sequence-Dependent Binding of RNA to the Nucleocapsid Protein of Moloney Murine Leukemia Virus†. *Biochemistry*. 2005;44(10):3735-3744. doi:10.1021/bi047639q

34. Gillet JP, Varma S, Gottesman MM. The Clinical Relevance of Cancer Cell Lines. *JNCI Journal of the National Cancer Institute*. 2013;105(7):452-458. doi:10.1093/jnci/djt007

35. Blackburn JS, Liu S, Raimondi AR, Ignatius MS, Salthouse CD, Langenau DM. High-throughput imaging of adult fluorescent zebrafish with an LED fluorescence macroscope. *Nat Protoc*. 2011;6(2):229-241. doi:10.1038/nprot.2010.170

Authors Bio:

Hi, I’m Jack Magson. I’m a 21-year-old undergraduate 3rd year biochemist with a passion for understanding the molecular intricacies of life. Beyond the lab, I’m an avid musician and was the proud president of the University of Bristol Big Band Society in '25/'26, where I get to indulge my love for jazz and big band music. When I’m not studying or making music, you’ll find me exploring open source projects and contributing to some of my own.

Do Relaxation and Stress management Interventions have Lasting Benefits for Hypertension: Subgroup Meta-analysis

Author: Laurence Ning

Programme: MB ChB Medicine

Affiliation: Bristol Medical School

Email: wk23815@bristol.ac.uk

Abstract

Objective: Current NICE guidelines identify inadequate evidence for the clinical significance of relaxation therapy effects on adults with hypertension. This review aims to investigate the persistence of effects on blood pressure after practice of relaxation and stress management techniques (RSMTs) have ceased, by performing a subgroup meta-analysis. Method: The primary outcome was the mean difference of change-from-baseline in intervention and comparator systolic blood pressure. The overall and subgroup effects were pooled using random-effects meta-analysis, with the subgroups defined by whether the outcome was measured during or after the intervention. Results: There were 81 included studies from the 179 screened. Overall, RSMTs produced a reduction in systolic blood pressure of 8.20 mmHg (95% CI: 6.20-10.20 mmHg). There was not enough evidence to support a subgroup effect (during: -8.62 mmHg (95% CI: -10.74 to -6.49), after: -7.41 mmHg (95% CI: -11.60 to -3.22); p = 0.60). The certainty of evidence is very low due to concerns with inconsistency, risk of bias and publication bias. Conclusion: We have very low certainty that RSMTs lead to a clinically meaningful reduction in systolic blood pressure, and there is little evidence to support any difference in effects based on whether interventions have ceased or were ongoing.

Keywords hypertension, relaxation therapy, stress management, lifestyle intervention

I am grateful to my supervisors, Monika Halicka and Katie Webster, for their generous support throughout the way.

Introduction

Hypertensionisa chronic condition diagnosed when blood pressure exceeds a clinical threshold[1]. While management often involves taking long-term antihypertensive medications, non-adherence is widespread with a global prevalence of 27-40%[2]. Lifestyle changes are an important part of UK treatment guidelines, recommended as a first-line strategy and complement to medications[3].

There is growing interest in the use of relaxation interventions for the management of hypertension, with the National Institute for Health and Care Excellence (NICE) recommending further research into its effectiveness in their 2019 guideline update [3]. In particular, there is a lack of research on the durability of effects post-intervention. A recently published paper[4] compared types of relaxation and stress management techniques through a systematic review and network meta-analysis. A secondary finding showed that benefits of interventions may diminish over time. Since included studies had a mix of outcomes measured during and after the intervention, this could contribute to the variability of reported efficacy. However, no explicit analysis was performed to strengthen the second finding or whether it could explain heterogeneity in the primary outcome.

This meta-analysis is a secondary analysis of the aforementioned review. It aims to evaluate whether the effects of relaxation and stress management techniques in adults with hypertension are sustained when they are no longer actively practised.

Methodology

Eligibility Criteria

179 studies included in the primary analysis were screened using the following eligibility criteria:

Types of Population

We included randomised controlled trials of adults (≥ 18 years old) with hypertension, defined as ≥140/90 mmHg or ≥135/85 mmHg for office or ambulatory/home measured blood pressure, respectively. We excluded data from pre-hypertensive populations to reduce potential heterogeneity.

Types of Interventions

We included studies evaluating all interventions categories from the primary analysis: autogenic training, biofeedback, breathing control, hypnosis, massage therapy, meditation, meditative movement, mindfulness, multicomponent interventions, music, progressive muscle relaxation and psychotherapy. We excluded studies that only contained data on supplementary or mixed interventions, and excluded data on these interventions from otherwise eligible studies, as these do not directly promote relaxation and stress management.

Types of Comparators

After partial exclusion of ineligible data from studies, we excluded studies that did not contain eligible data on at least one comparator and one intervention. We included studies with passive (no intervention) and non-specific (sham) comparators.

Types of Outcomes

We excluded data on diastolic blood pressure since the primary analysis found similar effect sizes on diastolic and systolic outcomes. We excluded trials that included booster interventions (follow-up sessions after the end of main intervention) as these do not clearly belong in the on or off treatment subgroups, and are hence not well suited to answering the research question. We excluded studies with follow-up durations ≥ 12 months to avoid subgroup over-representation and reduce heterogeneity. Where studies assessed outcome at multiple timepoints, we selected the latest timepoint for each eligible intervention or control arm to avoid double-counting participants.

Approach to Data Processing

All outcome data were prepared for analysis following guidance from the Cochrane Handbook for Systematic Reviews of Interventions [5].

We compared the follow-up duration and outcome measurement timepoints of each datapoint to allocate them to either the on-treatment or off-treatment subgroups. If the outcome measurement timepoint was equal to or less than the follow-up duration (outcome measured during or immediately at the end of intervention), the study was classified as belonging to the on-treatment subgroup. If the outcome measurement

timepoint was greater than the follow-up duration (outcome measured post-intervention), the study was classified as belonging to the off-treatment subgroup. Mean differences between change-from-baseline outcomes of intervention and control arms and their standard errors were used as effect measures for the meta- analysis. These were either directly extracted or calculated (see Appendix A).

Where studies had more than one eligible intervention or control arms, we pooled arm-level data recursively (see Appendix B) before calculating effect measures. This was to avoid the unit of analysis problem whereby participants would be double counted in analysis due to pairwise comparisons involving the same arms.

Where change-from-baseline means or standard deviations were missing, we would first attempt to calculate them from available baseline and follow-up data (see Appendix C). The pre-post correlation, which measures the correlation between baseline and post-intervention outcomes across participants [5], was taken to be 0.5 based on other studies in the main review.

If the standard deviation of baseline and/or follow up means were not available, such that we could not calculate the standard deviation of the change-from-baseline mean via the aforementioned step, we imputed this to be 12.026. We derived this from the average baseline standard deviation of systolic blood pressure from included studies in the primary analysis.

Meta-analysis Approach

The chosen effect measures for the meta-analysis were the mean difference between change-from-baseline means of intervention and control arms, and the standard errors of the mean difference. The data were fitted with a random effects model using the inverse variance method and a Hartung-Knapp (HK) adjustment, allowing for higher likelihood that the true effect is contained within the confidence interval. The restricted maximum likelihood (REML) method was used to estimate 〖tau〗^2.

The overall and subgroup meta-analyses were ran using the “meta” (version 8.2.1) package in R (version 4.5.1). The subgroup analysis compared the effects of relaxation intervention between outcomes measured during or immediately at the end of the intervention and outcomes measured after the intervention.

Assessment of Certainty of Evidence

We assessed certainty of evidence with the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) framework. Initially assumed to be high for a meta-analysis of randomised controlled trials, certainty of evidence of this review can be downgraded incrementally to moderate, low, and very low. The number of downgrades depends on the number of assessment domains where there are concerns, with the domains being risk of bias, inconsistency, indirectness, imprecision and publication bias.

For the evaluation of risk of bias, outcomes of the RoB 2 tool for individual studies were drawn from the primary analysis. Similarly, evaluation of indirectness was drawn from the corresponding domain outcome of the CINeMA (Confidence in Network Meta-Analysis) assessment from the original review for individual studies, based on alignment with a very similar PICO.

For the assessment of imprecision, we chose to use 5 mmHg as the minimally important difference for systolic blood pressure based on a recently published review [6].

Results

Of the 179 studies screened from the primary analysis, 81 met the inclusion criteria (Figure 1). We excluded 22 studies due to only having supplementary or mixed interventions, 13 studies due to being performed on pre-hypertensive populations, 42 studies due to not having at least one comparator and one intervention arm, 19 studies due to using booster interventions, and 2 studies due to long follow-up durations (≥ 12 months).

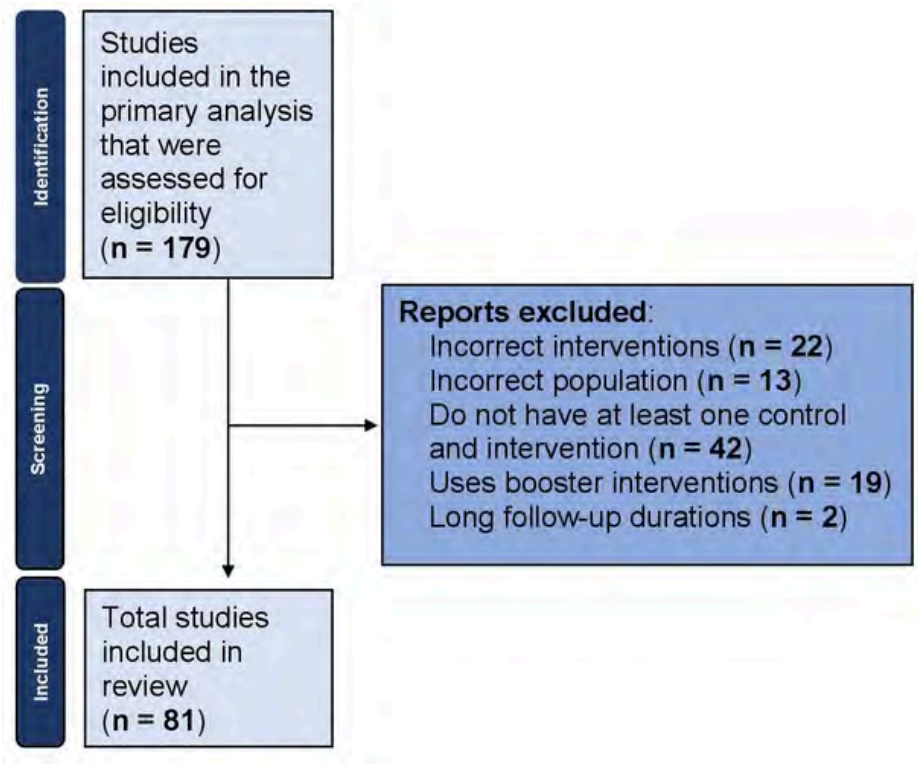


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Flow Diagram, showing the numbers of studies screened, excluded for different reasons and included in the review.

There was a total of 5041 participants across included studies. The mean age was 55.0 years old with an overall standard deviation of 8.87 years. There were 43 studies (53.1%) conducted in high income countries, while 20 studies (24.7%) and 18 studies (22.2%) were conducted in lower-middle and upper-middle income countries, respectively. Short follow-up durations (≤3 months) were used in 49 studies (60.5%) and medium follow-up durations (between 3 to 12 months) were used in 32 studies (39.5%). 73 studies (90.1%) used a passive control, while 49 (60.5%) studies used a non-specific control. The type of relaxation intervention as defined by the primary analysis and the number of included studies that evaluated them are meditative movement [23], biofeedback [15], breathing control [13], multicomponent interventions [8], progressive muscle relaxation [7], psychotherapy [7], mindfulness [6], meditation [3], massage therapy [2] and hypnosis [2].

Overall Meta-analysis

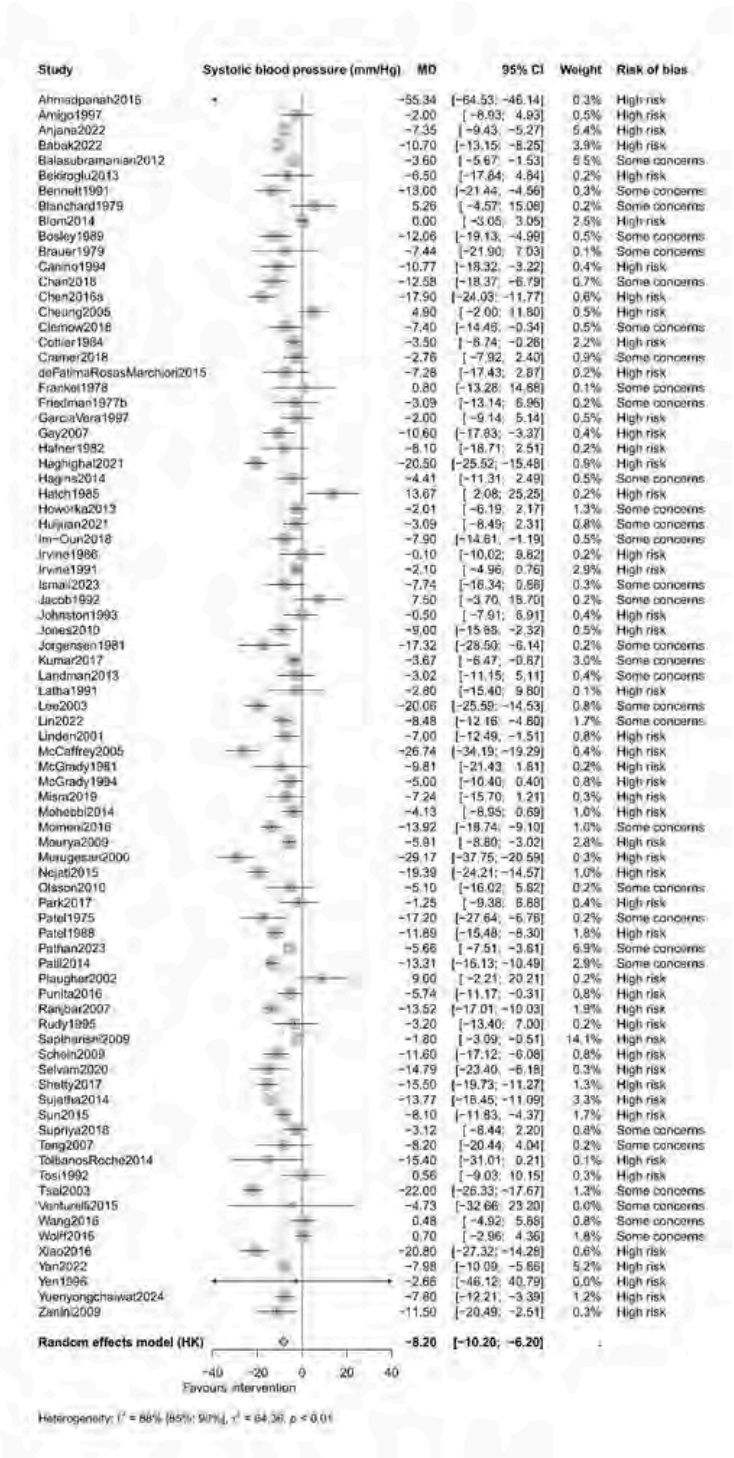


Figure 2. Forest plot of all included studies and the overall meta-analysis result, including heterogeneity statistics (HK: Hartung-Knapp, MD: mean difference, CI: confidence interval).

The overall meta-analysis (Figure 2) demonstrated that relaxation and stress management techniques produced a statistically and clinically significant reduction in

systolic blood pressure compared to controls. The point estimate was -8.1995 mmHg (95% CI: -10.2000 to -6.1991). However, heterogeneity is high, indicated by an I^2 statistic of 87.7% (95% CI: 85.3% to 89.6%).

Subgroup Meta-analysis

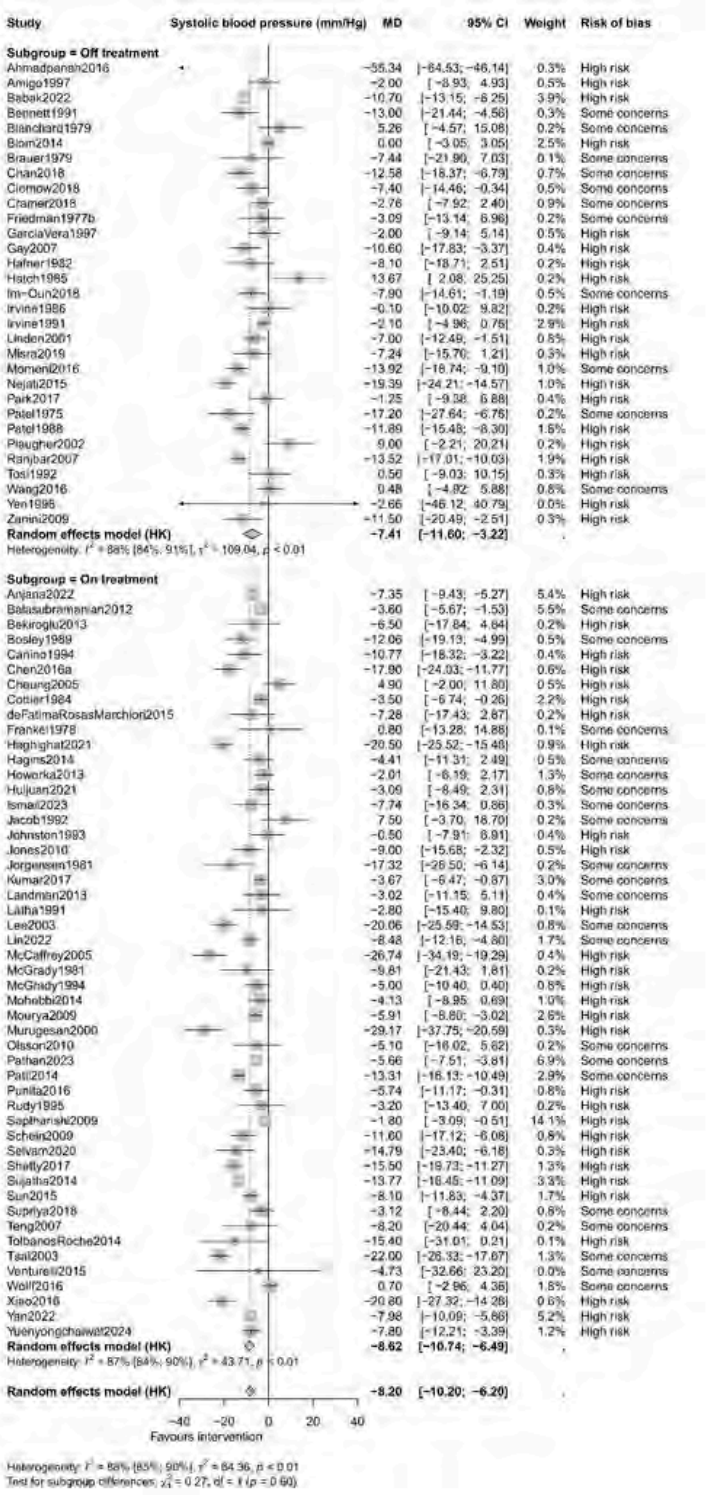


Figure 3. Forest plot of all included studies by subgroup and subgroup meta-analysis results, including heterogeneity and test for subgroup difference statistics (HK: Hartung-Knapp, MD: mean difference, CI: confidence interval).

The subgroup analysis (Figure 3) suggests on/off treatment grouping did not explain any heterogeneity in overall results. Results showed that there was not enough evidence to support that there was a difference between subgroups (Q= 0.27, p-value = 0.6030). I^2 values in the subgroup analysis were large and similar in both subgroups (on treatment: 87.4%, off treatment: 88.3%), indicating that there remained a substantial and comparable degree of heterogeneity in the subgroups.

GRADE Certainty of Evidence Assessment

GRADE domain	Overall Analysis	Subgroup Analysis
Risk of Bias	Downgraded by 1 level - 49 (60.5%) studies identified to have high risk of bias, 32 (39.5%) studies identified to have some concerns	See Overall Analysis.
Inconsistency	Downgraded by 1 level - high I^2 value indicating high percentage of variance due to heterogeneity	See Overall Analysis.
Indirectness	No major concerns.	No major concerns.
Imprecision	No concerns	Though not downgraded, the off treatment mean difference has a confidence interval that crosses the minimally important difference (-7.4143, 95% CI: -11.6042 to -3.2244), indicating reduced certainty of impact on cardiovascular outcomes.
Publication bias	Downgraded by 1 level - only studies in the English language were screened. There was some evidence that many	See Overall Analysis.

	studies were not prospectively registered and that the field is biased towards publishing of positive results.	

Figure 4. GRADE domain-specific judgements for overall and subgroup meta-analysis

The certainty of findings in this study is very low due to concerns for risk of bias, inconsistency and publication bias (Figure 4).

Discussion

Overall, relaxation and stress management techniques appear to yield a clinically meaningful yet highly uncertain reduction in blood pressure among adults with hypertension. The evidence is very uncertain as to whether effects are sustained after the intervention has ceased.

Strengths and Limitations of Evidence Included in Review

A strength of the included studies is their good alignment with the PICO components of the research question, whereby 60 of the 81 included studies had no concerns for indirectness.

However, there are concerns for the risk of bias in included studies. This was assessed for each study using the RoB2 tool, which consists of a series of questions under different domains of bias arising from trial design, conduct and reporting. Out of 81 studies, 49 were identified to have a high risk, while the rest were identified to have some concerns. Many studies had unsatisfactory descriptions of the randomisation and allocation process to intervention and comparator groups, creating a source of selection bias and confounding. Outcome assessors were frequently aware of treatment allocation, jeopardising internal validity. In some studies, there were no formal protocol for measuring blood pressure, which may introduce inconsistencies and errors in measurement and reporting.

Strengths and Limitations of the Review Process

A strength of the review process is the good precision of results. The overall effect confidence interval of -10.20 mmHg to -6.20 mmHg indicates that all plausible values lie above the minimally important difference of 5mmHg of systolic blood pressure reduction. This gives high certainty that the true effect of RSMTs on systolic blood pressure is clinically meaningful. Similarly, the on-treatment subgroup results (95%CI: -10.74 to -6.49) are precise according to the decided threshold. The exception to this certainty is the effect of the off-treatment subgroup (95%CI: -11.60 to -3.22), which cannot exclude the possibility of no clinically meaningful effect [7]. This makes the results of the subgroup analysis less imprecise than the overall, contributing to the uncertainty of whether clinically important blood pressure reductions persist off-treatment.

We are concerned about the risk of bias in the review process. Especially, only studies published in the English language were included as a result of the main review's search strategy, and positive results are more likely to be published in English. Moreover, there is evidence that the field of research is biased towards publishing positive findings. Both of these factors are sources of publication bias that can lead to an effect over-estimation. Many included studies were also found to lack prospective trial registrations which increases the risk of publishing positive data and outcome switching when initial methods fail to produce desired results.

Another limitation is the substantial and unexplained heterogeneity amongst studies. The large I^2 value and small p-value of Cochran's Q test in the overall meta-analysis show that there is high heterogeneity amongst included studies [7]. This review demonstrated that whether participants were still on relaxation therapies likely do not explain heterogeneity. This was also the case for factors investigated in the main review subgroup analyses of the main review, which stratified results by proportion of participants also on drug treatments, country level economic resource, age and severity of hypertension. Future reviews could stratify by type of RSMT as the main review demonstrated difference in effect sizes on blood pressure. However, substantial heterogeneity after subgroup analyses could suggest the influence of complex or unmeasured study-level differences, so future reviews should be cautious when pooling results and strive to avoid over-interpretation.

Implications of Results

As hypertension is a chronic condition with recognised long-term morbidity and mortality, the impact of therapeutic interventions on these risks is therefore of key interest. However, few RCTs include outcomes on cardiovascular events or mortality alongside blood pressure data. Studies with longer term follow-ups are often necessary to obtain this data.

Furthermore, since short-term benefits are likely to not be clinically useful for chronic conditions, the uncertainty of the persistence of blood pressure effects, combined with the lack of longer-term studies, means that the clinical value of RSMTs remains inadequately evaluated. Clinicians should be cautious and consider recommending relaxation interventions as part of broader lifestyle modifications.

Overall, future RCTs should prioritise including longer-term interventions and follow-ups (>1 year) to sufficiently evaluate long-term benefits and collect data on cardiovascular events, co-morbidities and mortality. They should also prioritise having robust methodologies, such as ensuring adequate randomisation and blinding processes, and ensuring prospective trial registration to avoid study-level bias and publication bias.

Agreements and Disagreements with Other Reviews

We only identified one other review [8] with mostly corresponding eligibility criteria and aims. However, likely due to being published when available RCTs in the field were very limited, it lacked the number of suitable studies for a robust systemic review. The author reported that many eligible studies were excluded for important methodological flaws, such as high loss-to-follow up rates and not using concurrent control groups. Lack of studies with larger scale and longer duration also prevented the author from evaluating the long-term endurance of treatment effects. Another review by Lee et al.9 synthesised 12 studies that focused only on mindfulness-based stress reduction techniques. It found a statistically significant (point estimate and confidence interval were both above 0 mmHg), but imprecise, reduction in systolic blood pressure, similar to our main review's findings for mindfulness. A subgroup analysis comparing outcomes measured immediately at the end and after the intervention showed no statistically-supported difference, similar to this review.

Conclusion

Relaxationand stress management techniques may lower systolic blood pressure in the short-term, but long-term durability of effects has very low certainty due to concerns for risk of bias, heterogeneity and publication bias in included RCTs. This contributes to further research recommended by NICE guidelines by providing an update on the current evidence base. It explores the persistence of blood pressure effects on adults with hypertension when few studies with prolonged interventions exist to assess long-term benefits against the chronic disease. The results point to inadequacy of the current evidence in determining the clinical value of RSMTs on hypertensive adults. Future research should prioritise robust trial designs and transparent reporting, performing targeted subgroup analyses to better understand sources of heterogeneity.

References

1. NHS. HighBlood pressure [Internet]. UK: NHS; 2024 [cited 2025 Oct 14] Available from: https://www.nhs.uk/conditions/high-blood-pressure/.

2. Lee EKP, Poon P, Yip BHK, et al. Global burden, regional differences, trends, and health consequences of medication nonadherence for Hypertension during 2010 to 2020: A Meta-Analysis involving 27 million patients. Journal of the american heart association. 2022;11(17). doi:10.1161/jaha.122.026582

3. NICE. Hypertension in adults: diagnosis and management [Internet]. UK: NHS; 2019 [cited 2026 Oct 14]. Available from: https://www.nice.org.uk/guidance/ng136/chapter/recommendations.

4. Webster KE, Halicka M, Bowater RJ, et al. Effectiveness of Stress management and relaxation interventions for Management of Hypertension and prehypertension: Systematic review and network Meta-Analysis. BMJ medicine. 2025;4(1):e001098. doi:10.1136/bmjmed-2024-001098

5. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editor(s). Cochrane Handbook for Systematic Reviews of Interventions [Internet]. Version 6.5. UK: Cochrane; 2024 [cited 2025 Oct 17]. Available from: www.cochrane.org/handbook.

6. Rahimi K, Bidel Z, Nazarzadeh M, Copland E, Canoy D. Blood pressure treatment: how low should you go? – Authors" reply. The Lancet. 2021;398(10312):1684-1685. doi:10.1016/s0140-6736(21)01856-0

7. Neumann I, Schünemann H. The GRADE Book [Internet]. Version 1.0. The GRADE Working Group; 2024 [cited 2026 May 5]. Available from: https://book.gradepro.org

8. Ebrahim S, Smith GD. Lowering blood pressure: a systematic review of sustained effects of non-pharmacological interventions. Journal of Public Health. 1998;20(4):441-448. doi:10.1093/oxfordjournals.pubmed.a024800

9. EKP L, NCY Y, Xu Z, Zhang D, Yu CP, Samuel Y.S. Wong. Effect and Acceptability of Mindfulness-Based Stress Reduction Program on Patients With Elevated Blood Pressure or Hypertension. Hypertension. 2020;76(6):1992-2001. doi:10.1161/hypertensionaha.120.16160

Appendices

Appendix A: Formulas used for calculating mean difference (1) and standard error (2)

$$MD = M_{I,change} - M_{C,change} \quad (1)$$

Where:

MD is mean difference

$M_{I,change}$ is change-from-baseline mean of intervention

$M_{C,change}$ is change-from-baseline mean of comparator

$$SE = \sqrt{\frac{n_I + n_C}{nI \times n_C} \times \frac{(n_I - 1) \times SD_{I,change}^2 \times (n_C - 1) \times SD_{C,change}^2}{n_I + n_C - 2}} \quad (2)$$

Where:

SE is standard error

nI is number of participants in intervention arm

nC is number of participants in comparator arm

$SD_{I,change}$ is change-from-baseline standard deviation in intervention arm

$SD_{C,change}$ is change-from-baseline standard deviation in comparator arm

Appendix B: Formulas used for pooling of number of participants (1), change-from-baseline mean (2) and standard deviation of change-from-baseline mean (3) where there are more than one intervention or control arms in a study. These are applied recursively where there are more than two arms.

$$n_{pooled} = n_0 + n_1 \quad (1)$$

$$M_{pooled} = \frac{n_0 M_0 + n_1 M_1}{n0 + n_1} \quad (2)$$

$$SD_{pooled} = \frac{(n_0 - 1)SD_{0,change}^2 + (n_1 - 1)SD_{1,change}^2 + \frac{n_0 n_1}{n0 + n_1}(M_{0,change}^2 + M_{1,change}^2 - 2M_{0,change}}{n0 + n_1 - 1}$$

Where:

SD^{pooled} is pooled standard deviation

n_0 isnumber of participants in the first arm

n_1 isnumber of participants in second arm

$SD_{0,change}$ is standard deviation of change-from-baseline mean in the first arm

$SD_{1,change}$ is standard deviation of change-from-baseline mean in the second arm

$M_{0,change}$ is change-from-baseline mean of the first arm

$M_{1,change}$ is change-from-baseline mean of the second arm

Appendix C: Formulas used for estimating the change-from-baseline mean (1) and standard deviation of the change from baseline mean (2) where they are not reported.

$$M_{change} = M_{fup} - M_{baseline} \quad (1)$$

Where:

M_{change} is change-from-baseline mean

M^{fup} is follow-up mean

$M_{baseline}$ is baseline mean

$$SD_{change} = SD_{baseline}^{2} + SD_{fup}^{2} - 2 \times Corr \times SD_{baseline} \times SD_{fup} \quad (2)$$

Where:

SD_{change} is standard deviation of change-from-baseline mean

$SD_{baseline}$ is baseline standard deviation

SD_{fup} is follow up standard deviation

$Corr$ is pre-post correlation

An Audit of WHO Surgical Checklist Compliance of University Hospitals Bristol and Weston NHS Foundation Trust with a Focus on the Bristol Royal Hospital for Children

Authors: Uliana Zybina, KiruUthayanan, Ashley Adjeroh

ORCID: 0009-0002-6121-2878

Programme: Medicine MBChB

Affiliation: Bristol Medical School

Email: po24898@bristol.ac.uk

Abstract

Introduction

The World Health Organisation (WHO) Surgical Safety Checklist (SSC) reduces surgical complications and improves communication within surgical teams. Despite strong evidence showing improved patient safety, usage remains inconsistent. Using data from University Hospitals Bristol and Weston NHS Foundation Trust (UHBW), this audit evaluated compliance with the SSC in paediatric surgery and compared this with adult theatres.

Method

Data were collected over 3 consecutive weeks in September 2025. A total of 232 surgical procedures were observed, including 80 paediatric and 152 adult cases. Compliance was assessed using sign-in, time-out and surgical site marking. Staff attentiveness during time-out was recorded, with time-outs not performed before knife-to-skin recorded as non-compliant.

Results

Paediatric surgery showed mean compliance of 90.7% (SD=6.4), highest during time-out (99.7%, SD=2.8) and lowest during sign-in (85.8%, SD=3.5). Adult surgery showed higher mean compliance of 95.5% (SD=10.5), with time-out highest (97.6%, SD=6.9). Inattentiveness occurred more often in paediatric cases (33.8%) than adult cases (13.8%).

We would like to thank anaesthetist Mat Molyneux for his support and guidance throughout this project. This experience provided valuable insight into patient safety and clinical research.

Conclusion

Adult theatres showed higher SSC compliance, although adherence was high in both groups. Targeted education and a tailored paediatric SSC may improve compliance.

Keywords WHO Surgical Safety Checklist, Clinical Audit, Paediatric Surgery, Patient Safety, Quality Improvement

We would like to thank anaesthetist Mat Molyneux for his support and guidance throughout this project. This experience provided valuable insight into patient safety and clinical research.

Introduction

While surgeryisfundamentally vital in-patient care, it poses significant risks [1]. Major complications,such as wrong-site surgery and post-operative infections, arise in 3 to 17% of surgicalprocedures, with at least half of them being preventable [2]. As a result, safety checks are essential in maintaining patient safety and high standards of care [3].

Implemented in 2007 by the World Health Organisation (WHO), the *Surgical Safety Checklist* (SSC) has proven to decrease the number of surgical complications and improve communication within surgical teams [4]. Usage of the SSC remains low despite substantial empirical data showing increases in patient safety and outcomes of surgery [3].

Using data gathered from the University Hospitals Bristol and Weston NHS Foundation Trust (UHBW), with an emphasis on the Bristol Royal Hospital for Children (BRHFC), this audit aims to evaluate how well paediatric surgical teams use the SSC and contrast this with adult theatres. The sign-in, time-out and the presence of the surgical mark were weighed to calculate mean compliance, with additional attention paid to the attentiveness of staff members during time-out.

Methodology

The data collected for this audit was gathered in the BRHFC over 3 consecutive weeks from mid-September 2025. The paediatric wards assessed in the BRHFC were Coastguard and Horizon. The adult data was collected during the first 3 weeks of September 2025 at St Michael's Hospital, Bristol Eye Hospital, and Bristol Royal Infirmary (BRI), with an extra day spent at Weston General Hospital. Because of the small and disproportionate size of the data pool, wards were not compared and were combined in the results to avoid sampling bias.

Compliance during sign-in and time-out was evaluated using a quantitative paper checklist that included when the time-out took place, the surgical team's attentiveness, and verification of the correct surgical site marking (Appendix 2).

The first assessed part of the SSC was the sign-in, completed before the patient could be anaesthetised. It required the compulsory presence of specific staff members, as well as completion of safety questions such as confirming the patient's identity, consent, and checking for allergies (Appendix 1).

As children are shown to experience higher pre-operative stress, the surgeon did not attend the sign-in process in paediatric cases in order to lessen patients 'anxiety [5]. This data was omitted to avoid skewed results. Additionally, the estimated blood loss (EBL) was not mentioned at this stage, as it was deemed negligible. Therefore, to provide a fair comparison to the adult data, EBL was omitted from the data analysis in paediatrics.

A second safety check was performed prior to the incision. The time-out reaffirmed the patient's identity, procedure, site and allergies. The time-out could have been done pre-scrubbing, pre-draping, post-draping, pre-knife-to-skin (KTS) and post- KTS, which was recorded in the audit. As it should be done post-draping and pre- KTS, time-outs done at other points were recorded as non-compliant [6]. The audit checklist used assessed whether the time-out was fully carried out, with any omitted questions recorded (Appendix 2). The inattentiveness of the staff members was measured by 3 categories: mild (e.g., checking equipment), moderate (e.g., being on the computer), and severe (e.g., not present in the room).

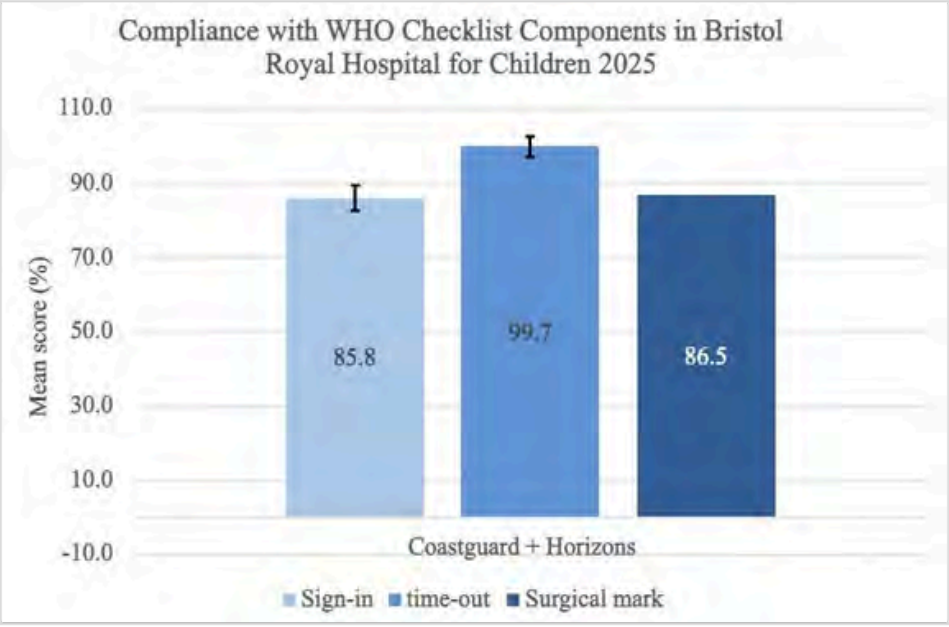
Finally, a mark of the site of surgery should be done in indelible pen prior to the patient going into the anaesthetics room, with the patient being involved and awake

[7]. To avoid a skew, any mark data for surgeries not requiring it (e.g., bilateral procedures) was excluded.

A total of 232 procedures were analysed (80 paediatric, 152 adult). Mean compliance rates and standard deviations were calculated. Group comparisons were performed using two-sample t-tests with significance set at $p<0.05$. Error bars represent standard deviations, and corresponding t-values and p-values are reported for all comparisons.

Results

(A)



(B)

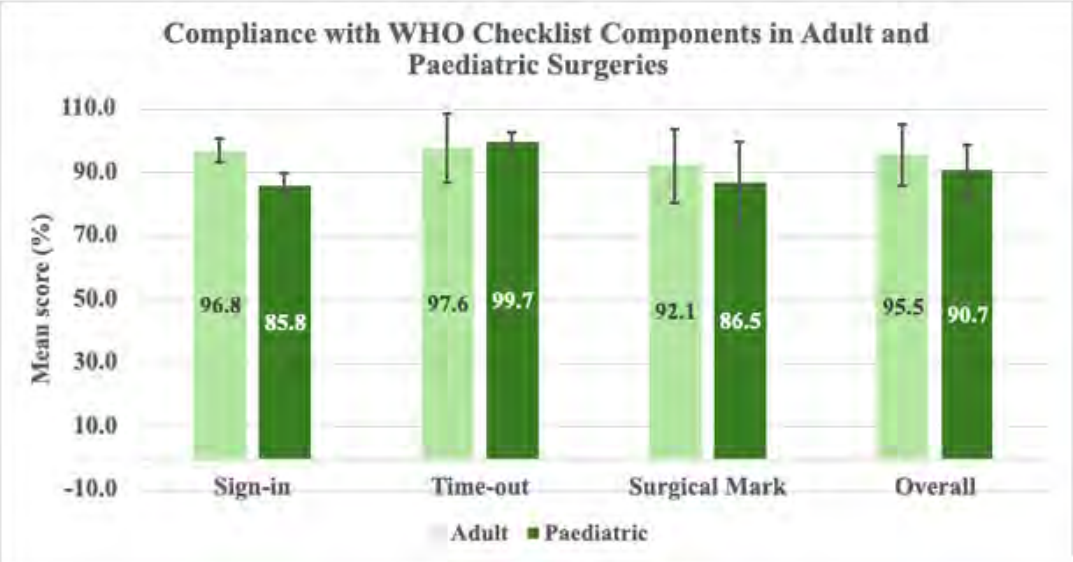


Figure 1. Compliance with WHO Surgical Safety Checklist components in paediatric and adult surgery (2025). (A) - Mean compliance scores for WHO checklist components in paediatric surgery. (B) - Comparison of mean compliance scores between paediatric and adult surgery.

Paediatric surgery scored a mean compliance score of 90.7% (SD=6.4). The highest compliance score of 99.7% (SD=2.8) was seen during time-out, while the lowest was shown by the sign-in at 85.8% (SD=3.5) (Figure 1A).

For the adult data, the overall mean compliance was 95.5% (SD=10.5). The best result was also scored by the time-out at 97.6% (SD=6.9), followed by the sign-in achieving 96.8% (SD=7.7) (Figure 1B).

	Adult (N=152)	Paediatric (N=80)
Surgeries where staff were inattentive	21 (13.8%)	27 (33.8%)
Number of staff inattentive (%)		
1	80.9	51.9
2	9.1	40.7
≥3	0	7.4
Degree of inattentiveness (%)		
Mild	28.6	81.5
Moderate	42.8	14.8
Severe	28.6	3.7

Table 1. Comparing the degree of inattentiveness during adult and paediatric surgery

Although only 13.8% of inattentiveness was recorded in adult surgeries, while paediatrics had 33.8%, adult inattentiveness was more severe. On average, paediatrics showed 3.5 times more inattentive staff per case than adults (0.15 and 0.53, respectively). While the majority of paediatric cases had mild distraction (81.5%), adults had more moderate (42.8%) and severe (28.6%) inattention.

Discussion

The audit found high compliance rates for adults (95.5%) and moderately high results for paediatrics (90.7%), both showing higher compliance rates than seen in similar audits (Figure 1B) [8]. A two-sample t-test showed that this difference (4.8%) is statistically significant ($p<0.05$) suggesting potential age group variability, but the magnitude is small. Both groups demonstrated strong adherence and a 4.8% gap within such a high-compliance range is unlikely to translate into a meaningful difference in surgical outcomes. [9].

Component-wise, paediatrics showed 2.1% lower compliance during sign-in, with high statistical significance ($p<0.001$)(Figure 1B). This could be explained by the children and families 'preoperative anxiety levels being significantly higher than adults '[5,10]. During the audit, we found that managing the anxiety of the family and child, as well as the child's resistance to anaesthetic induction, frequently distracted staff members during sign-in. Such interruptions correlate with oversight of checklist components and decreased compliance rates [11].

Apart from time-out, where paediatrics outperformed by 2.1%, adult procedures scored higher on all components (Figure 1B). This difference may be due to the checklist being originally developed for adult procedures. Familiarity with the standard checklist has been shown to increase compliance, whereas paediatric checklists require tailored surgical and anaesthetic approaches, leading to questions being omitted during sign-in [8].

Furthermore, audits take place more frequently in adult theatres, encouraging accountability, increased training, and improvement of checklist use over time [12,13]. Because staff in adult theatres are more aware of the audit's requirements, they are more vulnerable to the Hawthorne effect, which occurs when people alter their behaviour because they are aware that they are being evaluated [14].

The score for the surgical mark was higher for adult procedures than paediatric by 5.6% (Figure 1B). However, it is important to note that the SD was also the highest for that component (18.1 and 13.0, respectively). This indicates a high amount of variability and thereby lower statistical significance for that component, making it hard to say whether adult theatres would still yield higher results if the audit was repeated with a larger data set. We found that all of the mistakes that decreased compliance for this component in paediatrics were made by covering the mark during drapage. This difference could be explained by children's smaller body surface area, which makes it harder to drape the patient without covering the indelible pen, decreasing mean surgical mark compliance [15]. As correct site marking protocol is vital to prevent errors such as wrong site surgery, increased compliance is needed regarding this component in order to improve care standards [16].

Finally, paediatrics had 20% more surgeries where staff were inattentive (Table 1). We observed during the audit that while the adult data had more complex surgeries, the paediatric data had more common, shorter procedures like tonsillectomies and endoscopies. As the operating risk is lower for minor surgery, staff may perceive the time-out as less crucial, leading to omitted questions [2]. The fast procedure turnover rate may also cause staff to become task-orientated, decreasing their engagement with the SSC [11].

Limitations

Despite showing promising compliance rates, this audit had limitations. Out of 232 surgeries, only 80 were paediatric. The unequal sample size with almost a 2:1 ratio of adult to paediatric data introduced potential under-representation bias into the results. Furthermore, comparison between wards or specialities was not possible due to the limited dataset. Future studies should collect more well-balanced data from each ward and speciality, aiming for a larger sample size.

Moreover, all staff members were aware of the fact that the WHO audit was taking place, potentially paying more attention to the sign-in and time-out, making the data susceptible to the Hawthorne effect. This source of bias would impact the study's validity [14]. To counteract this, the audit should be done internally, by individual members of the team, without awareness from the rest of the staff members.

Although the level of inattentiveness during time-out was categorised into common behaviours (e.g., talking to staff or not being in the room), actions that did not fit into this category were up to the auditor to document as mild, moderate or severe. This introduced observer bias into the data. To avoid inter-rater variability in future studies, a wider variety of behaviours should be standardised and categorised into each subgroup prior to audit.

The study's next steps would be to address staff opinions and ideas regarding each component of the WHO checklist in order to determine the underlying causes of low compliance. This could be accomplished by using surveys to gather qualitative data, targeting staff members commonly present in theatre. Finally, future studies should

employ a random number generator to select a random operating room for daily audits, thereby enhancing objectivity and validity and minimising selection bias [17].

Conclusion

The purpose of this audit was to analyse and compare the compliance of UHBW's paediatric and adult surgical theatres with the WHO SSC. Overall, both groups demonstrated strong adherence (>90%), showing that the SSC is well-implemented throughout the UHBW.

Adult data achieved a 4.8% higher compliance rate. The biggest difference, 11%, was seen during sign-in, possibly caused by paediatric staff being distracted while managing children's pre-operative tension. Helpful implementations would include educating staff on child anxiety management techniques and creating a tailored paediatric SSC. By adding questions specific to paediatric care, staff members may perceive the checklist as more relevant, thereby improving overall adherence. Future studies should aim for a larger sample size of equal distribution to allow for comparison between different wards and specialities within adult and paediatric surgery.

References

1. Gasoma EBY, Marouf MA. Enhancing Surgical Outcomes: Evaluating the Impact of Implementing the World Health Organization Surgical Safety Checklist-A Prospective Cross-Sectional Study. *Surg J (N Y)*. Oct 2024;10(4):e53-e59. doi:10.1055/s-0044-1800917

2. Haynes AB, Weiser TG, Berry WR, et al. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine*. 2009-01-29;360(5)doi:10.1056/NEJMsa0810119

3. Badasa K, Abera M, Abebe M, et al. Improving utilisation of the WHO surgical safety checklist at Wollega University Referral Hospital: a multidimensional quality improvement project. *BMJ Open Qual*. May 7 2025;14(2)doi:10.1136/bmjoq-2023-002599

4. Ambulkar R, Ranganathan P, Salunke K, Savarkar S. The World Health Organization Surgical Safety Checklist: An audit of quality of implementation at a tertiary care high volume cancer institution. *J Anaesthesiol Clin Pharmacol*. Jul-Sep 2018;34(3):392-398. doi:10.4103/joacp.JOACP_328_17

5. Getahun AB, Endalew NS, Mersha AT, Admass BA. Magnitude and Factors Associated with Preoperative Anxiety Among Pediatric Patients: Cross-Sectional Study. *Pediatric Health, Medicine and Therapeutics*. 2020 Dec 16;11doi:10.2147/PHMT.S288077

6. WHO. *WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives*. 2009. *WHO Guidelines Approved by the Guidelines Review Committee*.

7. Quality PSa. In: Hughes RG, ed. *PatientSafetyandQuality:An Evidence-Based Handbook for Nurses*. 2008. *Advancesin PatientSafety* .

8. Habtie TE, Feleke SF, Terefe AB, et al. Beyond compliance: examining the completeness and determinants of WHO surgical safety checklist - a systematic review and meta-analysis. *BMC Health Services Research 2025 25:1*. 2025-04-04;25(1)doi:10.1186/s12913-025-12569-0

9. Finch S, Cumming G. Putting research in context: understanding confidence intervals from one or more studies. *J Pediatr Psychol*.Oct2009;34(9):903-16. doi:10.1093/jpepsy/jsn118

10. Liu W, Xu R, Jia Je, et al. Research Progress on Risk Factors of Preoperative Anxiety in Children: A Scoping Review. *International Journal of Environmental Research and Public Health 2022, Vol 19, Page 9828*. 2022-08-09;19(16)doi:10.3390/ijerph19169828

11. Freundlich RE, Bulka CM, Wanderer JP, Rothman BS, Sandberg WS, Ehrenfeld JM. Prospective Investigation of the Operating Room Time-Out Process. *Anesth Analg*. Mar 2020;130(3):725-729. doi:10.1213/ANE.0000000000004126

12. WHO. WHO's patient-safety checklist for surgery. *Lancet*. Jul 5 2008;372(9632):1. doi:10.1016/S0140-6736(08)60964-2

13. Mahajan RP. The WHO surgical checklist. *Best Pract Res Clin Anaesthesiol*. Jun 2011;25(2):161-8. doi:10.1016/j.bpa.2011.02.002

14. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol*. Mar 2014;67(3):267-77. doi:10.1016/j.jclinepi.2013.08.015

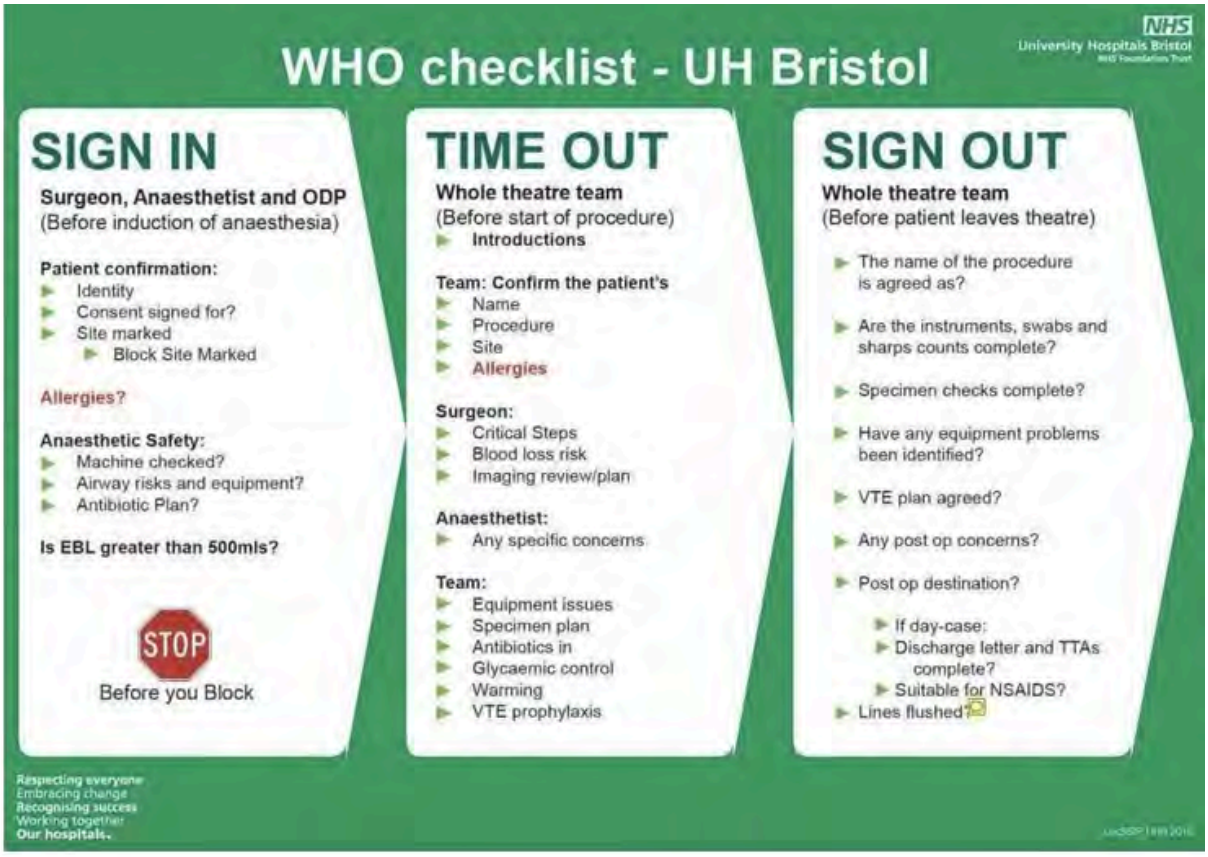
15. Mears SC, Dinah AF, Knight TA, Frassica FJ, Belkoff SM. Visibility of surgical site marking after preoperative skin preparation. *Eplasty*. Jul 16 2008;8:e35.

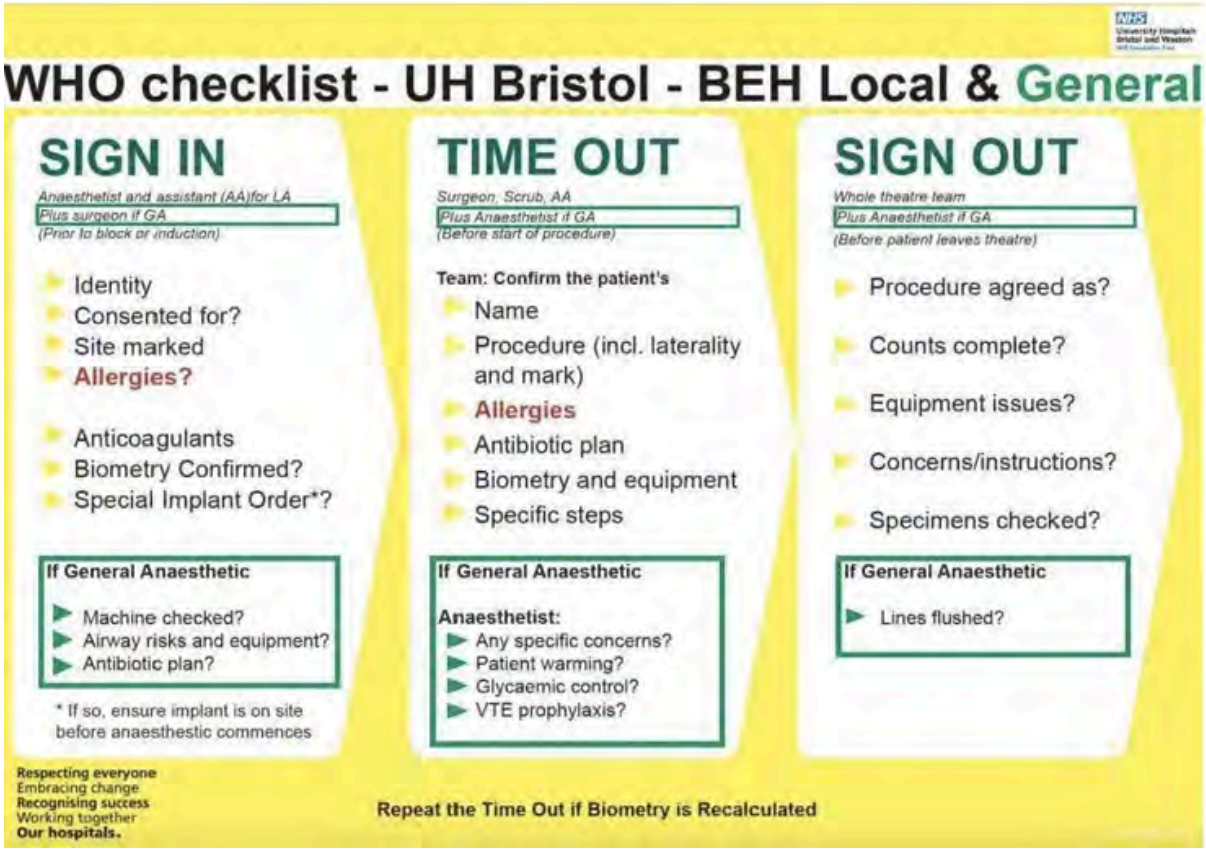
16. M P, A M, A G. The WHO safer surgery checklist time out procedure revisited: Strategies to optimise compliance and safety - PubMed. *International journal of surgery (London, England)*. 2019 Sep;69doi:10.1016/j.ijsu.2019.07.006

17. Gitelis ME, Kaczynski A, Shear T, et al. Increasing compliance with the World Health Organization Surgical Safety Checklist-A regional health system's experience. *Am J Surg*. Jul 2017;214(1):7-13. doi:10.1016/j.amjsurg.2016.07.024

Appendix

Appendix 1: The WHO SSCs used at the UHBW.





Appendix 2: The paper checklist used to assess WHO SSC compliance rates.

WHO Checklist and Marking Audit 2025

Sign in:	Y	N
Surgeon present		
Anaesthetist present		
ODP present		
Patient identity confirmed		
Procedure confirmed		
Consent checked		
Allergies checked		
Anaesthetics safety questions asked		
Estimated blood loss gauged		

Date:.....
.....
Theatre Suite and number.....
Surgeon (grade):.....
Anaesthetist (grade):.....
Specialty:.....
.....
T number:
.....

Time out:

Was it carried out?			
Was it carried out fully? (If 'no', record omitted questions below)			
Surgeon from sign in present at Time Out?			
Was the senior surgeon present at Time out?			
Who led the checklist?	Surgeon	ODP	an

The Mark:
Prior to AR: Y N
Indelible pen: Y N
Visible prior drape: Y N
Visible after drape: Y N
If No…remarked?: Y N
Comment:

If so, when:

Pre-scrubbing	
Pre-draping	
Post-draping, pre-KTS	
Post-KTS	

Quality of time-out:

All members of team paused and attentive	
Number of people inattentive	
Role/ name of person	Level of inattention (please circle as appropriate)
1.	Mild/ Moderate/ Severe
2.	Mild/ Moderate/ Severe
3.	Mild/ Moderate/ Severe
4.	Mild/ Moderate/ Severe

Questions omitted in time out:
-
-

Authors Bio:

My name is Uliana Zybina and I am currently a second-year medical student at the University of Bristol. My strong interest in improving patient care has led me to undertake a surgical checklist audit in order to contribute to enhancing health and safety in the operating theatre. While in theatre, I enjoyed seeing how important effective team communication and engaging with patients in meaningful interaction are in delivering high-quality care.

Towards Reliable Synthetic Circuit Design: Challenges, Modelling Approaches, and Future Directions

Authors: George Gould

ORCID: 0009-0003-0775-6935

Programme: Biology BSc

Affiliation: School of Biological Sciences

Email: jt23155@bristol.ac.uk

Abstract

Reliable synthetic circuit design remains limited by unpredictable host-circuit interactions, resource competition, and weakly emergent behaviours. Modelling tools assess circuit viability and behaviour, but they currently either lack the resolution and scale to be accurate and capture global effects or are too computationally intensive for routine use. This review evaluates the potential of selectively detailed whole-cell models to balance biological fidelity with practicality by integrating molecular simulations with deterministic and machine learning predictions and coarse-grained representations. Such models provide detection of harmful emergent behaviours and offer a unified platform for testing diverse circuit types. Expanding orthogonal toolkits is a parallel strategy for circuit reliability, reducing crosstalk and improving modularity. Despite this, incomplete biological knowledge, unmodelled pathways and imperfect insulation still act as barriers to reliable design, keeping empirical circuit validation essential. Overall, hybrid modelling and insulated design accelerate the build-test cycle, enhancing circuit reliability before implementation into living cells.

Keywords Synthetic Biology, Cell Modelling, Synthetic Circuits

Introduction

The concept of gene networks as analogous to circuits emerged in the 1960s when gene control was explored in bacterial regulatory systems[1]. Analysing the Lac operon first characterised how bacteria regulate the synthesis of specific proteins in response to environmental changes[2]. Research following this key paper documented how these systems operate through interactions that regulatory proteins such as repressors, activators and transcription factors have with genetic elements such as promoters and operators[3,4].

Based on the understanding of the roles of genetic regulatory elements, the fundamental ambition of synthetic biology has been to impose control theoretic principles- frameworks for designing predictable systems through feedback and regulation- onto biological systems[5]. Producing artificial gene networks that behave predictably holds substantial potential, with applications in biosensing, therapeutics, and biofuel and pharmaceutical production industries[6]. Increasing reliable design is crucial to unlocking the full potential of synthetic biology; however, this ambition has consistently been challenged by the inherent stochasticity and emergent complexity of biological systems[7]. This review synthesises early circuit design and its shortcomings, the emergence of automation and modelling, and the current landscape for biological circuit design. It will address challenges the field has faced, to what extent they have been overcome, and suggestions for future directions.

Early Circuit Design

The first synthetic biological circuits emerged at the start of the 21st century and confirmed the functionality of synthetic circuits within living organisms. The Repressilator was synthesised from three synthetic repressor systems to generate oscillations that periodically repress and induce protein expression[8]. Another study from the same year designed a toggle switch from two repressible promoters that mutually inhibit one another, resulting in a system that can be switched between two states[9]. Although these studies were revolutionary, marking the start of the field of circuit design, they relied on manual design and trial-and-error methods, limiting their scalability to just simple modules with a small number of regulatory building blocks[10]. Because of this, they serve mostly as a proof of concept rather than viable methods for synthetic design and construction.

Furthermore, early circuits exhibited unexpected noise- variations from expected outputs- highlighting immediate challenges regarding biological unpredictability and demonstrating that simple, logic-based design principles alone cannot account for the stochasticity of biological systems[11]. Stochastic effects arise from crosstalk between components, causal interactions, nonlinear component relationships and resource competition within the cell[12,13,14]. These qualities contribute to weak emergence, which describes behaviours that emerge from a whole system that are hard to predict when looking at its individual parts[15]. Bedau[15] emphasised how weakly emergent qualities make creating circuits that produce expected behaviours difficult when they are placed in the context of a whole cell, and how this unpredictability hindered early design efforts. Whilst this paper highlights an early roadblock, modern methods have demonstrated that uncertainty in circuit design from emergent behaviours can be reduced to an extent, utilising a combination of modelling, automated design tools and standardised part libraries[16,17].

Emergence of Automated Design Tools

Automated tools using Computer-Aided Design (CAD) for circuit generation became a focus of synthetic biology research in the 2010s, representing a shift from manual, unreliable circuit design to a computationally driven and predictive process. AutoBioCAD is an early tool which pioneered automated circuit assembly using evolutionary algorithms and stochastic models[18]. Despite AutoBioCAD's innovation as a concept, it faced limitations via small genetic part libraries and lacked predictive accuracy when circuits were designed in different contexts- factors that ultimately restricted its mainstream use.

One of the first widely recognised and successful software tools was Cello, which generates circuits based on the user's specific logical function request, using genetic components from a library of standardised parts[19]. Cello then simulates circuit performance *in silico* using deterministic models, outputting the most promising circuits. Whilst initial results of Cello seemed successful with 92% of outputs behaving as predicted, an important caveat is that simulations are carried out on a local and isolated scale, omitting interactions with the global cellular environment. Cello does not consider resource competition, coupling of components and potential host interactions at a global scale, which may give rise to emergent behaviours[14,13].

This means that despite impressive trial results, predictions may fail when circuits outside the intended design space are generated, causing circuit failure. As a result, Cello is locally constrained by the types of circuits it can reliably design.

The Infobiotics Workbench (IBW) advances on Cello's shortcomings by modelling a larger genetic network and integrating Gillespie's stochastic simulation algorithms[20]. The improved modelling allows IBW to handle biological stochasticity and gene expression dynamics more accurately than Cello's deterministic models. Combined with broader network simulations, IBW's design space is less confined, reducing the tool's specificity and bringing generalised circuit design a step closer. Despite being more practical, IBW's generation is still limited to small, well-defined circuits, as it similarly cannot capture whole-cell processes.

This evolution shows that whilst integrating broader, stochastic modelling into design software has improved their reliability and scope, the same overarching challenges persist. Firstly, these modelling frameworks assume a degree of statistical predictability- a condition that biological systems rarely meet, and a simplification that sacrifices accuracy[5]. Secondly, integrated CAD predictive testing operates on too small a scale to predict broad emergent effects[21]. Whilst these models may suffice for current circuit design, their inaccuracies will become a future limitation as circuits increase in size and complexity. Considering this, testing models should increase their predictive scope from local networks to global cellular environments, increasing the range and size of circuits they can test and improving robustness against emergent behaviours. These limitations motivate the need for more complex and encompassing biological simulations such as whole-cell models (WCMs).

Whole-Cell Models

Whole-cell modelling was first proposed at the turn of the century, with the first complete and reliable models functional by the early 2010s[22,23]. These models integrate a cell's entire genome, proteome, transcriptome and metabolome, constructing a model of the whole cell, its chemical reactions and the physical processes that influence their rates[21]. They therefore offer a solution for shortcomings of local modelling methods by reliably testing a circuit's functionality within a whole cell, and supporting circuit scaling, allowing increasingly larger circuits to be evaluated within a global context.

Despite their promising applications to synthetic biology, WCMs introduce several challenges. When modelling an entire cell, comprehensive experimental data is required for quantifiable measurements of every protein, metabolite and reaction. Currently, datasets are fragmented and often exhibit inconsistencies such as variable units and experimental conditions[24]. As a result, only select organisms have the comprehensive libraries required to create a complete WCM, and unknown genes and interactions remain a major source of unpredictability[25]. This highlights that modelling does not currently replace empirical testing, but serves to accelerate the design process reliably, generating circuits that must then be verified *in vivo*.

Additionally, simulating an entire cell remains an expensive task, and the time, resource and computational requirements have kept true WCMs bespoke, and not accessible or reproducible for routine circuit design applications[26]. A potential solution is using 'effective models', which aim to simplify WCMs through similar deterministic frameworks as the aforementioned CAD tools but implemented on a larger biological scope to capture more global effects[27]. These utilise fixed equations to represent changes in molecule concentrations- for example, using a single deterministic equation for rate of protein synthesis per ribosome, rather than modelling the binding of every ribosome and mRNA as a detailed WCM would. Another approach is coarse-graining, which simplifies biological data by averaging over molecular detail, capturing just the main interactions[28]. These techniques could be used in combination, with general biological noise coarse-grained and important molecular processes abstracted using deterministic models. Whilst these techniques unload a significant amount of computational burden, the simplifications and abstractions of biological data may limit their predictive reliability[29].

The goal, therefore, is a maximally efficient model that balances biological fidelity with computational cost. This requires determining which details are essential to model fully to capture emergent behaviours, and which can be abstracted to alleviate computational burden. Whilst computational bottlenecks can be alleviated in this way, incompleteness of molecular datasets for high-resolution simulation remains the most fundamental challenge, as it introduces gaps that no algorithm can overcome, keeping empirical testing necessary.

Machine Learning to Supplement Models

Current studies are additionally using machine learning (ML) to minimise the run-times and computational limitations of WCMs. These models reliably predict metabolic fluxes and genotype-phenotype relationships in a fraction of the time it would take a WCM to simulate them in molecular detail[30]. A study testing the viability of reduced genome *E. coli* cells used ML surrogates to predict cell division, resulting in a 95% reduction in model run time compared to using a WCM alone[31]. Whilst the paper focuses on genome reduction viability, the core advancements occur in the computational and methodological processes, which can be applied to circuit testing models. Additionally, ML can derive interaction potentials across varying levels of resolution, allowing for the seamless combination of atomic and coarse-grained models[32]. These advantages make ML a powerful tool for future hybrid modelling, enabling faster simulation and bridging different scales of modelling resolution.

Despite this, ML surrogates require large amounts of good quality training data to provide accurate predictions[33]. Consequently, they fail when model parameters fall outside their training data due to inaccurate extrapolation[34]. These limitations highlight the need for collaborative efforts to generate training datasets if ML methods are to be used for circuit design applications. Overall, given datasets are comprehensive, ML is not a replacement but a powerful complement to deterministic models: ML is most effective when employed to accelerate specific, well-understood modules, whilst deterministic models offer crucial interpretability through mechanistic equations. Both should be deployed to alleviate computational strain, reserving detailed molecular models for critical processes.

Generative AI in Circuit Design

Generative artificial intelligence (AI) is emerging as a further development in circuit design. Whilst conventional ML surrogates generally test pre-designed circuits, AI may be utilised to create novel circuits, representing a fundamental leap in design automation. Large language models (LLMs) and deep generative models are pre-trained on extensive collections of biological sequence data. This allows models to process and reproduce the underlying regulatory and biophysical 'grammar' of biological systems, enabling the generation of novel circuits with predicted functionality[35].

Recent advancements have been demonstrated by CELLM, which integrates LLMs with Cello's software, allowing circuits to be designed and optimised using natural language inputs[36]. The ability to turn text descriptions into functional designs significantly lowers the technical barrier for researchers, reducing the need for specialised bioengineering expertise and highlighting the future potential of AI for accessible and routine circuit design. Despite this, AI models share the same limitations of ML, requiring comprehensive, high-quality training data to avoid inaccurate extrapolations beyond their training distributions and produce reliable outputs. Until sufficient datasets are curated, their integration with mechanistic models will remain essential to ensure not only computationally plausible but biologically reliable circuits are generated. Overall, CELLM is a promising start to the integration of LLMs within design suites, underpinning a fundamental shift from the design-build-test-learn cycle to an approach where models can propose circuit architectures guided by learned biological principles, rather than extensive manual research.

Modelling Approach	Key Examples	Strengths	Limitations
Deterministic CAD Tools	AutoBioCAD,[18] Cello[19]	Automated design, high initial accuracy (92%); fast *in silico* simulation	Local scale, omits global cellular context, context-dependent failures
Stochastic CAD Tools	IBW[20]	Handles biological noise, broader scope, reduced design-space constraints	Still limited to small, well-defined circuits, cannot capture global processes
Whole-Cell Models (WCMs)	Karr *et al.*[23]	Full cellular context, detects broad emergent behaviours, supports circuit scaling	Computationally expensive, requires complete molecular datasets, not yet routine
Coarse-Grained / Effective Models	Sechkar *et al.*,[28] Stumpf[27]	Reduced computational cost, captures main global interactions	Sacrifices resolution, abstractions may limit predictive reliability
Machine Learning Surrogates	Gherman *et al.*[31,32]	Fast prediction, can bridge modelling scales, 95% reduction in run-time demonstrated	Requires large training datasets, poor extrapolation outside training data can cause inaccuracies
Generative AI	CELLM[36]	Natural language-driven design, learns biological grammar for *de novo* part generation	Data-dependent, limited extrapolation, requires integration with mechanistic models

Table 1: Summary of modelling approaches for synthetic circuit design, key examples, their strengths and main limitations.

Engineering Reliable Circuits: Orthogonal Parts

Even with predictive tools becoming increasingly accurate, synthetic circuits exhibit unpredictable context effects- when new modules are placed within a cellular environment, it can change their behaviours and generate unexpected noise through crosstalk with host components[10,12]. Orthogonal parts are engineered to interact with only specific partners, creating insulated communication channels that function independently of endogenous pathways[13]. Current research focuses on developing orthogonal toolkits to minimise context effects of synthetic modules, vital for making circuits behave more predictably in whole-cell contexts and restoring a degree of modularity to increase circuit scalability.

A 2019 study created twelve orthogonal transcription factors for 'marionette' *E. coli* cells, each specific to its own promoter and inducer, so expression of these genes could be controlled independently[37]. Building on this, insulated parts have been combined to engineer tuneable logic gates[38]. That study demonstrates how NOT gates were used to turn off gene expression in the presence of certain chemical cues, and tested various gate combinations for orthogonality, ensuring no cross-interference occurred. Whilst these studies show promising results for orthogonal part reliability, they can still exhibit crosstalk to a certain extent due to natural weak binding affinities that components may have with non-target parts[39]. As these studies test within controlled conditions, they may not predict real-world conditions, as weak, non-cognate binding is unavoidable even when parts are insulated.

Furthermore, studies have exhibited variations from expected outputs despite circuits being correctly insulated, through indirect burden on the host due to resource competition[40]. This module coupling occurs when synthetic circuits consume cellular resources, leaving less for another module and indirectly affecting its outputs[13]. This can be detrimental; for example, if synthetic circuits put a strain on RNA polymerases and ribosomes, essential host genes may experience transcriptional or translational bottlenecks, resulting in expression defects or cell failure. To improve robustness of orthogonal parts, future research should address both the 'leaky' insulation that persists- resulting in off-target effects- and the coupling that arises

indirectly through resource competition. These shortcomings reinforce the continuing need for computational approaches like WCMs to predict crosstalk and emergent behaviours that cannot be fully eliminated by orthogonal parts alone. Overall, orthogonal libraries to insulate circuit networks complement modelling strategies, forming a dual approach to improving modular and predictable circuit design.

Future Directions

So far, circuit testing has relied on either low-resolution models that are too coarse to be predictive, or high-resolution models optimised only for specific design parameters. Future models should balance accuracy with practicality, simulating important processes in molecular detail, and using different modelling scales such as coarse-graining, deterministic models and ML surrogates for broader details. Previous work in other fields has addressed this with hybrid models at subsystem scales[32]. These hybrid resolution models allow for the selective use of detailed, molecular simulation only where it is necessary[41,42,43,44]. Whilst hybrid WCMs are not currently optimised for circuit testing, they could offer accurate, scalable testing as synthetic circuits grow more complex in the future.

The next step is combining CAD and modelling into one standardised design and testing environment. If hybrid WCMs demonstrate optimised accuracy whilst retaining computational speed, they could be embedded into design software, facilitating a streamlined end-to-end platform for circuit construction and testing. The critical advancement would lie in the ability to test circuits against a full-cell model, not only checking for broad emergent behaviours but also supporting generalisability. Any circuit- whether metabolic, regulatory or biosensing- could be evaluated within the same platform as it maps the entire cell. Whilst each host needs a distinct model, that model could test any circuit designed for that organism, reducing reliance on distinct, bespoke frameworks for each new circuit, and accelerating design across different biological contexts. Additionally, Generative AI may be implemented into the CAD stages, accelerating part characterisation and circuit proposal. These would be validated against the hybrid WCM, creating a seamless design-test pipeline. Importantly, future AI use remains dependent on overcoming the same empirical data limitations that constrain conventional models.

To create a WCM with this hypothetical level of versatility, tuning features would be required for manual control over which subsystems are simulated in molecular detail, depending on their relevance to the circuit being engineered. Whilst spatially adaptive tuning does not yet exist for whole-cell contexts, current modular hybrid frameworks such as Vivarium allow researchers to model separate subsystems of the same network at different resolutions, using high-resolution submodels where desired[45]. Whilst further progress would be needed to make hybrid WCM tuning a reality, current multiscale model architectures highlight the feasibility of selective local refinement within cell models, forming a foundation for general, all-purpose hybrid WCM-CAD tools[46,43].

For a CAD-modelling framework to meet its full potential as a standardised all-in-one design tool, expanding part libraries is necessary to support the growing size and complexity of circuits[47]. Currently, part libraries suffer from narrow host contexts and limited characterisation data, resulting in unpredictable behaviour when integrating parts into new configurations[48]. This reduces the predictive ability of models and generalisability of design tools. Future progress towards standardised and generalisable design depends on collaborative efforts to build larger, orthogonal and well-characterised part libraries, providing design software with a richer, more reliable arsenal of parts. This development will be critical for scaling synthetic circuits, moving from bespoke construction to modular and automated biological engineering.

Although modelling approaches greatly improve predictive design, they cannot currently replace empirical testing. Even the most accurate models omit numerous unknown genes and pathways, meaning tests within living cells are still mandatory to ensure they survive, replicate and behave as expected[49,25]. This limitation additionally highlights circuit insulation as a parallel strategy to improve reliability. Through minimising unwanted circuit-host interactions they reduce emergent effects, increasing predictability and easing the demands placed on modelling accuracy[50].

The need for reliable and repeatable circuit design is motivated by the implications these advancements would have in medicine, biotechnology and biological research. Therapeutics would benefit from circuits capable of sensing and responding dynamically to disease markers, enabling more accurate cell therapies than currently

exist. An example of this is mitigating off-target toxicity, where healthy cells are attacked, in CAR-T cell therapy[51]. Circuit design holds potential for diagnostic purposes using biosensing circuits within engineered bacteria. These may detect early signs of disease *in situ*, such as mutations associated with cancers. In the case of colorectal cancer, the ability to detect disease signs at the site of pathology delivers accuracy and speed that conventional stool screening cannot match[52]. In biomanufacturing, reliable metabolic circuits would achieve more accurate control over biosynthetic pathways, demonstrated to improve pharmaceutical yields and reduce production costs[53]. Future progress towards reliable synthetic circuit production is therefore an essential prerequisite for synthetic biology to reach its full potential within the medicine and biomanufacturing fields.

Conclusion

Synthetic circuit design presents unique modelling challenges, as the disturbance of natural cell systems often induces problematic emergent behaviours through context effects and resource competition. Whilst testing models are frequently too limited in scale or resolution to accurately predict emergent behaviours, molecular WCMs remain prohibitively slow and expensive for routine and integrated testing.

The future calls for hybrid models with selective resolution, so systems relevant to a circuit can be modelled accurately, with broader cellular processes modelled in coarser detail or predicted using deterministic or ML models. Modelling the entire cell offers the dual advantage of capturing broad emergent effects and introducing a general framework for varying circuit types, reducing reliance on application-specific models. Additionally, it provides a scalable testing environment for future circuits; as synthetic modules become larger, they can still be tested within the full cellular context. Generative AI tools are beginning to automate circuit proposal within these frameworks, shifting design from manual iteration toward data-driven inference. Alongside modelling advances, orthogonal toolkit expansions will further improve reliable design, reducing unwanted circuit-host interactions.

Whilst insulated design and whole-cell simulations offer significant progress towards reliable and scalable circuit design frameworks, unknown and poorly characterised genes and pathways, simplified cellular processes and imperfect orthogonality restrict perfect predictions, and keep empirical testing essential for circuit verification.

Rather than replacing experimental tests, modelling and insulation improvements should be viewed as a tool for accelerating the build-test cycle, filtering out low-performing designs and increasing the number of circuits that behave as intended *in vivo*.

References

1. English MA, Gayet RV, Collins JJ. Designing biological circuits: synthetic biology within the operon model and beyond. *Annu Rev Biochem*. 2021;90:221-244. doi:https://doi.org/10.1146/annurev-biochem-013118-111914
2. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*. 1961;3(3):318-356. doi:https://doi.org/10.1016/s0022-2836(61)80072-7
3. Englesberg E, Squires C, Meronk F. The L-arabinose operon in Escherichia coli B/r. *Proc Natl Acad Sci USA*. 1969;62(4):1100-1107. doi:https://doi.org/10.1073/pnas.62.4.1100
4. Morse DE, Mosteller RD, Yanofsky C. Dynamics of synthesis, translation, and degradation of trp operon messenger RNA in E. coli. *Cold Spring Harb Symp Quant Biol*. 1969;34:725-740. doi:https://doi.org/10.1101/sqb.1969.034.01.082
5. Barajas C, Del Vecchio D. Synthetic biology by controller design. *Curr Opin Biotechnol*. 2022;78:102837. doi:https://doi.org/10.1016/j.copbio.2022.102837
6. Khalil AS, Collins JJ. Synthetic biology: applications come of age. *Nat Rev Genet*. 2010;11(5):367-379. doi:https://doi.org/10.1038/nrg2775
7. Del Vecchio D, Dy AJ, Qian Y. Control theory meets synthetic biology. *J R Soc Interface*. 2016;13(120):20160380. doi:https://doi.org/10.1098/rsif.2016.0380
8. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*. 2000;403(6767):335-338. doi:https://doi.org/10.1038/35002125
9. Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. *Nature*. 2000;403(6767):339-342. doi:https://doi.org/10.1038/35002131
10. Purnick PEM, Weiss R. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol*. 2009;10(6):410-422. doi:https://doi.org/10.1038/nrm2698

11. Kelwick R, MacDonald JT, Webb AJ, Freemont P. Developments in the tools and methodologies of synthetic biology. *Front Bioeng Biotechnol*. 2014;2:60. doi:https://doi.org/10.3389/fbioe.2014.00060

12. Cardinale S, Arkin AP. Contextualizing context for synthetic biology. *Biotechnol J*. 2012;7(7):856-866. doi:https://doi.org/10.1002/biot.201200085

13. Grunberg TW, Del Vecchio D. Modular analysis and design of biological circuits. *Curr Opin Biotechnol*. 2020;63:41-47. doi:https://doi.org/10.1016/j.copbio.2019.11.015

14. Qian Y, Huang HH, Jiménez JI, Del Vecchio D. Resource competition shapes the response of genetic circuits. *ACS Synth Biol*. 2017;6(7):1263-1272. doi:https://doi.org/10.1021/acssynbio.6b00361

15. Bedau MA. Weak emergence drives the science, epistemology, and metaphysics of synthetic biology. *Biol Theory*. 2013;8(4):334-345. doi:https://doi.org/10.1007/s13752-013-0139-6

16. Chakraborty D, Rengaswamy R, Raman K. Designing biological circuits: from principles to applications. *ACS Synth Biol*. 2022;11(4):1377-1388. doi:https://doi.org/10.1021/acssynbio.1c00557

17. Singh V. New Frontiers and Applications of Synthetic Biology. Amsterdam: Elsevier; 2022.

18. Rodrigo G, Jaramillo A. AutoBioCAD: full biodesign automation of genetic circuits. *ACS Synth Biol*. 2012;2(5):230-236. doi:https://doi.org/10.1021/sb300084h

19. Nielsen AAK, Der BS, Shin J, *et al.* Genetic circuit design automation. *Science*. 2016;352(6281):aac7341. doi:https://doi.org/10.1126/science.aac7341

20. Konur S, Mierla L, Fellermann H, *et al.* Toward full-stack in silico synthetic biology. *ACS Synth Biol*. 2021;10(8):1931-1945. doi:https://doi.org/10.1021/acssynbio.1c00143

21. Purcell O, Jain B, Karr JR, Covert MW, Lu TK. Towards a whole-cell modeling approach for synthetic biology. *Chaos*. 2013;23(2):025112. doi:https://doi.org/10.1063/1.4811182

22. Tomita M. Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol*. 2001;19(6):205-210. doi:https://doi.org/10.1016/S0167-7799(01)01636-5

23. Karr JR, Sanghvi JC, Macklin DN, *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell*. 2012;150(2):389-401. doi:https://doi.org/10.1016/j.cell.2012.05.044

24. Goldberg AP, Szigeti B, Chew YH, *et al.* Emerging whole-cell modeling principles and methods. *Curr Opin Biotechnol*. 2018;51:97-102. doi:https://doi.org/10.1016/j.copbio.2017.12.013

25. Macklin DN, Ruggero NA, Covert MW. The future of whole-cell modeling. *Curr Opin Biotechnol*. 2014;28:111-115. doi:https://doi.org/10.1016/j.copbio.2014.01.012

26. Georgouli K, Yeom JS, Blake RC, Navid A. Multi-scale models of whole cells: progress and challenges. *Front Cell Dev Biol*. 2023;11:1260507. doi:https://doi.org/10.3389/fcell.2023.1260507

27. Stumpf MPH. Statistical and computational challenges for whole cell modelling. *Curr Opin Syst Biol*. 2021;26:58-63. doi:https://doi.org/10.1016/j.coisb.2021.04.005

28. Sechkar K, Steel H, Perrino G, Stan GB. A coarse-grained bacterial cell model for resource-aware analysis and design of synthetic gene circuits. *Nat Commun*. 2024;15(1):1814. doi:https://doi.org/10.1038/s41467-024-46410-9

29. Weiße AY, Oyarzún DA, Danos V, Swain PS. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc Natl Acad Sci USA*. 2015;112(9):E1038-E1047. doi:https://doi.org/10.1073/pnas.1416533112

30. Gherman IM, Abdallah ZS, Pang W, *et al.* Bridging the gap between mechanistic biological models and machine learning surrogates. *PLOS Comput Biol*. 2023;19(4):e1010988. doi:https://doi.org/10.1371/journal.pcbi.1010988

31. Gherman IM, Sharma K, Rees-Garbutt J, *et al.* Accelerated design of Escherichia coli reduced genomes using a whole-cell model and machine learning. *Cell Syst*. 2025;16(10):101392. doi:https://doi.org/10.1016/j.cels.2025.101392

32. Feig M, Sugita Y. Whole-cell models and simulations in molecular detail. *Annu Rev Cell Dev Biol*. 2019;35:191-211. doi:https://doi.org/10.1146/annurev-cellbio-100617-062542

33. Costello Z, Martin HG. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ Syst Biol Appl*. 2018;4(1):19. doi:https://doi.org/10.1038/s41540-018-0054-3

34. Sivakumar N, Mura C, Peirce SM. Innovations in integrating machine learning and agent-based modeling of biomedical systems. *Front Syst Biol*. 2022;2:959665. doi:https://doi.org/10.3389/fsysb.2022.959665

35. Kim N, De Carluccio G, Zhang K, Collins JJ. Generative AI for synthetic biology: Designing biological parts, circuits, and genomes. *Cell systems*. 2026;17(2):101533. doi:https://doi.org/10.1016/j.cels.2026.101533

36. Abello Castillo L, Gutiérrez Pescarmona M. CELLM: Bridging Natural Language Processing and Synthetic Genetic Circuit Design with AI. *ACS Synthetic Biology*. 2025;14(9):3799-3803. doi:https://doi.org/10.1021/acssynbio.5c00391

37. Meyer AJ, Segall-Shapiro TH, Glassey E, Zhang J, Voigt CA. Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nat Chem Biol*. 2019;15(2):196-204. doi:https://doi.org/10.1038/s41589-018-0168-3

38. Jiang T, Teng Y, Li CY, *et al.* Establishing tunable genetic logic gates with versatile dynamic performance by varying regulatory parameters. *ACS Synth Biol*. 2023;12(12):3733-3742. doi:https://doi.org/10.1021/acssynbio.3c00554

39. Lu TK. Engineering scalable biological systems. *Bioeng Bugs*. 2010;1(6):378-384. doi:https://doi.org/10.4161/bbug.1.6.13086

40. De Marchi D, Shaposhnikov R, Gobaa S, *et al.* Design and model-driven analysis of synthetic circuits with the Staphylococcus aureus dead-Cas9 as a programmable transcriptional regulator. *ACS Synth Biol*. 2024;13(3):763-780. doi:https://doi.org/10.1021/acssynbio.3c00541

41. Kar P, Gopal SM, Cheng YM, *et al.* PRIMO: a transferable coarse-grained force field for proteins. *J Chem Theory Comput*. 2013;9(8):3769-3788. doi:https://doi.org/10.1021/ct400230y

42. Renevey A, Riniker S. Improved accuracy of hybrid atomistic/coarse-grained simulations using reparametrised interactions. *J Chem Phys*. 2017;146(12):124131. doi:https://doi.org/10.1063/1.4979128

43. Ward MD, Nangia S, May ER. Evaluation of the hybrid resolution PACE model for the study of folding, insertion, and pore formation of membrane associated peptides. *J Comput Chem*. 2017;38(16):1462-1471. doi:https://doi.org/10.1002/jcc.24694

44. Wassenaar TA, Ingólfsson HI, Prieß M, Marrink SJ, Schaefer LV. Mixing MARTINI: electrostatic coupling in hybrid atomistic–coarse-grained biomolecular simulations. *J Phys Chem B*. 2013;117(13):3516-3530. doi:https://doi.org/10.1021/jp311533p

45. Agmon E, Spangler RK, Skalnik CJ, *et al.* Vivarium: an interface and engine for integrative multiscale modeling in computational biology. *Bioinformatics*. 2022;38(7):1972-1979. doi:https://doi.org/10.1093/bioinformatics/btac049

46. Kar P, Feig M. Hybrid all-atom/coarse-grained simulations of proteins by direct coupling of CHARMM and PRIMO force fields. *J Chem Theory Comput*. 2017;13(11):5753-5765. doi:https://doi.org/10.1021/acs.jctc.7b00840

47. Bervoets I, Van Brempt M, Van Nerom K, *et al.* A sigma factor toolbox for orthogonal gene expression in Escherichia coli. *Nucleic Acids Res*. 2018;46(4):2133-2144. doi:https://doi.org/10.1093/nar/gky010

48. Aldulijan I, Beal J, Billerbeck S, *et al.* Functional synthetic biology. *Synth Biol*. 2023;8(1):ysad006. doi:https://doi.org/10.1093/synbio/ysad006

49. Rees-Garbutt J, Chalkley O, Landon S, *et al.* Designing minimal genomes using whole-cell models. *Nat Commun*. 2020;11(1):836. doi:https://doi.org/10.1038/s41467-020-14545-0

50. Darlington APS, Kim J, Jiménez JI, Bates DG. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat Commun*. 2018;9(1):695. doi:https://doi.org/10.1038/s41467-018-02898-6

51. Lu L, Xie M, Yang B, Zhao W, Cao J. Enhancing the safety of CAR-T cell therapy: synthetic genetic switch for spatiotemporal control. *Sci Adv*. 2024;10(8):eadj6251. doi:https://doi.org/10.1126/sciadv.adj6251

52. Cooper RM, Wright JA, Ng JQ, *et al.* Engineered bacteria detect tumor DNA. *Science*. 2023;381(6658):682-686. doi:https://doi.org/10.1126/science.adf3974

53. Li Y, Yang Y, Li L, Tang K, Hao X, Kai G. Advanced metabolic engineering strategies for increasing artemisinin yield in Artemisia annua L. *Hortic Res*. 2024;11(2):uhad292. doi:https://doi.org/10.1093/hr/uhad292

Authors Bio:

My name is George Gould and I am a third year Biology undergraduate at the University of Bristol. My interest in synthetic biology was developed through my second year modules on molecular biology and genetics, in which I was fascinated with the regulatory machinery that underlies gene expression. When the opportunity arose to explore circuit design for my dissertation through my supervisor's research, the topic immediately stood out to me. The unrefined nature of the field and innate stochasticity of living systems makes synthetic circuit design an open problem with enormous potential, particularly in medical applications such as cell therapies.

Philosophy and Humanities

Meet the Subject Editor

Allegra Gervasi
Philosophy and Humanities

Allegra is currently completing her Master's degree in Philosophy at the University of Bristol. Previously, she received her Bachelor's in Physics and Philosophy at the University of Leeds. Having focused her studies on analysing complex scientific concepts, she has developed an interest in science communication and journalism. She feels that the work of the philosopher, that is, clarifying and communicating meaningful ideas, is akin to the roles of journalist and editor. She enjoyed being the Philosophy and Theology Subject Editor and reading the manuscripts, which all covered lively and socially impactful areas of philosophy. She hopes to continue working in the field of journalism to facilitate conversations about science and beyond.

Interpreting and assessing the validity of Durkheim's Rules under conditions of structural injustice – a standpoint theory approach

Author: Sian Amekuedi

ORCID: 0009-0005-4156-523X

Programme: MScSocial Science Research Methods

Affiliation: SchoolofSociology, Politics and International Studies

Email: ea16634@bristol.ac.uk

Abstract

This article interrogates the validity of Durkheim's Rules of Sociological Method under conditions of structural injustice. We argue that Durkheim's Rules can only be considered partially valid. Understanding Durkheim's rules as built upon an epistemic foundation that implicitly assumes universal observability, we suggest that this makes social research vulnerable to socially embedded ignorance. Drawing on standpoint theory and a contemporary case study, we highlight that such oversights can lead to epistemic and ethical harms. We then turn to standpoint theory for a potential solution, proposing an epistemic shift in line with contemporary medical research. Such an epistemological shift will bring with it methodological consequences. Hence, we speculate on a potential reformulation of Durkheim's fundamental rule that the sociologist must "treat social facts as things". Using Durkheim's own morphological/physiological distinction, we propose a reformulation of this rule, Rule 1*. The aim of Rule 1* is to encourage greater reflexivity and awareness of situated knowledge. We conclude that under conditions of structural injustice, Durkheim's Rules may be rendered conditionally valid.

Keywords

Durkheim,Social Epistemology, Standpoint theory, Methodology, Structural Injustice

Introduction

Durkheim's (2013) *Rules of Sociological Method* aimed to serve as a sociological manifesto. It offered an early framework for sociological research design. Durkheim's ultimate goal was for sociology to be a moral science. A scientific discipline capable of diagnosing social pathologies that undermine social solidarity (Lukes, 1985, pp. 193,226). Through diagnoses, sociology could prescribe directions for social reform and eventual social order (McLennan, 2011, p. 63).

We suggest that Durkheim's rules are only partially valid under conditions of structural injustice. However, they can be rendered conditionally valid when reformulated. Validity is assessed in relation to Durkheim's diagnostic aim. We suggest that structural injustice generates potential epistemic limitations. First, we establish the relationship between Durkheim's epistemology and methodology. Secondly, we introduce a case study and consider the validity of Durkheim's rules under conditions of structural injustice. Thirdly, standpoint theory is introduced as a comparison to signal Durkheim's epistemic limitations – suggesting only partial validity. Finally, we consider whether Durkheim could adopt a standpoint-based epistemology. A reformulation of his fundamental rule is suggested, potentially allowing conditional validity.

For clarification, our understanding of Durkheim focuses on the Rules of Sociological Method. Other works will only be used illustratively. Durkheim's primary rule will be the focus, with further rules not being discussed in depth. Throughout, we refer to and build upon standpoint theory.

Durkheim at a glance:

This section outlines Durkheim's methodology and the epistemological assumptions upon which the validity of Durkheim's rules depends. Methodologically, Durkheim was a naturalist; he believed methods of the natural sciences should be replicated in sociology (Baert, 2005, p. 12). For Durkheim, scientific status required starting from observable and measurable data. Hence, the social fact:

Externality *- social facts exist outside of any individual consciousness.*

Constraint *– social facts impose rules and expectations, sometimes without our awareness.*

Generality *- social facts cannot be reduced to individual actions.*

Durkheim (2013, p. 7-9) asserts that sociologists must treat social facts as things. We will call this “Rule 1”. By “things”, Durkheim entails that social facts are impenetrable to the observer’s introspection and thus require external methodological study. Three corollaries immediately followed. Firstly, preconceptions must be eliminated. Secondly, social facts must be defined by their external observable characteristics. Thirdly, social facts should be studied independently of individual manifestations. Together, these corollaries aimed to ensure research objectivity. Rule 1 is the basis from which the subsequent rules follow. The subsequent rules establish a quantitatively focused methodology aiming to produce sociological ‘proof’ grounded in claims of causality. Durkheim’s epistemology can be framed as a mixture of scientific rationalism and positivism (Lukes, 1985, pp. 72-76; Lukes, 2013, p. xii). Broadly, this position has been characterized by Bordo (1986) and Haraway (1988) as valuing rationality, objectivity, detachment, and a commitment to universal observability – a ‘view from nowhere’. This is evident in Rule 1 and its corollaries. Durkheim aims to ground sociological arguments in systematic reasoning, explicitly aiming for an objective method. Through this, he seemingly assumes that the trained observer can bracket off these biases when examining social facts.

So, Rule 1 is the basis for Durkheim’s methodology, and his methodology is seemingly built upon his epistemology. Together, these suggest that the validity of Durkheim’s rules is dependent upon their underlying epistemological validity.

Turning to a contemporary example, this provides the intuitive grounds for the limitations of the seeming epistemological assumptions that ground his rules.

Chambers et al. as a case study:

Chambers et al. (2021) look to understand how structural racism influences reproductive health. This case study aims to indicate the validity of alternative approaches in the context of racism. Specifically, how such approaches allow access to knowledge that risks being excluded under Durkheim's framework.

Chambers et al. (2021) note that racism/misogynoir operates at a societal level, imposes a constraining force, and has generality across ethnically minoritized women. By Durkheim's own criteria, racism could be understood as a social fact. Similarly to Durkheim, Chambers et al.'s central aim is to prescribe policy changes to combat a social pathology. Methodologically, however, they deviate significantly from Durkheim's prescriptions. Their research design involves the following steps:

1. Defining the research problem.
2. Understanding their positionality.
3. Sampling participants based on lived experience.
4. Data collection via focus groups.
5. Data analysis.
6. Checking findings with an advisory board in the field.
7. Gathering feedback from participants to affirm the representativeness of findings.

Chambers et al. (2021) implement a small-N approach prioritising depth over generalisability. In contrast to Durkheim, they avoid the assumption that biases can be bracketed away through an objective method. Instead, they aim for reflexivity, seeking to mitigate and recognize the impact of external power dynamics. This is first evident in step 2 - by acknowledging their positionality, Chambers et al. encourage themselves to interrogate how this has come to shape the research process. In the case of step 3, this aims to ensure that research findings are derived from those with lived experience. Steps 7 and 6 moreover, look to integrate whether participants and the wider community have been adequately represented by researcher attitudes alongside assumptions.

As Bryman (2008, pp. 391-392) notes, however, such studies are also limited by their subjectivity and replicability. From a Durkheimian perspective, this small sample would provide an extremely limited basis for an inductive generalisation. Moreover, its reliance

on lived experience undermines its objective basis. Durkheim (2013 p. 24) instead prioritises generalisability reflected in his use of official secondary data such as government statistics. By doing this, Durkheim sought detached, repeatable causal explanations.

Each of these approaches requires respective trade-offs and fundamentally different approaches to researcher bias. They each bring with them alternative epistemological assumptions. Chambers et al. sacrifice generalisability to access knowledge that purely quantitative methods can obscure. As suggested in the next section, Durkheim's epistemic commitments are vulnerable to blind spots. Unfortunately, these blind spots can have ethical consequences under conditions of structural injustice.

Standpoint Theory and Epistemic Blind Spots:

This section examines standpoint theory to explain how Durkheim's epistemological assumptions may limit the applicability of his rules. Appealing to standpoint theory, this section critiques the assumption of universal observability under conditions of structural injustice.

Statements of positionality within research (like those of Chambers et al.) are epistemic consequences of standpoint theory (Freeman, 2019). Standpoint theory has been characterized by Crasnow (2014) as maintaining three key theses. Firstly, knowledge is always tied to social positionality. Secondly, those in marginalized positions have access to knowledge invisible to dominant social groups. And thirdly, the achievement thesis – The standpoint is not automatic; it is cultivated through political and collective work. Methodologically, this can manifest as placing lived experience as the object of study, as done by Chambers et al. (2021).

Together, this immediately contrasts with Durkheim's seeming epistemological assumptions. In particular, his seeming assumption of universal observability. Haraway (1988, pp. 581-584) suggests that by assuming universal observability, the roles of societal power dynamics in shaping knowledge are ignored. Thus, the researcher's

positionality can come to shape how they interpret data. Hence, steps 2, 6, and 7 in Chambers et al.'s (2021) research design.

Under conditions of structural injustice, a researcher assuming universal observability may unknowingly reproduce the social pathologies Durkheim is committed to diagnosing. Admittedly, Durkheim (2008) was not entirely blind to the existence of structural discrimination. In his account of antisemitism, he acknowledges that it meets each of his criteria of a social fact and is hence structural in nature. Nevertheless, Durkheim saw structural discrimination as resulting from factors such as social crises. What he seemingly didn't foresee was the possibility of social pathologies being embedded within the researcher's epistemology, shaping their ways of knowing.

Epistemic ignorance illustrates one way that social pathologies may persist undetected within an epistemology and act as a constraining force. It can be described as socially supported and strategic instances of not knowing that result in the reproduction of existing power structures. It includes bias, conceptual blind spots, denial, and systematic direction (Tuana, 2006; Mills, 2017). Crime statistics exemplify how the methodological consequences of ignorance may manifest in research.

Spohn (2015) argues that through the war on drugs, black men faced disproportionately harsher policing. This included harsher sentencing and racial profiling. Similar to Durkheim's suicide example, biases have shaped what data looks like. Naturally, this meant that subsequent crime rates would reflect patterns of enforcement over offending. Durkheim's suicide study indicates how blind spots may arise.

In Suicide, Durkheim partially overlooks biases that shape the production and reliability of data. Durkheim (2002, pp. xxxix-xl) opens with a warning of common prejudices of the public that distort researchers' understanding. However, Durkheim's own study relied heavily on the use of secondary statistics, assuming their neutrality. Though, as Montuschi (2014, pp. 135-138) highlights, Durkheim underestimates the role of moral norms in shaping secondary data. For example, the societal stigma around suicide may have led to families lying.

The implication here is that Durkheim may lack the tools for identifying bias in secondary data, leading it to be taken as neutral evidence. Returning to the example of racialised policing, when a group is profiled and faces harsher sentencing, this is going to be reflected in official crime statistics. Accordingly, the use of biased data is likely to produce biased research. Hence, in the case of structural injustices, such oversights while claiming objectivity risk obscuring bias rather than challenging it. Unfortunately, the inability to identify biases presents a potential epistemic barrier to Durkheim's diagnostic aim.

Critically, Tuana (2006) argues that epistemic ignorance can be a systematic consequence of assuming universal observability. This can occur despite claims of objectivity. The Durkheimian researcher, then, may risk unknowingly reproducing structural bias within research from dominant social categories that shape data. For example, in the case of crime statistics. Therefore, Rule 1 and its corollaries may be insufficient on their own to adequately guard against systematically embedded epistemic ignorance. Such ignorance can contribute to ethical harm.

The British Sociological Association (2017, pp. 4-6) highlights the researcher's responsibility to safeguard those affected by their work. Likewise, the need to consider the effects and use of the research. Jensen (1969) for instance, attempted to correlate IQ with race. Through a different methodology, he drew upon 'objective' positivist methods – obscuring his own bias.

Responding to Jensen, Flynn (1980, pp. 15-20) asserts that such social research does not exist in an ethical vacuum. It can provide the racist with a 'rational justification' of their attitudes and/or racist policy. Hence, commitments to neutrality and objectivity can obscure embedded biases in research. Research can be mobilized to intentionally, or otherwise produce ethical harm. With epistemologically embedded ignorances, it follows that Rule 1's corollaries may be insufficient to remove researcher bias. Having understood Durkheim's epistemic limitations, the next section looks to the potential answer from standpoint theory.

Conversely, while standpoint epistemologies cannot guarantee the elimination of epistemic ignorance, they can offer a conceptual framework for uncovering it. Using the

example of the women's health movement, (Tuana, 2006) alludes to ignorances being uncoverable by collective political work. This mirrors Collins's (2000, pp. 260-269) assertion that dialogue is a necessary ingredient for developing the standpoint. Methodologically, this is reflected in Chambers et al.'s (2021) implementation of focus groups.

Blakemore (2011) describes radical feminist consciousness-raising as women gathering to understand their experiences of misogyny. Through this, they would develop a conceptual framework. Montell (1999) flags the consciousness-raising power of focus groups. This is through factors such as the creation of participant solidarity and dialogue that would not arise through the power dynamics of the interview. Through focus groups, Chambers et al. facilitate collective meaning-making, fostering political dialogue. Thus, Chambers et al present a potential methodological response to the problem of ignorance. For Durkheim, however, Chambers et al's use of focus groups challenges his doubt of introspection. Montell (1999, p. 51) suggests that collective meaning-making can emerge through collective individual reflection. For (Chambers, et al., 2021), participants are collectively performing acts of introspection to foster dialogue. In doing so, participants are accessing situated knowledge of a social fact that Durkheim's quantitative methodology may not capture.

The challenges in this section suggest that Durkheim's epistemology is vulnerable to blind spots that may manifest as epistemic ignorance. The existence of this epistemic barrier implies that when put into practice, a Durkheimian could unknowingly reflect structural injustices. Hence, the epistemology that grounds Durkheim's rules potentially threatens their validity. Standpoint theory, however, provides epistemic tools for identifying these blind spots. The final section looks to assess whether Durkheim's rules may be rendered valid under epistemological revision.

(Re)Interpreting Durkheim

This section speculates on the consequences of Durkheim adopting a standpoint-based epistemology. In doing so, it suggests that Durkheim's rules can be rendered conditionally

valid. Instead of rejecting Rule 1 outright, this section offers an alternative formulation. Developments within medical epistemology provide an illustrative comparative case. Since the publication of *The Rules of Sociological Method*, health epistemology has experienced what Kuhn (1970) would call a paradigm shift. Specifically, it has seen a greater openness to standpoint epistemologies. Historical biases stemming from the universal observer have led to failures in understanding phenomena, causing individual harm. Grasswick (2017), for example, highlights a historical androcentric understanding of heart disease as a particular case of this phenomenon.

In response, at least two related subsequent developments can be seen within more contemporary research that reflect an epistemological shift. Firstly, health research is seeing increased scepticism towards the universal observer, as reflected in the example of McGregor's (2020) call for a greater diversity of practitioners. The implication here is that greater epistemic diversity can assist in combating shared reality bias. Anderson (2012, pp. 167-170) describes this as the convergence of perspectives and judgements amongst those who commonly interact. Anderson adds that when marginalised knowers are excluded from enquiry, the interpretive resources can be insufficient for understanding the experiences of those marginalised knowers. Secondly, as we have seen with (Chambers, et al., 2021), some areas of medical research have adopted alternative methodologies more suited to capturing standpoint-dependent knowledge.

Durkheim's epistemology can be vulnerable to similar problems. However, Durkheim's methodology was the means through which his diagnostic goal could be achieved (Baert, 2005, pp. 10-13; Giddens, 1985, pp. 10-11) . Moreover, the intellectual climate contemporary to Durkheim was dominated by positivism, with the natural sciences seen as the most reliable routes to knowledge (Lukes, 1985, p. 67). From this perspective, Durkheim's epistemological commitments could be understood as historically situated. In line with contemporary medicine, this perhaps suggests making a similar epistemic shift. This could equip Durkheim with tools to identify epistemic ignorance, thereby encouraging the conditional validity of his rules. Adopting such a shift raises the question of whether his rules could be adapted.

Methodologically, this would not require Durkheim to abandon quantitative methods. As Neitz (2021, pp. 58,68) note, some standpoint theorists are overtly open to methodological plurality. Sprague & Zimmerman (1989) assertion puts this into perspective. They forward that qualitative methods can also reproduce dominant ideology through researcher privilege. Additionally, they suggest that quantitative methods can be a powerful tool for identifying patterns of structural inequality.

Hence, while Durkheim would not need to abandon his methodology entirely, it would need to be used as one tool amongst many. It would, of course, need to be used reflexively, with a greater openness to how data categories can be socially produced. Under these conditions, his rules may maintain conditional validity while recognizing the potential for epistemic ignorance.

If the axioms of standpoint theory are to be accepted, Durkheim will need to reformulate rather than abandon Rule 1. As Rule 1 serves as the foundation of his methodology, abandoning it means abandoning the remaining rules. Therefore, any reformulation of Rule 1 must retain how social facts are treated. But it must account for how knowledge is accessed according to standpoint theory. On this basis, the following reformulation of Rule 1* is proposed as a speculative solution:

Rule 1*:

- *Treat social facts* as things (reflexively). If and only if morphological.
- Treat social facts as things (reflexively) with standpoint-dependent features only uncoverable via collective introspection under *specific conditions.* If and only if physiological.

Durkheim (2013, pp. 25-27) describes morphological social facts as material conditions that shape society. For example, population density or resource distribution. Physiological social facts are understood as dynamic aspects of life, such as norms, public ideology, or laws. Racism and instances of structural injustice would therefore fall into this category. Physiological social facts can give rise to morphological social facts. For example, racism may impact child mortality rates (Chambers, et al., 2021). Moreover, physiological social facts can generate further physiological social facts, as evidenced through racism's role

in the creation of segregationist laws. In each case, structural social facts can produce further social facts. Hence, distinguishing social facts through Rule 1* requires considering how structural social facts may have brought others into being and what investigation should look like.

In the case of morphological social facts, Rule 1* maintains that knowledge is not accessible via introspection. For example, if conducting a study of population density. Importantly, however, this does not always entail that quantitative methods are suitable for examining these social facts. Spatial arrangements, for example, may be studied both quantitatively and qualitatively. Moreover, the reflexivity clause is added to signal that these social facts remain epistemically non-neutral. Reflexivity becomes a necessary tool for interrogating how categories and measurements are produced.

Lammer (2024) provides an illustration of what quantitative reflexivity could look like. It begins with transparency about why a research topic was selected, and its relation to researcher positionality. Because positionality shapes enquiry, this entails that social research can benefit from greater diversity amongst researchers. At each subsequent stage of research, Lammer adds that researchers must transparently and critically examine how their standpoint, biases, and assumptions may shape the study. If we extend this to the context of secondary data, such as crime statistics, the reflexivity clause requires researchers to acknowledge structural forces that may have shaped the production of that data.

Nevertheless, while morphological social facts can indicate patterns of structural injustice, they cannot uncover how structural injustices are experienced by those subjected to them. This highlights the value of methodological pluralism, as different kinds of social fact may require different forms of investigation.

In the case of physiological social facts, introspection is restricted to when the standpoint may arise. These will be cases concerning the study of a structural social fact, such as racism. Here, collective introspection aims to provide the means for accessing standpoint-dependent knowledge. Importantly, this introspection must be combined with the dialogue needed for standpoint. By specifying “only uncoverable under certain conditions”, we allow the three theses of standpoint theory outlined by Crasnow (2014) to be maintained.

In both physiological and morphological cases, Rule 1* aims to encourage criticality of the power dynamics that shape social facts of each kind; thus, the reflexivity clause. Reflexivity furthermore is a methodological commitment of standpoint theory. Through adopting standpoint theory, this necessitates the adoption of alternative epistemic virtues. The switch from (weak) objectivity to strong objectivity is a clear example of this. Harding (1991, pp. 138-163; 1992) forwards that objectivity is not gained through neutrality, but reflexivity. Instead, it's gained through understanding how power dynamics shape knowledge. It further requires epistemic diversity and the rejection of universal observability.

The reformulation proposed in this section is intentionally limited to Rule 1. Rule 1 is the methodological foundation of Durkheim's rules. So, this suggests that the subsequent rules and corollaries could plausibly retain conditional validity under an epistemic shift. However, what these would look like and indeed whether it would be possible remains an open and speculative question. It may require further revisions to Durkheim's rules and this is outside the scope of this essay. It could, however, be an area for future investigation.

Conclusion:

Durkheim's goal of establishing sociology as a diagnostic force remains a noble goal, still pursued in social research. In addition, his formulation of the social fact effectively produced an early conception of systemic injustice. This is a significant achievement for the time, one that deserves greater recognition. Despite this, when considering racism as a social fact, potential epistemic and ethical issues arise. Durkheim's epistemic commitment to detachment, objectivity, and universal observability potentially allows epistemic vices. Namely, epistemic ignorance. Epistemology shapes methodology, and together they can shape ethical outcomes. From this perspective, Durkheim's rules risk being an insufficient route to his aims because of Durkheim's epistemic foundations. Standpoint theory, on the other hand, may offer a solution. Adopting it, Durkheim could gain an epistemology more suited to achieving his diagnostic aims. Of course, adopting a new epistemology brings baggage. In particular, a new set of epistemic virtues and

vices as well as a need for methodological reformulation. In this essay, we have established what a reformulation of Rule 1 may look like – Rule 1*. If Durkheim accepts this reformulation, it could provide the validation of Rule 1 without discarding the entire methodology by virtue of pluralism. On this basis, we should consider Durkheim's rules conditionally valid, dependent upon epistemic revisions.

References

Anderson, E., 2012. Epistemic Justice as a Virtue of Social Institutions. *Social Epistemology,* 26(2), pp. 163-173.

Baert, P., 2005. *Philosophy of the Social Sciences: Towards Pragmatism.* 1st ed. Cambridge/Malden: Polity Press.

Blakemore, E., 2011. *Consciousness-Raising Groups and the Women's Movement.* [Online]
Available at: https://daily.jstor.org/consciousness-raising-groups-and-the-womens-movement/
[Accessed 2025 12 05].

Bordo, S., 1986. The Cartesian Masculinization of Thought. *Signs,* 11(3), pp. 439-456.
British Sociological Association, 2017. *Statement of Ethical Practice.* Durham: British Sociological Association.

Bryman, A., 2008. *Social Research Methods.* 3rd ed. Oxford: Oxford University Press.
Chambers, B. D. et al., 2021. Black Women's Perspectives on Structural Racism across the Reproductive Lifespan: A Conceptual Framework for Measurement Development.

MaternalandChildHealth Journal, 25(3), pp. 402-413.

Collins, P. H., 2000. *Black Feminist Thought:Knowledge,Consciousness,andthe PoliticsofEmpowerment.* 2nd ed. New York and London: Routledge.

Crasnow, S., 2014. Feminist Standpoint Theory. In: N. Cartwright & E. Montuschi, eds. *Philosophy of Science: A New Introduction.* Oxford: Oxford University Press, pp. 145-161.

Durkheim, É., 2002. *Suicide: A Study in Sociology.* 2nd ed. London: Routledge.
Durkheim, E., 2008. Anti-Semitism and Social Crisis. *Sociological Theory,* 26(4), pp. 321-323.

Durkheim, E., 2013. The Rules of Sociological Method. In: S. Lukes, ed. *The Rules of Sociological Method.* Second ed. Basingstoke: Palgrave Macmillan, pp. 3-120.

Flynn, J. R., 1980. *Race, IQ and Jensen.* 1st ed. London, Boston and Henley: Routledge & Kegan Paul.

Freeman, E., 2019. *Feminist Theory and Its Use in Qualitative Research in Education.* [Online]
Available at:
https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefor

e-9780190264093-e-1193
[Accessed 30 11 2025].

Giddens, A., 1985. *Durkheim.* 4th ed. London: Fontana Paperbacks.

Grasswick, H., 2017. Epistemic injustice in science. In: I. J. Kidd, J. Medina & J. Gaile Pohlhaus, eds. *The Routledge Handbook of Epistemic Injustice.* Abingdon-on-Thames : Routledge, pp. 313-323.

Haraway, D., 1988. The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies,* 14(3), pp. 575-599.

Harding, S., 1991. *Whose Science? Whose Knowledge? Thinking from Women's Lives.* 1st ed. New York: Cornell University Press.

Harding, S., 1992. Rethinking Standpoint Epistemology: What is Strong Objectivity. *The Centennial Review,* 36(3), pp. 437-470.

Jensen, A. R., 1969. How Much Can We Boost IQ and scholastic achievement?. *Harvard Educational Review,* 39(1), pp. 1-123.

Kuhn, T. S., 1970. *The Structure of Scientific Revolutions.* Second ed. Chicago: The University of Chicago Press.

Lammer, L., 2024. Why we should all be reflexive: The case for reflexivity in quantitative science. *Journal of the Student Network for Open Science,* 12 December .

Lukes, S., 1985. *Emile Durkheim: His Life and Work.* 2nd ed. Stanford: Stanford University Press.

Lukes, S., 2013. Introduction to This Edition. In: S. Lukes, ed. *The Rules of Sociological Method.* Basingstoke: Palgrave Macmillan, pp. xi-xxxiii.

McGregor, A. J., 2020. Gender Inequity in Academic Medicine: A Visible Sign of a Disseminated Disease. *Journal of Women's Health,* 29(2), pp. 137-138.

McLennan, G., 2011. *Story of Sociology.* First ed. London: Bloomsbury Academic.

Mills, C. W., 2017. *Black Rights/White Wrongs: The Critique of Racial Liberalism.* 1st ed. Oxford: Oxford University Press.

Montell, F., 1999. Focus Group Interviews: A New Feminist Method. *The National Women's Studies Association Journal,* 11(1), pp. 44-71.

Montuschi, E., 2014. Scientific Objectivity. In: N. Cartwright & E. Montuschi, eds. *Philosophy of Science A New Introduction.* Oxford: Oxford University Press, pp. 123-144.

Neitz, M. J., 2021. Feminist Methodologies: Feminist standpoint analysis. In: S. Engler & M. Stausberg, eds. *The Routledge Handbook of Research Methods in the Study of Religion.* London: Routledge, pp. 57-71.

Spohn, C., 2015. Race, Crime, and Punishment in the Twentieth and Twenty-First Centuries. *Crime and Justice,* 44(1), pp. 49-97.

Sprague, J. & Zimmerman, M. K., 1989. Quality and Quantity: Reconstructing Feminist Methodology. *The American Sociologist,* 20(1), pp. 71-86.

Tuana, N., 2006. The Speculum of Ignorance: The Women's Health Movement and Epistemologies of Ignorance. *Hypatia,* 21(3), pp. 1-19.

Authors Bio

Sian Amekuedi is an MSc Social Science Research Methods student. Their undergraduate was in Mathematics and Philosophy, and since then, they have maintained a substantial interest in social epistemology, particularly standpoint theory. Sian especially enjoys the social epistemology of science and social science. Through writing this, they came to realise that they had judged had Durkheim far too harshly. Recently however, they have been developing a growing curiosity about social network analysis. Outside of academics, Sian really enjoys DJing, boredom, and Doctor Who.

Susan Moller Okin on Women's Choices: Freedom of Choice or False Consciousness?

Author: Antonia Milla

ORCID: 0009-0000-9631-8897

Programme: MScPoliticalTheory; BA Politics and French

Affiliation: SchoolofSociology, Politics and International Studies

Email: a.milla@lse.ac.uk

Abstract

The aim of this paper is to critique Susan Moller Okin's assumptions, which I argue undermine women's freedom of choice. This creates an internal tension in Okin's work, as she herself advocates for liberal, Rawlsian policy, and focuses heavily on freedom of choice. This tension describes the two sides of Okin's work, which is torn between liberal and radical feminist theory. I use normative theory to discuss why Okin's assumptions of false consciousness have problematic implications. My critique adopts a broadly liberal perspective. I first establish how Okin's more liberal work highlights lack of freedom of choice as a major problem in heterosexual families. I then go on to argue that Okinian female false consciousness extends to career and childcare choices, beyond what previous critics have argued. Finally, I analyse Okin's solutions for equality of outcome in families and workplaces. I conclude that Okin's focus on the freedom of choice is undermined by her more radical ideas.

Keywords: feminist political theory, liberalism, freedom of choice, female labour, employment discrimination

Introduction

Susan Moller Okin's feminist political theory focuses heavily on freedom of choice. This normative article argues that this focus is compromised as she constrains this freedom herself. I critique Okin as I argue that she makes assumptions of female *false consciousness* in more of her work that she acknowledges. False consciousness is defined as beliefs held by women which contradict their own interests and are shaped by and complicit in their own sexist oppression. *Freedom of choice* is defined as women's autonomous opportunity to make life decisions. I explore this as a negative freedom, i.e. the possibility of choosing without interference. My critique of Okin adopts the broadly liberal perspective she herself mostly defends. The first section establishes how the liberal Okin initially highlights a lack of freedom of choice as a major problem in heterosexual families. Section two argues that Okinian female false consciousness extends to career and childcare choices. The final section discusses Okin's solutions for equality of outcome in families and workplaces. I conclude that Okin's focus on the freedom of choice as part of her diagnosis is undermined by her more radical ideas. I also discuss a distinction between Okin's ideal and non-ideal theory.

The liberal Okin's focus on choice

Okin contradicts herself as she seeks to reconcile her liberal convictions with more radical normative solutions. This section highlights the importance the liberal Okin places on the negative freedom of choice. This supports my later critique of Okin's inconsistency and radicalism. Okin's liberal focus on how gender constrains choices includes a Rawlsian policy to 'protect those (perhaps mostly, but not exclusively, women) who **chose** to undertake the bulk of unpaid family work' (2005: 240, emphasis added). This element of choice implies that her normative vision includes at least the option for more women to be mainly responsible for childcare should they wish. Okin also centres women's ability to choose in policy proposals. Parental leave and subsidised childcare (Ibid.: 244-245) are endorsed by Okin so women 'like men, **can** work in paid jobs without being exploited and discriminated against' (Ibid.: 245, emphasis added). This shows Okin's analysis that exploitation and discrimination constrain women's negative freedom. She sees a problem

with interfering in women's free choices. Okin's concern here seems to be women's ability to choose their own path. This thread of women's negative freedom of choice being constrained runs through Okin's empirical evidence too, such as society not valuing childbearing as a resource equal to traditional wage labour (2016: 72). This disallows women the breadth of choices available to men. Their undervalued labour leaves women financially dependent on their partners and constricts their ability to flourish in society as a man does. Based on Okin's diagnosis of the problem, we can thus imagine a different future for women. Okin describes how ideally women may choose a traditional path of 'unpaid family work' but retain the same breadth of options as men. Okin further underscores this liberal outlook when she says family justice can be achieved through unintrusive, society-focused policies like state-subsidised childcare, gender-neutral parental leave, and reformed divorce law (2005: 240-241; 246). Family justice thus requires '[n]o household spies nor 'kitchen police'' (Ibid.: 246). All this points towards a liberal feminist approach that opens up women's choices. This should apply even if they choose traditional gender roles at home or at work, so long as they make their choice freely.

False consciousness

Despite her liberal focus on choice and endorsement of non-intrusive policy, there is a more radical current in Okin's work.

Okin has previously been critiqued for promoting the idea that women suffer from false consciousness in relation to women in the Global South. In an earlier paper Okin explicitly names her concern for 'false consciousness' in the context of those specific group differences negatively influencing feminist epistemology (1994: 5). Specifically, Okin argues that an outside observer might be better able to support those women in the Global South who do not recognise their own suffering but have their welfare diminished nonetheless (Ibid.: 18-19). In line with this concern for diverse cultural practices, Okin also proposes much more intrusive restrictions on religious freedom, potentially including women's practices (2005: 242), which can also be interpreted as her interpreting women's beliefs as misguided. Okin's turn to false consciousness concerning cultural issues has

been heavily criticised. According to Elizabeth Wingrove, Okin introduces the notion of false consciousness because she would be more likely to defend and maintain a more objective distance between an issue and those trying to diagnose it. Wingrove sees this particular error by Okin as isolated to this context. She argues it is rooted in Okin's methodological background as a political scientist, which then emerges in this normative context. (Wingrove, 2009: 63-64) Whether Wingrove is correct in her diagnosis or not, restricting religious choice, where religion is freely chosen, is hardly in line with the liberal concern for negative freedom of choice in which Okin grounds much of her analysis. Wingrove points out this issue in the context of Okin's more controversial 1994 paper. However, I now argue that the issue of false consciousness can *also* be found in Okin's feminist work concerning the kind of liberal Western state she inhabited. Specifically, I argue that Okin can be read as seeing false consciousness in women's and girls' career choices. False consciousness is an idea originally from Marxism, which was particularly popular among Women's Liberation activists in second-wave feminism. The idea is that the oppressed, in this case women, become deluded and mistakenly adopt the belief that they benefit from their own oppression. This line of argument has been applied to female practices such as make-up and housework. (Widdows, 2022: 28-29) My argument in regard to Okin is *not* that she endorses false consciousness to describe all women's current circumstances. After all, Okin herself (2016: 68-69) demonstrates that many women currently benefit materially from being stay-at-home parents because of the wage gap and related employment issues. My argument is *instead* that in an idealised world where gender no longer constrains women in this way, Okin would regard housewives as operating under false consciousness. This is based on the more radical strand in Okin's work, which undermines the way she champions freedom of choice elsewhere.

First, let me set out that Okin says:

> '**A just future** would be **one without gender**',
> where
> 'it would be a cause for surprise, and **no little concern**, if men and women were **not equally responsible** for domestic life or if children were to spend **much more time with one parent** than the other',

and where

‘men and women [participate] in **more or less equalnumbers** in every sphere of life, from infant care to **differentkinds of paid work** ’ (Ibid.: 79).

Note that the very concept of a stay-at-home parent would be concerning to Okin, whichever gender they may be. Equally, Okin wants women and men to participate in different professions equally, no matter their personal preferences. Here Okin moves from non-ideal into ideal theory: the idealisation comes in as Okin’s imagined future lacks key elements of unjust gender oppression present in our actual society in order to achieve this equality. In this ideal theory she proposes a world where women and girls have *complete equality of outcome no matter their choices*, a point which I elaborate in section 3. This also means that, if in an ideal world (seemingly also one without gender oppression) women would still want to have these traditional roles, it would be concerning for Okin. This, in turn, means that *any and all* differences between women and men, which Okin describes in her quote, are due to gender oppression, as they would stop existing in an *ideal* world. We can therefore assume that any women who feel that they are choosing a traditional gender role freely in this scenario do so out of false consciousness. This also has implications for women who may be less constrained in their choices nowadays due to economic or social privilege, as they would also not freely choose traditional roles according to Okin’s framework.

Inow turn to my example. Okin’s arguments around *women’s career and childcare choices* and *girls’ choices of profession* exemplify the analysis of female false consciousness which runs throughout her work. Regarding women, Okin argues that sexist power relations lead to discrimination in the workplace and women choosing lower-paying professions (2016: 69). Because of these negative consequences, 'wives (...) **come to perceive** themselves as benefiting from giving priority to their husbands’ careers.’ (Ibid.: 68 emphasis added). Okin assumes that these wives’ choices are never freely or deliberately made but exclusively based on gender oppression. In many cases this will likely be correct and these women’s choices are therefore at times based on correct assumptions. Many women will currently be perceiving objective facts. However, we will

see that, if women were to continue seeing traditional gender roles as favourable in a world without this kind of gender oppression, it would be acting against their own best interests. Okin believes that it is women's own perception which ultimately leads them into oppression.

Indeed, with their careers suffering, these women then turn towards childcare, as they '**see** the benefits of domestic life **as** greater than the costs', leading to mutually reinforcing inequalities at home and work (Ibid.: 69 emphasis added). Okin's continued argument for women's false consciousness is also evident in the words I have emphasised, which introduce subjective uncertainty and leave open the possibility of Okin disagreeing with these women's own realities. While many women might have chosen differently in an ideal world, we cannot assume on their behalf that all would. Leading to these choices, says Okin, are more problematic feminine assumptions. Young girls' socialisation towards marriage and motherhood makes them more likely to make choices that inhibit their future career progress or earnings (Ibid.: 64-65), such as regarding their education and 'degree of purposiveness about careers' (Ibid.: 65). Consequently, a world of 'occupational sex segregation' emerges, with women working in lower-paid fields and roles without advancement possibilities (Ibid.: 67). The empirical evidence Okin presents plausibly explains the cultural pressures women face and girls' low confidence in choosing professions (2016: 65). However, as section three shows, this cannot show why each individual woman makes her choices. Okin nevertheless seems to *assume* that, once freed from gendered constraints and discrimination, women as a group would *want to* make the exact same choices as men, with concerning implications. How do we move to Okin's *ideal* world, which she describes as genderless and perfectly equal?

The two Okins on women's choices

This section discusses how Okin's support of strict gender equality of outcome undermines the very feminist freedom of choice she defends. Such a prescriptive outcome is problematic no matter its implementation precisely because of Okin's diagnosis of *female false consciousness* I have established. Okin champions Amy Gutman's proposal of gender quotas for positions in primary schools to mostly equalise roles (Okin, 2016:

81). Taking this idea much further, as discussed above, in Okin's desired future 'men and women [participate] in more or less equal numbers in every sphere of life, from infant care to different kinds of paid work' (Ibid.: 79). This quote is crucial for my critique of Okin and, as I will show, exemplifies why her assumption of false consciousness is so problematic. Following Okin, every occupation and role within these occupations, as well as the structuring of families, would therefore be not just without gendered expectations, but follow relatively strict *equality of outcome*. This is no longer a future of choice, but one of very morally demanding quotas, no matter how they are implemented. Considering that Okin presents an unresolved debate of how far gender is due to nature or nurture (Ibid.: 80), it is troublesome to ascertain through coercive policy whether 50% of a nation's women would want to work a certain profession, or whether the same number of women and men would want to do the same amount of childcare in a world without restrictive gender stereotyping. But even beyond this, Okin's ideal outcome would require a heavily restricted world on the individual level, where career choices are guided by gender as a qualification. Okin simply assumes that in our non-ideal world a high number of women and girls are *only* making these choices because of gendered constraints and that in an ideal world they would choose otherwise. However, the policy consequences of Okin's counterfactual assumption are demanding and intrusive. Section one explained how Okin denies requiring much direct state intervention in the family, and she underscores how family justice can be achieved through *non-intrusive* policy (2005: 246). Okin creates the illusion of a state who does not police gender in intrusive ways but still endorses a false consciousness view in an idealised world without gender oppression. But what could be more intrusive than for the government to change somebody's *mind*? Okin insists on holding on to Rawlsian liberalism (Ibid.: 240) even when her normative outcomes seem to be so irreconcilably opposed to the freedom of choice inherent to both liberalism and her own liberal arguments.

Okin's assumption that taking gendered oppression away would lead women and men to make the exact same choices explains her gender abolitionism. Okin says '[a] just future would be one without gender' (2016: 79) and argues that schools must 'promote the minimization of gender.' (Ibid.: 81), in line with her Rawlsian endorsement of schools as a space that reproduces the societal sense of justice (Okin, 2005: 243). However, Okin seems to conflate *gender* and *gendered oppression* and makes the controversial

assumption that oppression is gender's only aspect without justification. It is highly implausible that most women would agree with her, which is where false consciousness comes in. If our conception of womanhood is purely negative, which Wingrove calls the 'de Beauvoir dilemma' of Okinian false consciousness regarding Okin's earlier work (Wingrove, 2009: 63), Okin's practical proposals of equality of outcome make more sense. Surely many women's professional choices are indeed impacted or created by the power relations Okin describes.

However, we cannot assume that, in an ideal world, every woman, or even most women, would be in a different profession or living different childcare arrangements through choice, as Okin does. A wholly negative conception of womanhood is highly contentious and without backing it up and defending it against those who see gender as having positive *and* negative aspects, Okin opens herself up to the criticism that she wishes to strip those women who disagree with her of the exact thing she finds lacking in her diagnosis: freedom of choice. This creates an internal inconsistency as Okin insists on Rawlsian liberalism as the solution to female oppression (2005: 240). Furthermore, her egalitarianism has the potential to create a world of extreme state interventionism limiting women's life plans, founded on a *counterfactual empirical assumption* (women as a demographic group in a world freed from gendered constraints would desire the exact same things as men). Okin cannot, and does not try to, back this up. This further undermines the autonomy of those women Okin tries to defend, as Okin simply assumes their life and career choices to have been principally led by wider gender injustices through false consciousness.

We must not forget that there is a more liberal side to Okin. In her earlier writings she stresses the importance of allowing a pluralism of beliefs. Abolishing marriage, for example, would therefore disadvantage those who seek more traditional structures and gendered division of labour; although those traditional choices remain socially problematic (Okin, 1989: 180). It is possible that Okin's views simply changed over time to allow for

gender abolitionism (2016: 79) and restrictions on religious freedom (2005: 242). However, as Okin's later writings display equally concerning contradictions, it seems likely that the issue is simply that Okin wishes to hold on to the liberal tradition despite her radical views, which she cannot justify within the tradition. The crucial issue with Okin's

diagnostic focus on freedom, then, is that she herself wishes to constrain freedom. It is exactly Okin's assumptions of false consciousness which makes her diagnosis of gender constraining freedom so demanding and any resulting outcomes so intrusive and radical. She does not simply want to establish conditions of choice. Okin's arguments that a just future would be genderless make clear that her ideas around equality of outcome are not a feature of transitional justice but instead its final product. Okin does not justify that interventionist gender abolitionism is the only way to achieve this state.

Finally, I want to note the nuanced complexity of the topic I have discussed. Epistemic equality may be a real issue for *some* women, as they are not given all the information they need in order to make informed choices about their future, which leads to gendered oppression. Freedom of choice can only exist where women have all the facts before they make their choices. Furthermore, it is crucial for us to combat any situation where women are either coerced or make career choices based on societal pressures and pre-existing inequalities. However, assuming traditional gender choices are *always* a case of false consciousness will inevitably lead to very strict and intrusive policy. More importantly, Okin restricts the very value she advocates: freedom of choice for women.

Conclusion

To conclude, my article first explained how the liberal of the two Okins focuses on freedom of choice, which was later shown to be contradictory. I then moved into a critique of Okin's diagnosis through sections two, on false consciousness, and three, on how Okin undermines women's freedom of choice. To conclude, Okin's diagnosis and its focus on freedom of choice are flawed because of her assumption of women's false consciousness, because Okin uses it to inconsistently promote further restrictions on women's freedom of choice, and because Okin does not defend her call for a genderless society.

References

Okin, Susan Moller (1989) *Justice, Gender, and the Family*, New York, NY: Basic Books.
Okin, Susan Moller (1994) 'Gender Inequality and Cultural Differences', *Political Theory*, 22 (1): 5-24.

Okin, Susan Moller (2005) ''Forty acres and a mule' for women: Rawls and feminism', *Politics, Philosophy & Economics*, 4 (2): 233-248.

Okin, Susan Moller (2016) 'Justice, Gender, and the Family', in Engster, Daniel and Metz, Tamara (eds) *Justice, Politics, and the Family*, New York, NY: Routledge, pp. 63- 87.
Widdows, Heather (2022) 'No Duty to Resist: Why Individual Resistance Is an Ineffective Response to Dominant Beauty Ideals', *Proceedings of the Aristotelian Society*, 122 (1): 27–46.

Wingrove, Elizabeth (2009) 'Of Linchpins and Other Interpretive Tools: Reconsidering Okin's Method', in: Satz, Debra and Reich, Rob (eds) *Toward a Humanist Justice: The Political Philosophy of Susan Moller Okin*, Oxford: Oxford University Press, pp. 55-66.

Authors Bio

Antonia Milla is an MSc Political Theory student at LSE and recent BA Politics and French graduate from Bristol. Her research focuses primarily on democratic theory and migrants' rights to political participation. This also includes the points at which the two sides of her undergraduate degree intersected, such as multiculturalism, language rights, and the European Union. More recent interests also include democratic rights in the context of climate change and bioethics, and rights to cultural spaces. In her free time, she enjoys all kinds of dance, playing the ukulele, and spending as much time with her favourite dogs as possible.

More than just ‘bad manners’ and dropping ‘f-bombs’: The pharmakon approach and Duterte’s ambivalent populist politics classified as a family resemblance

Author: Stefanija Kovacevic

ORCID: 0009-0000-8942-2004

Programme: BSc Politics and International Relations

Affiliation: School of Sociology, Politics and International Studies

Email: stefanija.kovacevic02@gmail.com

Abstract

This article addresses the existing literature’s gap that overlooks classifying Duterte’s populism as simultaneously style, strategy, and ideology, whilst under-emphasising that his populism may be in fact fascism. The article directly challenges Duterte’s populism as ostensibly ‘bad manners’, proposing a family resemblance approach to capture its uniqueness. A family resemblance classification is supported by contextualising Duterte’s populism as a pharmakon and his ‘curing’ and ‘poisoning’ of democracy, strengthening an ‘ambivalent’ characteristic of his populist politics, reinforcing the requirement for a variety of populism approaches. With the pharmakon notion justifying the family resemblance, this enables classifying Duterte’s populism to a greater extent as style, which also accounts for degrees of his fascism. Literature on Duterte’s populism between 2016 and 2022 is analysed, thematically organised in relation to theoretical approaches of Moffitt and Ostiguy; Weyland and Roberts; and Rooduijn and Mudde. The article contributes to populism literature by offering a flexible conceptual framework for understanding populist leaders, whilst further situating East Asian politics within populism studies. Through Duterte’s case, future assessments of populist cases should attempt to incorporate parallel discussions on democracy, developing the theoretical basis of populism as an ambivalent response to the infrastructure of democracy.

Keywords: Populism, Duterte, Family Resemblance, Pharmakon

Introduction

Classifying Duterte's populism as 'bad manners' limits understanding his dynamic populist politics. Subsequently, this article adopts a family resemblance approach which captures the diverse empirical evidence of his populism across overlapping dimensions of style, strategy, and ideology. The article justifies this approach by contextualising Duterte's populism as a pharmakon, where evidence of both 'curing' and 'threatening' democracy reinforces the requirement of multiple populism approaches to acknowledge the theme of 'ambivalence', which singular approaches, such as 'bad manners', overlook. Combining these frameworks enables understanding Duterte's case of politics as populist style *emphasised* amongst other approaches (Tormey, 2018, p.270). Despite the importance of and initiating the article with recognising Duterte's populism as a pharmakon, the article acknowledges and classifies Duterte's populism as 'bad manners', where the performance of crisis is discussed. In discussing the crisis of the 'War on Drugs', the requirement of a family resemblance is introduced, leading to discussions of Duterte's populism as strategy and ideology. Finally, I discuss Duterte and fascism, which reinforces the importance of a pharmakon framework, and the validity of emphasising populist style. The family resemblance notion is reinforced throughout by provided evidence of Duterte's populism, relating to style, strategy, ideology, and fascism discussions, demonstrating a 'curing' and 'poisoning' of democracy. Overall, I conclude that by highlighting Duterte's populism's ambivalence through the pharmakon and the family resemblance, populism can become more theoretically distinguishable within political studies.

Duterte's populism as a pharmakon means a requirement for a family resemblance approach

Family resemblance is complex and overlapping networks of smaller collectivities, which together form multitudes that encapsulate empirical dynamism, that being the uniqueness of concepts like populism (Wittgenstein cited in Gakis, 2025, pp. 184 – 195; Binev, 2023, p.525). I argue that understanding Duterte's populism as a pharmakon derives, firstly, from his populism's emphasis in crisis. Subsequently, employing a pharmakon theoretical lens contextualises and strengthens the inherent multidimensionality of his populist politics, supporting the essence of the family

resemblance approach's characteristic of uniqueness. Since his politics is ambivalent regarding 'curing' or 'poisoning' democracy, this establishes how 'bad manners' is insufficient, solely capturing an aspect of his threatening capacity; where other approaches must be employed to develop the varying facets of 'threat' as well as 'cure'.

Populism as a pharmakon is understood as populism being a symptom of crisis that is both a poison and a cure to democracy (Tormey, 2018, p.260). Seeing populism as bad and a 'threat' to democracy is equally as overblown as embracing it as a deepener; where ambivalence must be embraced (Tormey, 2018, pp. 260 - 261). This problematises a single approach to categorising populists, where style, strategy, and ideology deem populism as threatening, whilst Laclau's discourse approach finds positive democratic influences (Tormey, 2018, p.260); singular approach definitions and classifications eroding the essence of populism and its ambiguity.In Duterte's case, specifying an approach is certainly neither viable nor useful, exemplified by both his threatening *and* deepening of democracy (Timberman, 2016, p. 143; Mietzner, 2019, pp. 371 -378) – an ambivalence that I reinforce throughout the article regarding each populist approach and Duterte's politics. Importantly, a pharmakon contextualisation grants the notion of populism as one kind of politics of one approach emphasised amongst other approaches (Tormey, 2018, p.270); where, although Duterte's populism is principally a family resemblance, populist style *can* be emphasised. Overall, the article's ambivalent positionality towards categorising Duterte's case, supports and elaborates Urbinati and how populism displays family resemblances, specifically to oppositive political systems of democracy and fascism (2019, p. 115).

Justifying Duterte's performance of 'bad manners'

'Bad manners' is contextualised as a component within populist style, which is a symbolically mediated performance made to audiences, to create and navigate fields of power comprising the political, including features of appealing to 'the people' versus 'the elite', and the construction of crisis (Moffitt, 2016, p.28). The fact that the leader conducts performative aspects renders style partially synonymous with the discourse of rhetoric (Moffitt, 2016, pp.40 - 41). However, it maintains its distinguishability,

emphasising ‘performative’ features through its relation to and focus on performances within heavily mediatised contemporary politics (Moffitt, 2016, pp. 28 – 39; Manucci, 2017, p.469).

Duterte’s performance of ‘bad manners’

Populism’s ‘bad manners’ builds upon the appeal to the people, where populists coarsen political rhetoric to highlight urgent matters to the ‘people’ (Moffitt, 2016, p.44). Moffitt collaborates with Ostiguy, defining ‘bad manners’ as the flaunting of the ‘low’, which is an antagonistic performance by the leader to establish a socio-cultural appeal, often flaunting body language and gestures deeply linked with a society’s history (2016, p.44; 2017, p.74). The identified ‘low’ is relationally and culturally understood in relation to the ‘high’ of a given society, often positioned and interpreted in relation to the left-right spectrum (Ostiguy, 2017, pp.86 - 87).The socio-cultural component of the ‘low’ is the coarse language, whilst the political-cultural component is leaders’ performances being valued for ‘ballsyness’ (Ostiguy, 2017, pp. 79 – 83), where successful appeals often require highlighting the leader’s machismo (Moffitt, 2016, p.66) – crucial to Duterte. The leader within this approach often *performs* ‘bad manners’, identified by characteristics of maintaining both ordinariness, seeming as one of the people, and extraordinariness, seeming above them (Moffitt, 2016, p. 52).

Iagree that Duterte’s populism can be classified as the performance of ‘bad manners’, where Curato and Encinas-Franco confirm Duterte’s ‘low’ behaviour as anger, the socio-cultural dimension, and hypermasculinity, the political-cultural dimension (2017, pp. 142 – 145; 2022, p.107). Anger can be exemplified by Duterte using the ‘f-bomb’ in key speeches, whilst calling Barack Obama, Pope Francis and Ban Ki-moon ‘son[s] of a whore’ (Curato, 2018, p. 209). This is in addition to coarse language and stating that ‘[he] will eat [drug addicts] alive’ (PCOO cited in Encinas-Franco, 2022, p. 115). Duterte’s machismo is more prevalent (McCargo, 2016, p.188; Guevarra and Arcilla, 2020, p.491), exemplified by two accounts of him objectifying women. The first, admitting to ogling the vice president’s legs at a cabinet meeting, and the second, kissing a Filipino migrant woman at an overseas meeting (Curato, 2018, p.209; Encinas-Franco, 2022, p.107). This confirms ‘bad manners’ in the performance of extraordinariness through sexuality (Encinas-Franco, 2022, p.118; Moffitt, 2016, p. 52).

Subsequently, to a greater degree the classification of Duterte's populism as 'bad manners' and its performative aspect, is underscored by new media complementing populist performances. For example, the Philippines has one of the highest rates of social-media use in the world, where politics inevitably unravels online (Dressel and Bonoan, 2019, p.143). Understanding this context *and* the global relation to the political valorising more the spectacular, 'bad manners' more readily and successfully establishes as an appeal to the 'people' (Moffitt, 2016, p.39) in the Philippines, reinforcing Duterte's ability to perform extraordinariness through sexuality. Overall, the performance of 'bad manners' classifies Duterte's populism, exemplified through coarse language, machismo, and the mediatised context.

Duterte's populism as largely populist style: the performance of crisis and critique

Performance of crisis is an additional feature of style, alongside 'bad manners', that must be mediated, injecting urgency into populists' messages (Moffitt, 2016, pp. 45, 113 – 120), which can also classify Duterte's populism, exemplified by the 'War on Drugs' crisis. The crisis was an important aspect of Duterte's populism, where he declared to 'save' the nation from the disaster of illegal drugs (Curato, 2017, pp. 148 – 149). Curato supports that the crisis was performed (2017, pp. 148 – 149), which is verified by the pre-established Philippine media use, which indicates a privileging performance and mediatisation. Moreover, the crisis additionally demonstrates the combination of populist style features of 'low' behaviour of coarsened language within the performance of crisis, challenging the assumption that 'bad manners' is insufficient for encapsulating Duterte's populism.

Nevertheless, I problematise populist style's applicability to the Philippine context regarding the left-right spectrum and 'the low' (Ostiguy, 2017, p.73). The Philippines is defined as a cacique democracy, contextualising Duterte's style as neither left nor right-wing (Capuno, 2020, p.265; Curato, 2020, 6:43 – 7:50). Since the Philippines cannot account for left-right terminology, I problematise how the 'low' can be identified, since the spaces of the orthogonal low-high and left-right axis combined determine the political opposite of populism in politics: the high (Ostiguy, 2017, pp.86 – 87). This challenges Moffitt's conclusion that populist style is applicable to a variety of cases -

a result of overlooking Asia-Pacific cases (2016, pp.41 - 45).However, Moffitt *does* outline that the 'low' breaks historical society norms of a relevant context (2016, p.44), therefore accounting for cases beyond the left-right spectrum, where culture is prioritised (Ostiguy, 2017, pp.86 – 87). Duterte's low behaviour, of coarse language encapsulating machismo, exemplified this by breaking the 'high' culture norm of politeness, upheld by old family alliances of great wealth, where even asking 'who are you' (sin ka ba) is considered rude (Wurfel, 1979, p.233; Agpalo, 1975, p.11; Pinches, 1997, p.107; Bernard, 1973, p.363). This limitation of specifying the historical society norm that the populist challenges question the justification of a family resemblance approach, partially eroding the validity of style interpretations. Subsequently, understandings of Duterte's 'low' behaviour lose credibility in the inability to locate the 'high' that is being challenged, which further questions in what ways and whether democracy was at all deepened or threatened, weakening the pharmakon justification. However, to a greater extent, populist style sufficiently encapsulates Duterte's populism's performance of 'bad manners' and crisis, accounting also for left-right axis challenges. The highly mediatised Philippine context ultimately emphasises Duterte's populist style, where spectacular performance is contextually privileged, reinforced in subsequent sections that support Duterte's populism as a pharmakon. Importantly, style demonstrates Duterte's populism as a pharmakon, establishing the ambivalence argument, Curato confirming that Duterte's machismo presents an attack against democratic institutions, having had destabilising impacts for democracy (2019, p.261). However, Duterte's 'bad manners' can also be perceived as a cure to democracy in shaping the government to be more representative (Lijphart, 2012, pp. 1-2), where his swearing and misogynistic events have channelled the true voice of the people (Montiel, Uyheng and de Leon, 2022, pp. 437 – 438; Bernard, 1974). Czech verifies this, where Duterte's 'bad manners' were deemed an authentic expression of the people's resentment and anger towards the establishment (2022, p.212).

Duterte's populism as strategy

Moreover, Duterte's 'War on Drugs' crisis can be classified as populist strategy, where mass followers were mobilised to undermine the elite who were enabling the drug crisis. Populism as a political strategy is when a personalistic or charismatic leader seeks government power based on large numbers of unorganised followers, vying to win or exercise government power, ensured by the constant mobilisation of the people

(Weyland, 2017, pp.50- 57).Social communications media are increasingly used to reach these masses of the ‘people’ (Weyland, 2017, p.57), where failure of representative institutions to mediate between citizens and the state facilitates populism as a strategy (Roberts, 1995, p.113).

The failure of representative institutions facilitating strategy were evident; where the failure of liberal democratic promises was confirmed as the reason for Duterte’s rise (Thompson cited in Kusaka, 2017, p.51). The adoption of followers, based on this failure to mediate between citizens and the state, was evidenced by Filipinos voting for Duterte as a courageous leader capable of tackling ineffective political leadership (Timberman, 2016, p.135). Itao elaborates that the crisis ensured dedicated Filipino supporters, who identified Duterte as a supporter of democracy, where ‘constant’ mobilisation was demonstrated by Duterte’s consistent popularity with Filipinos trusting his leadership (cited in Kreuzer, 2020, p.18; Maboloc cited in Kreuzer, 2020, p.19). Crucially, Curato highlights the instrumentalisation of crisis for strategic purposes, where in her ethnographic study, before the ‘War on Drugs’, Filipinos did not view the ‘drug problem’ as a pressing national issue (Curato cited in Cielo, 2022, p.273). Alongside the crisis, Duterte demonstrated reaching and mobilising masses of followers through media (Weyland, 2017, p.57). For example, young people volunteered to ‘like’ Facebook posts favourable to Duterte, whilst the ‘Dutertards’ tag – short for ‘Duterte’s retards’ – was deployed across social media, garnering record-breaking political rallies composed of disparate groups of citizens (Curato, 2018, pp.210 - 212, 2017, p.148).

Overall, Duterte’s populism can be classified as a strategy, demonstrating mass followers being led by a charismatic leader against the ‘elite’. However, Duterte’s mass mobilisation, especially through the social media context, could also be mistaken for fans admiring a celebrity, relating to populist style and the performance of extraordinariness (Moffitt, 2016, p.85) – a duality encapsulated by the family resemblance approach. A family resemblance approach is strengthened by viewing evidence of Duterte’s strategy as a pharmakon. Kreuzer exemplifies the ‘cure’ dimension, where everyday Filipinos believed that Duterte could deepen democracy for good through institutional changes that challenged the elite , which could enable greater participation in decision-making and social equality for the people (Quimpo,

2005, p.230). Rubio-Perez and Tuaño (2025) verify this, where a survey conducted by Jesuit schools outlined that 62% of Filipinos believed that Duterte was good for democracy in challenging institutions and upholding transparency and accountability. Fundamentally, Duterte's attempted changes of institutional checks in challenging the elite of the 'War on Drugs' crisis 'cured' democracy by making the government more representative and responsive to the interests of the people (Lijphart, 2012, pp. 1-2). However, Duterte's response to the institutional failings of the elite can also demonstrate the 'poison', eroding the democracy's constitutional checks on institutional power, which are important for democratic stability(Akkerman, 2003, pp.156 – 157).

Duterte's populism and the lowest common denominator approach

Drawing on Rooduijn and Mudde (2014; Mudde and Rovira Kaltwasser, 2017), 'The War on Drugs' can be classified as ideology, where the crisis constructed an 'us' versus 'them' dichotomy, between the 'pure' people, and the drug addicts *and* the 'elite' (Curato, 2016, p.94). Rooduijn characterises populism as the emphasis of the people, the criticism of the elite, and populists proclaiming a serious crisis (2014, p.572). Despite Rooduijn's analysis not focusing on Asian cases, Duterte is still incorporated in proclaiming a serious crisis. Furthermore, Mudde focuses the 'us' and 'them' between the 'people' and the 'elite'. The 'people' of the crisis relates to Mudde's 'common people', who vindicate the dignity of groups, who are being objectively excluded from power due to their sociocultural status – in the Filipino case drug addicts - whilst the 'elite' are those working against the interests of the country – the Filipino elite protecting the addicts (Mudde and Rovira Kaltwasser, 2017, pp. 7 – 13).

Curato demonstrates the 'us' of the 'pure' people of Duterte's populism through the concept of 'penal populism', where people engaging in the politics of anxiety and hope constructed drug users as 'dangerous others' (cited in Kusaka, 2017, p.51), confirming 'pureness' through the community cohesion of this other (Curato, 2016, p.100). Linking to Rooduijn, the people were emphasised by Duterte justifying the anti-drug campaign as a matter of life-or-death for everyday Filipinos (Kine, 2017, p.25). Both Capuno and Muhs reflect how the 'elite', those opposed to the 'pure people', were identified as the

‘them’ by Duterte by being criticised as the enablers of the drug crisis (2020, p.265; 2022, p.154).

Duterte’s populism can be classified as ideology - a lowest common denominator approach. However, the construction of the dichotomy between the ‘people’ and ‘elite’ can also be interpreted as style, where ‘pureness’ can be defined as a feeling of moral virtuosity (Ostiguy, 2017, p.91). Hence, a family resemblance approach is most adequate to capture all the facets of Duterte’s populism. Duterte’s populism as an ideology supports the pharmakon, where the everyday ‘pure’ Filipino ‘people’ believed demonising drug addicts was for the better of democracy (Curato, 2016, p.100). The pharmakon notion is further elaborated, where the elite were understood as supporting the West and liberal ideologies in perpetuating the crisis, Duterte’s populism and the construction of the ‘them’ deepened democracy by challenging the colonial history of ‘Imperial Manila’ derived from US imperialism (Pernia, 2022, p.77; Quimpo, 2005, p.235). However, these challenges to ‘Imperial Manila’ present the alternative side of the pharmakon in eroding the essence of the democratic state.

Questioning Duterte’s populism as fascism

Interestingly, literature on Duterte’s populism explores the argument that he is a fascist, instead of a populist. Bello defines fascism as a charismatic individual, deriving strength from a heated multi-mass base, and proposing a political project that contradicts democratic values (2017, p.21). Discussions on strategy demonstrate Duterte’s charisma and his multi-mass base, as well as how he threatens democracy. Meanwhile, a 2020 rally opposed specifically Duterte’s fascist measures (Bello, 2017, p. 22; Guevarra and Arcilla, 2020, p.489). However, the strategy approach distinguishes populism as fully personalistic, where supporters follow the leader’s whims, whilst fascism establishes full-scale despotism (Weyland, 2017, p.48), as demonstrated by the ‘Dutertards’. Populist style also outlines that fascism governs in a disciplined manner from the state down and ends elections once it wins them (Ostiguy, 2017, p.83). Full-scale despotism was not evident, where during Duterte’s administration, the Philippines continued to hold elections in 2019, where candidates from parties opposed to him in fact failed to win any of the Senate seats up for election (Dressel and Bonoan, 2019, p.134).

Questioning Duterte's degree of fascism re-introduces the emphasis of populist style. Style does not view populism as a binary and acknowledges the degree of populism that is employed across contexts, where some actors are not as populist as we think (Gidron and Bonikowksi cited in Moffitt, 2016, p.46). Thus, populist style can account for Duterte's degree of fascism, compared to strategy and ideology, rationalising earlier arguments of Duterte's populism as largely populist style, whilst confirming a family resemblance of fascism being amongst strategy, ideology, and style approaches. Fascist arguments further demonstrate how Duterte's populism can be interpreted as a threat to democracy, and so a pharmakon, reinforcing the ambivalence of his populist politics, and so justifying a family resemblance.

Conclusion

In conclusion, I agree that Duterte's populism can be classified as the performance of 'bad manners', however, this is not sufficient. A family resemblance approach *is* sufficient for encapsulating Duterte's populism's empirical dynamism and overlapping networks of approaches, particularly emphasising populist style. This has been enabled by understanding Duterte's populism as a pharmakon, where the approaches demonstrate and so support his populism's threat to and deepening of democracy; reflecting the ambivalent characteristic. Importantly, populist style is emphasised in underpinning Duterte's 'bad manners', his performance of crisis, the followers of strategy ,constructing the 'pure' of the people as an ideology, and accounting for fascist elements by understanding populism in degrees. Most convincingly, I conclude that understanding Duterte's populism as a pharmakon enables a classification of a family resemblance, recognising his unique populist brand and its various facets. Employing this broad approach aids the distinguishing of populism from other political phenomena, where singular approaches restrict analytical depth for populist cases and populism studies (Tormey, 2018, p.260).

References

Agpalo, R. E. (1975) 'Legitimacy and the Political Elite in the Philippines', *Philippine Political Science Journal*, 2(2), pp. 1–16. Available at: doi: 10.1080/01154451.1975.9753907.

Akkerman, T. (2003) 'Populism and Democracy: Challenge or Pathology?', *Acta Politica International Journal of Political Science,* 38, pp. 147–159. Available at: https://doi.org/10.1057/palgrave.ap.5500021

Bello, W. (2017) 'The Spider Spins His Web: Rodrigo Duterte's Ascent to Power', *Philippine Sociological Review*, 65, pp. 19–47. Available at: http://www.jstor.org/stable/45014308 (Accessed: 2 December 2025).

Bernard, M. A. (1974) 'Philippine Culture and Social Values', *Philippine Studies*, 22(3/4), pp. 363–373. Available at: http://www.jstor.org/stable/42634879

Binev, B.S. (2023) 'Post-neoliberal Populism in Latin America and Eastern Europe: Recognizing Family Resemblance', *Studies in Comparative International Development*, 59(3), pp. 517–581. Available at: https://doi.org/10.1007/s12116-023-09384-1.

Capuno, J.J. (2020) 'Dutertenomics: Populism, Progress, and Prospects', *Asian Economic Policy Review*, 15(2), pp. 262–279. Available at: https://doi.org/10.1111/aepr.12296.

Cielo, C. (2022) 'Social Inequalities and Political Organization in the Philippines', in H.-J. Burchardt and J. Leinius (eds), *(Post-)colonial Archipelagos: Comparing the Legacies of Spanish Colonialism in Cuba, Puerto Rico, and the Philippines*, University of Michigan Press, pp. 259 – 282.

Curato, N. (2016) 'Politics of Anxiety, Politics of Hope: Penal Populism and Duterte's Rise to Power', *Journal of Current Southeast Asian Affairs*, 35(3), pp. 91–109. Available at: https://doi.org/10.1177/186810341603500305

Curato, N. (2017) 'Flirting with Authoritarian Fantasies? Rodrigo Duterte and the New Terms of Philippine Populism', *Journal of Contemporary Asia*, 47(1), pp.142-153. Available at: DOI: 10.1080/00472336.2016.1239751

Curato, N. (2018) 'The Power and Limits of Populism in the Philippines', *Current History*, 117(800), pp. 209–214. Available at: https://www.jstor.org/stable/48614362 (Accessed: 2 December 2025)

Curato, N. (2019) 'TOXIC DEMOCRACY? THE PHILIPPINES IN 2018', *Southeast Asian Affairs*, pp. 261–274. Available at: https://www.jstor.org/stable/26939699 (Accessed: 11 December 2025).

Curato, N. (2020) *8. Nicole Curato on Duterte and Populism in the Philippines* [Podcast]. 20 July. Available at: https://podcasts.apple.com/us/podcast/8-

nicole-curato-on-duterte-and-populism-in-the-philippines/id1516388187?i=1000485525373 (Accessed: 3 December 2025).

Czech, F. (2022) ‘Between Legitimation and Support of Rodrigo Duterte. Understanding Enduring Popularity of the President of the Philippines’, *Polish Sociological Review*, (218), pp. 207–224. Available at: https://www.jstor.org/stable/27249688

Dressel, B. and Bonoan, C.R. (2019) ‘Southeast Asia’s Troubling Elections: Duterte Versus the Rule of Law’, *Journal of Democracy*, 30(4), pp. 134–148. Available at: https://muse.jhu.edu/pub/1/article/735447 (Accessed: 13 December 2025).

Encinas-Franco, J. (2022) ‘The presidential kiss: Duterte’s gendered populism, hypermasculinity, and Filipino migrants’, *NORMA*, 17(2), pp. 107-123. Available at: DOI: 10.1080/18902138.2022.2026107

Gakis, D. (2025) ‘Wittgenstein's “Family Resemblances” and Their Political Potential’, in L. Rasiński, A. Biletzki, L. Koczanowicz, A. Pichler and T. Wallgren (eds), *Wittgenstein and Democratic Politics Language, Dialogue and Political Forms of Life,* New York, NY: Routledge, pp. 182 – 196.

Guevarra, A. R., and Arcilla, M. (2020) ‘The Source of Actual Terror: The Philippine Macho-Fascist Duterte’, *Feminist Studies*, 46(2), pp. 489–494. Available at: https://doi.org/10.15767/feministstudies.46.2.0489

Jayasuriya, K. (2020) ‘The Rise of the Right: Populism and Authoritarianism in Southeast Asian Politics’, *Southeast Asian Affairs 2020,* (1), pp. 43-55. Available at: https://muse.jhu.edu/article/754739 (Accessed: 12 December 2025).

Kine, P. (2017) ‘Philippine President Rodrigo Duterte’s “War on Drugs.”’, *Harvard International Review*, 38(3), pp. 24–27. https://www.jstor.org/stable/26528678 (Accessed: 2 December 2025).

Kreuzer, P. (2020) *3. Support For the Patron-Strongman Who Delivers*. Available at: https://www.jstor.org/stable/resrep26192.6?seq=2 (Accessed: 7 December 2025).

Kusaka, W. (2017) ‘Bandit Grabbed the State: Duterte’s Moral Politics’, *Philippine Sociological Review*, 65, pp. 49–75. Available at: http://www.jstor.org/stable/45014309 (Accessed: 2 December 2025).

Lijphart, A. (2012) *Patterns of democracy: government forms and performance in thirty-six countries*. 2nd edn. New Haven: Yale University Press.

Manucci, L. (2017), ‘Populism and the Media’, in C. Rovira Kaltwasser, P. Taggart, P. Ochoa Espejo and P. Ostiguy (eds), *The Oxford Handbook of Populism*, Oxford: Oxford University Press, pp. 73 – 98.

McCargo, D. (2016) ‘Duterte’s Mediated Populism’, *Contemporary Southeast Asia*, 38(2), pp. 185–190. Available at: http://www.jstor.org/stable/24916626 (Accessed: 2 December 2025).

Mietzner, M. (2019) ‘Movement leaders, oligarchs, technocrats and autocratic mavericks’, in C.d.l. Torre (ed.) *Routledge Handbook of Global Populism.* Milton Park, Abingdon, Oxon: Routledge, Taylor & Francis Group, pp. 370 – 381.

Moffitt, B. (2016) *The global rise of populism: performance, political style, and representation*. Stanford, California: Stanford University Press.

Montiel, C. J., Uyheng, J., and de Leon, N. (2022) ‘Presidential Profanity in Duterte’s Philippines: How Swearing Discursively Constructs a Populist Regime’, *Journal of Language and Social Psychology*, 41(4), pp. 428-449. Available at: https://doi.org/10.1177/0261927X21106578

Mudde, C. and Rovira Kaltwasser, C. (2017) *Populism: a very short introduction*. New York, NY: Oxford University Press.

Muhs, L. (2022) ‘Duterte's "War on Drugs" Rhetoric: Consolidating Power through Penal Populism’, *Journal for the Study of Radicalism*, 16(2), pp. 149-164. Available at: https://muse.jhu.edu/article/869477. (Accessed: 2 December 2025).

Ostiguy, P. (2017), ‘Populism: A Socio-Cultural Approach’, in C. Rovira Kaltwasser, P. Taggart, P. Ochoa Espejo and P. Ostiguy (eds), *The Oxford Handbook of Populism*, Oxford: Oxford University Press, pp. 73 – 98.

Pernia, R. A. (2022) ‘Beyond Duterte: Conceptualizing a Research Agenda of Populism Research in the Philippines’, *Philippine Political Science Journal* , 43 (1), pp. 57-90, Available at: https://doi.org/10.1163/2165025X-bja10028

Pinches, M. (1997) ‘Elite democracy, development and people power: contending ideologies and changing practices in Philippine politics’, *Asian Studies Review*, 21 (2-3), pp.104-120, Available at: DOI: 10.1080/03147539708713166.

Quimpo, N. G. (2005) ‘Oligarchic patrimonialism, bossism, electoral clientelism, and contested democracy in the Philippines’. Review of *ThePolitics ofBanking inthe Philippines;Capital, Coercion, and Crime:Bossism inthe Philippines;Elections and Democratization in thePhilippines; TheCommunist Party ofthe Philippines 1968-1993:AStoryof Its Theoryand Practice* , by P. Hutchcroft, J. Sidel, J. Franco, and K. Weekley, 37 (2), pp. 229 – 250. Available at: https://doi.org/10.2307/20072884

Roberts, K. M. (1995) ‘Neoliberalism and the Transformation of Populism in Latin America: The Peruvian Case’, *World Politics*, 48(1), pp. 82–116. Available at: http://www.jstor.org/stable/25053953 (Accessed: 2 December 2025).

Rooduijn, M. (2014) ‘The Nucleus of Populism: In Search of the Lowest Common Denominator’, *Government and Opposition*, 49(4), pp. 572–598. Available at: https://www.jstor.org/stable/26350350. (Accessed: 2 December 2025)

Rubio-Perez, B. and Tuaño, R. (2025) *Why Filipinos think Marcos Sr. and Duterte were “good’ for democracy* [Podcast]. 23 May. Available at: https://www.youtube.com/watch?v=ZRteJ5TjZVM (Accessed: 24 May 2026).

Timberman, D.G. (2016) ‘The Vote in the Philippines: Elite Democracy Disrupted?’, *Journal of Democracy*, 27(4), pp. 135-144. Available at: https://dx.doi.org/10.1353/jod.2016.0069.

Tormey, D. (2018) ‘Populism: democracy's *Pharmakon?’*, *Policy Studies*, 39(3), pp. 260-273, Available at: DOI: 10.1080/01442872.2018.1475638.

Urbinati, N. (2019) ‘Political Theory of Populism’, *Annual Review Political Science*. 22, pp. 111-127. Available at: https://doi.org/10.1146/annurev-polisci-050317-070753

Weyland, K. (2017) ‘Populism: A Political-Strategic Approach’, in C. Rovira Kaltwasser, P.A. Taggart, P.O. Espejo, P. Ostiguy (eds), *The Oxford Handbook of Populism,* Oxford, United Kingdom: Oxford University Press, pp. 48 – 72.

Wurfel, D. (1979) ‘Elites of wealth and elites of power. The changing dynamic: A Philippine Case Study’, *Southeast Asian Affairs*, pp. 233–245. Available at: http://www.jstor.org/stable/27908379.

Authors Bio

My name is Stefanija Kovacevic, and I am a 2026 graduate from the University of Bristol’s School of Sociology, Politics, and International Studies. Over the course of my degree, I have developed a research interest in the Global South, where I am fascinated in analysing this group of countries’ role in international politics, alongside the degree of influence they can exercise. This has led to a specific interest in regional blocs of the African Union and ASEAN. Beyond academics, I am devoted to my chess playing, as well as language learning, where this year I hope to include Arabic.

Slipping Through the Nets of Language: Character and Circulation in *Jacob's Room*

Author: Lily Prole

ORCID: 0009-0003-4540-8309

Programme: MAEnglish Literature

Affiliation: School of Humanities

Email: lilyprole@gmail.com

Abstract:

This paper examines the problem of representing character in *Jacob's Room* (1922), arguing that Virginia Woolf develops a narrative method that resists the containment of identity within language. While criticism has emphasised the novel's fragmentation and the displacement of Jacob's voice, often reading these as formal absence or narrative loss, less attention has been paid to how these strategies operate as a sustained critique of classificatory thinking. This paper argues that these strategies underpin a method of 'netting', which resists fixed characterisations, assembling only partial impressions of character. Drawing on Woolf's life writings and close analysis of taxonomic imagery, this paper demonstrates how acts of classification, whether scientific or social, consistently fail to stabilise Jacob's identity. These failures expose the limits of language as a system that seeks to fix meaning. Rather than abandoning representation, however, Woolf reconfigures it by privileging movement, repetition, and circularity over static description. In *Jacob's Room*, meaning is generated absence and fragmentation; the goal of capturing character is replaced with an ongoing process of perceptual return.

Key words: Virginia Woolf, Jacob's Room, Modernism, Classification, Character

‘It is no use trying to sum people up’, remarks the narrator of *Jacob’s Room* (1922).[1] ‘One must follow hints, not exactly what is said, nor yet entirely what is done’ (p. 123). In her 1924 essay ‘Mr Bennett and Mrs Brown’, Virginia Woolf writes that the essence of a person always slips beyond narrative grasp; it is ‘from the gleams and flashes of this flying spirit that [the writer] must create solid, living, flesh-and-blood Mrs Brown. Sadly he must allow that the lady still escapes him’.[2] In *Jacob’s Room* the attempted capture of the elusive titular protagonist appears futile, as Jacob escapes the act of being pinned down. By exploring the attempts to capture Jacob, this paper will argue that Woolf suggests that we cannot contain the essence of a person. Perhaps, then, ‘it is no use trying to sum people up’ (p.123). However, Woolf does not abandon the effort entirely, she presents a literary methodology of observing glimpses of a person’s essence. I term this process netting – akin to catching glimpses of insects – acknowledging that something will always slip through these nets of language.

I use the term ‘netting’ as it emerges from Woolf’s own reflections on her writing process. This is then mirrored in the lepidoptera imagery that pervades *Jacob’s Room*. Her diary reflects her longing to record the ‘myriad [of] impressions [she] net[s]’.[3] Yet, she acknowledges in a letter that, ‘sentences are only an approximation, a net one flings over some sea pearl which may vanish; and if one brings it up it won’t be anything like what it was when I saw it’.[4] The sea pearl, glimpsed beneath the surface, can never be recovered intact; once brought to the surface, into language, it loses something essential. Woolf, therefore, presents writing as both an act of retrieval and of inevitable loss, where the very attempt to fix perception entails its distortion. When discussing *Jacob’s Room*, critics have used terms like ‘capturing’ and ‘pinning’, however, ‘netting’ more precisely evokes this sense of inevitable failure. It is clear to the user of the net that there are holes in it, the captured specimen easily escapes and slips out the net. In the same way, Woolf’s use of language inevitably leads to escape of her subjects. She is aware of this failure and frustrated by it, writing in a 1927 essay,

---

[1] Virginia Woolf, *Jacob’s Room*, ed. by Urmila Seshagiri (Oxford University Press, 2022) p. 123. All further references are to this edition.
[2] Woolf, ‘Mr Bennet and Mrs Brown’, in *Selected Essays*, ed. by David Bradshaw (Oxford University Press, 2009) pp. 35-36.
[3] Woolf, *The Diary of Virginia Woolf: Vol. 3, 1925-1930*, ed. by Anne Oliver Bell and Andrew McNeillie (Hogarth, 1980) p. 6.
[4] Woolf, *The Letters of Virginia Woolf: Vol. 4, 1929-1931*, ed. by Nigel Nicolson and Joanne Trautmann (Hogarth, 1978), p. 223.

'life is always and inevitably much richer than we who try to express it'.[5] Therefore, the essence of a person, like Woolf's fictious Mrs Brown, can never be fully captured; we must 'content with a scrap of her dress or a wisp of her hair', as Woolf explains in her 1924 essay, 'Character in Fiction' .[6] The paper will conclude by demonstrating that Woolf's subjects spiral, return, and flicker in and out of perception, much like the moths and butterflies in *Jacob's Room*.

Characters attempt to classify Jacob, as if he were a specimen. The classification of Jacob by other voices reflects Suzanne Raitt's argument that the novel 'abandons the project of developing its protagonist's voice altogether, and instead experiments with the voices of others speaking in his place'.[7] Woolf foregrounds the instability of such classification through Mrs. Norman's speaking for Jacob, as she attempts to construct his identity through fragmentary details: his 'socks', 'tie', 'mouth', 'eyes', and reading habits (p. 22). Her observations move restlessly between interpretation and projection, repeatedly revising her assessment of Jacob, shifting from fear, imagining men as inherently 'dangerous', to admiration, perceiving him as 'nice, handsome, interesting, distinguished' (p. 22). These contradictory judgments expose how little of Jacob himself is grasped. Mrs. Norman catches only disconnected impressions which she stitches together into a narrative shaped by her own anxieties, desires, and maternal associations with 'her own boy' (p. 22). Woolf makes this explicit in this passage, writing that 'nobody sees any one as he is… They see a whole – they see all sorts of things – they see themselves' (p. 22), revealing that such acts of perception disclose more about the observer than the observed. The futile attempt to classify Jacob parallels his own engagement with taxonomy, when identifying an insect, he notes that, 'the upper wings of the moth which [he] held were undoubtedly marked with kidney-shaped spots of a fulvous hue' (p. 16). However, using language to classify insects is similarly inconclusive as Jacob consults a scientific text, only to find that it is incomplete: 'Morris called it "an extremely local insect found in damp or marshy places". But Morris is sometimes wrong. Sometimes Jacob, choosing a very fine pen, made a correction in the margin' (p. 17). Just as he isolates traits to

[5] Woolf, 'The Narrow Bridge of Art', in *Granite and Rainbow*, ed. by Leonard Woolf (Harcourt, 1958), pp. 11-23 (p. 23).
[6] Woolf, 'Character in Fiction', in *Selected Essays*, p. 37.
[7] Suzanne Raitt, 'Virginia Woolf's Early Novels: Finding a Voice', in *The Cambridge Companion to Virginia Woolf*, ed. by Susan Sellers (Cambridge University Press, 2010), pp. 29–48 (p. 31).

categorise insects, Mrs. Durrant attempts to define Jacob by isolating aspects of his appearance. However, like the attempted classification of insect, this project is inconclusive. Mrs Durrant struggles to place him, finding it 'difficult to say which seat in the opera house was his, stalls, gallery, or dress circle. A writer? He lacked self-consciousness. A painter?' (p. 55). However, Woolf suggests that such acts of classification are inherently flawed as they are inconclusive. Jacob is repeatedly refracted through the interpretive habits of others who only ever construct him through partial access, reinforcing Raitt's claim that the novel disperses classification by privileging a plurality of mediated perspectives over a stable protagonist. These moments show that Woolf repeatedly frustrates any stable attempt to classify Jacob, as scientific, social, and narrative descriptions all break down into uncertainty. Both taxonomy and social observation rely on selecting fragments and assuming coherence, but Woolf exposes this as an illusion shaped by the observer's projections.

The novel's refusal to pin Jacob down linguistically reflects the instability of language itself. Mrs Durrant's contradictory assessments of Jacob, calling him both "distinguished-looking" and "extremely awkward" (p. 55), highlight the limits of words, as each attempt to define Jacob offers only a fragment of him. There is an inherent failure at pinning words down to classify a person, which echoes the sentiments of Woolf's essay, 'Craftsmanship' (1937), where she describes words as 'the wildest, freest, most irresponsible, most unteachable of all things. Of course, you can catch them and sort them […] but words do not live in dictionaries'.[8] Likewise, individuals do not live within the constraints of language; they slip between definitions, resisting fixed meaning. Just as a pinned insect is no longer a living being, a person, once confined by words, is inevitably reduced to something incomplete, a sample specimen. This mirrors Rachel Sarsfield's argument that 'the question of how to "pin down" life in words without destroying it, to "pin" the moth without killing' is a 'perennial dilemma' for Woolf.[9] Woolf's awareness of the limits of language suggest that she does not seek to fully capture her subjects linguistically, she fears this destroys them. However, there are moments where this does not seem to be a dilemma but a sense that there is value in attempted capture. What emerges is a process of 'netting', a provisional mode of

[8] Woolf, 'Craftsmanship', in *SelectedEssays*, p. 89.
[9] Rachel Sarsfield, quoted by Christina Alt, *Virginia Woolf and the Study of Nature* (Cambridge University Press, 2010) p. 168.

apprehension in which language can only ever catch fleeting impressions, like insects briefly held before slipping through the holes of the net. Meaning is therefore never securelyretained but always escaping, so that representation becomes an encounter with lossas much as with retrieval. As Woolf herself writes in a 1922 letter, 'beauty […] isonly got by the failure to get it'.[10] Woolf resists closure, embracing the beauty foundinthe attempt, rather than the achievement.

Although complete capture of a person seemingly fails, Woolf reimagines the process, valuing movement over fixity. Christina Alt argues that Woolf shifts from a taxonomic approach to one more akin to ethology, writing that, for Woolf 'only fleeting glimpses of a moving subject were possible or, indeed, desirable'.[11] And so, Woolf conceives of 'writing as a process of release rather than capture'.[12] Whilst I agree that Woolf rejects capture, release implies a linear movement away from the subject. Her narrative method does not simply release its subjects but returns to them, in an ongoing circular engagement. Christine Froula conceptualises Jacob as 'an elusive being no net of words can capture', yet Woolf's novel enacts netting in a different way to Froula's conception: it is an ongoing attempt at catching glimpses rather than a final grasp of its subject. It is an acknowledgement that something will always slip through this net, drawing from her diary where writes of her desire to note down some of the 'myriad impressions [she] net[s]'.[13] Netting is not about possession but about attentiveness, a process that values fluidity over finality. Capture is not the goal; rather, by watching subjects spiral, we perceive their vitality.

Woolf's interest in fleeting moments over containment is further evident in her diaries as she recollects insect hunting as a child. However, Woolf herself describes a sense of detachment as she notes how 'the rest of us have rather departed from that profession'.[14] She deflates the seriousness of the hunt, mimicking military language, addressing the 'leader' and how they 'advance slowly.[15] She then describes the failed

---

[10] Woolf, *TheLettersofVirginiaWoolf:Vol.2, 1912-1922* , ed. by Nigel Nicolson and Joanne Trautmann (Hogarth, 1976) p. 599.
[11] Alt, p. 168.
[12] Alt, p. 179.
[13] Woolf, *TheDiaryofVirginiaWoolf:Vol.3, 1925-1930*, p. 6.
[14] Woolf, *APassionateApprentice:TheEarly Journals, 1897-1909* , ed. by Mitchell Alexander Leaska (Hogarth, 1990) p. 145.
[15] Woolf, *APassionateApprentice* , p. 144.

capture of a moth, however, Woolf does not express any disappointment at its escape. She focuses on the fleeting motion of the moth, writing that they 'gazed one moment on his splendour, & then uncorked the bottle. I think the whole procession felt some unprofessional regret when, with a last gleam of scarlet eye & scarlet wing, the grand old moth vanished'.[16] It is in the moth's escape that makes it so interesting for Woolf. A parallel moment occurs in the novel, when Jacob observes how 'the red underwing had circled round the light and flashed and gone. The red underwing had never come back' (p. 17). The moth does not simply appear and vanish: it circles, it flashes, and then it disappears, a sequence of movements that enacts its ephemeral nature. The repetition of 'red underwing' emphasises its presence even as it is lost, echoing Woolf's journal entry in its focus on motion rather than possession. Through both passages, Woolf underscores the value of observing creatures in motion, privileging the vitality of experience over the permanence of capture.

Woolf's fascination with circling glimpses permeates *Jacob's Room*, reflected in her repeated imagery of insects in motion:

> Back came the sun, dazzlingly. It fell like an eye upon the stirrups, and then suddenly and yet very gently rested upon the bed, upon the alarum clock, and upon the butterfly box stood open. The pale clouded yellows had pelted over the moor; they had zigzagged across the purple clover. The fritillaries flaunted along the hedgerows. (p. 17)

The butterflies are animated by dynamic verbs: as they 'pelted', 'zigzagged', 'flaunted', emphasising movement over stillness. The interplay of light and motion suggests an ephemeral, fragmented way of seeing, privileging glimpses over permanence. The open butterfly box symbolises failed containment, reinforcing the idea that life, like these insects, resists confinement. This is echoed in Jacob's own unsuccessful attempt at capture as he 'had seen a white admiral circling higher and higher round an oak tree, but he had never caught it' (p. 17). Woolf contrasts movement with capture, demonstrating that true vitality exists in motion, not in preservation. At Cambridge, white butterflies reoccur, circling again, 'in front of him two white butterflies circled higher and higher round the elm tree' (p. 27). Unlike earlier passages where insects were classified: 'Jacob caught the pale clouded yellows' (p. 16), these butterflies remain unnamed, marked only as white. By stripping the butterflies of taxonomic

[16] Woolf, *A Passionate Apprentice*, p. 145.

specificity, Woolf shifts the emphasis from classification to circular movement itself. The repeated image of the white butterflies itself enacts a kind of circling, formally mirroring the motion it describes. As Urmila Seshagiri argues, 'repetition only collapses meaning. Language, doubled and trebled, is as powerless to establish narrative direction'.[17] Moreover, repetition resists linear progression and instead produces a continual return. Meaning does not accumulate but circles back on itself, refusing resolution. The repetition does not clarify but leaves Jacob, and the reader, in an ongoing, ungraspable flux. Through the recurring imagery of insects circling, Woolf enacts a linguistic circling, suggestive of life's ungraspable essence. This repeated return also mirrors Woolf's broader process of netting, where perception continually reaches towards its object without ever fully securing it. Like the butterflies that circle beyond capture, Jacob's essence remains perpetually in motion, glimpsed only partially before slipping away again.

Diana Swanson cautions against over-interpreting nature in Woolf's work, arguing that, 'references to animate and inanimate parts of the nonhuman world are in excess of what would be sufficient for symbolic uses'.[18] However, in *Jacob's Room*, the circling motion of the natural world is also mirrored in human experience, thereby allowing for symbolic resonance. Like glimpsing at insects in flight, Jacob's presence is refracted through Clara Durrant's perception of the leaves as 'one green one, spread like the palm of a hand, circled down past Jacob's head' (p. 48). Jacob is not the focal point, he exists only in relation to the leaf's movement. Structurally, the sentence defers his presence, positioning him as secondary to nature. The leaf's circling motion mirrors Jacob's own elusiveness: he seems to hover just beyond full comprehension. As Raitt notes, 'most of the main events of the novel occur obliquely, just outside our range of vision'.[19] By placing Jacob on the periphery of perception, Woolf suggests that identity is not something fully graspable, but something we net aspects of, continually formed through movement. Woolf's fascination with fleeting moments is evidenced in a 1922 letter, where she writes that, 'no one can see [the human soul] whole […] The best of us catch a glimpse of a nose, a shoulder, something turning

[17] Seshagiri, Introduction, Woolf, *Jacob's Room*, pp. xi-xliii (p. xxxvi).
[18] Diana L. Swanson, "The Real World': Virginia Woolf and Ecofeminism', in *Virginia Woolf and the Natural World*, ed. by Kristin Czarnecki and Carrie Rohman (Clemson University Press, 2011), pp. 24–34 (p. 32).
[19] Raitt, p. 42.

away, always in movement'.[20] The idea that life can only be perceived in fragments is not a resignation but an aesthetic principle. She continues, 'it seems better to me to catch this glimpse, than […] make large oil paintings of fabulous fleshy monsters'.[21] Through this intertwining of natural movement and human perception, Woolf suggests that identity, like the circling leaf or the flight of insects, can only ever be netted in fleeting,partial glimpses rather than complete forms.

Later, Jacob's body again remains absent from direct perception as 'a man wheeled a trolley past Jacob's legs so near that he almost grazed them' (p. 104). Woolf presents Jacob's physical presence indirectly, emphasising what almost happens rather than what does. The focus shifts from Jacob to the movement around him. He exists not as a fully realised subject but as a peripheral figure, partially obscured by external actions. As Alt writes he 'remains an elusive presence, sighted rather than caught'.[22] Jacob is repeatedly only partially netted by perception, as though Woolf's narrative can merely brush against him rather than securely draw him into focus. Jacob's absence is further explored in the same passage, where Woolf writes: [U]nder the arc lamps in the Gare des Invalides, with one of those queer movements which are so slight yet so definite, which may wound or pass unnoticed but generally inflict a good deal of discomfort, Jinny and Cruttendon drew together; Jacob stood apart. They had to separate. (p. 104)

The arc lamps that illuminate the scene, positioning the characters as moths circling a flame, reinforcing the motif of fleeting movement. Woolf's description of 'queer movements', 'so slight yet so definite', evokes the delicate, imperceptible shifts of an insect adjusting its wings. The contrast between drawing together and standing apart suggests an unresolved cycle of attraction and separation, a kind of social and psychological circling. These encounters are only briefly netted before dispersing again, escaping stable emotional or narrative containment. This motif of circular movement extends to intellectual activity. At Cambridge, men write while 'their heads went round in a circle as the pen moved' (p. 32). This synchronised, repetitive motion evokes an insect-like quality, as if the act of thinking and writing is itself trapped in a cyclical pattern. Meanwhile, Jacob is described as 'wagging his head slowly from side to side; and then, taking out his penknife, he dug the point of it again and again into a

[20] Woolf, *TheLetters of Virginia Woolf* , p. 598.
[21] Woolf, *TheLetters of Virginia Woolf* , p. 598.
[22] Alt, p. 181.

knot in the table' (p. 33). His repetitive, restless gestures suggest a circling energy that refuses containment. Like an insect's erratic flight, he returns to the same point without ever settling. Unlike a pinned specimen, Jacob becomes an insect briefly slowed down by a net before slipping away.

Just as Jacob's restless gestures suggest an entrapment in repetitive patterns, the motif of movement extends beyond individual actions to shape the very structure of *Jacob's Room*. Woolf's repeated descriptions of Jacob's room mirror this tension between containment and motion. The room reappears at the end of the novel: 'listless is the air in an empty room, just swelling the curtain; the flowers in the jar shift. One fibre in the wicker arm-chair creaks, though no one sits there' (p. 142). The word 'listless' is ambiguous, it denotes stillness, yet it also implies a kind of restless suspension, as if the room is caught in a liminal state. Harvena Ritcher argues that Jacob is 'ironically "collected," like one of the specimens in his butterfly box, in Flanders Field'.[23] I argue that the room itself can be read as functioning as a site of failed containment. As Seshagiri writes, 'the room contains everything and nothing […] Jacob lives amidst the ruins of a narrative arc that can neither guide nor contain him'.[24] In her notes while drafting the novel, Woolf writes, 'let us suppose that the Room will hold it all together'.[25] Woolf's remarks seem both hopeful and ironic, suggesting that what remains is not a fixed essence of Jacob, but a space of lingering impressions. The room becomes a paradoxical space, both a container and a void, like Jacob's butterfly boxes, filled with remnants, yet emptied of certainty. The vibrations within the room, such as the air 'swells', the chair 'creaks', and the flowers 'shift', persists, rendering Jacob both present and absent. This tension between preservation and movement extends beyond Jacob's room to the very structure of *Jacob's Room*. Just as the room vibrates, Woolf's language quivers, resisting the finality of containment. Jacob's room and *Jacob's Room* function like butterfly boxes: intended to contain and preserve, yet ultimately inadequate for capturing true vitality. Through these layers of circling imagery, Woolf challenges the notion that identity, experience, or narrative can

[23] Harvena Richter, 'Hunting the Moth: Virginia Woolf and the Creative Imagination', in *Virginia Woolf: Revaluation and Continuity: A Collection of Essays*, ed. by Ralph Freedman (University of California Press, 1980) pp. 13-28 (pp. 16-17).
[24] Seshagiri, p. xxiii.
[25] Woolf, *Virginia Woolf's Jacob's Room: The Holograph Draft*, ed. by Edward Bishop (Pace University Press, 2010) p. 1.

ever befully contained. Jacob, like the moths and butterflies, remains in motion, alwaysslipping through the nets of language, only ever glimpsed.

We vibrate around Jacob, drawn into a perpetual, circular motion. The narrator notes that ‘something is always impelling one to hum vibrating, like the hawk moth, at the mouth of the cavern of mystery’ (p. 57). Woolf’s phrase to ‘hum vibrating’, evokes a kind of suspended, restless hovering, neither still nor in flight, but caught in an ongoing, trembling anticipation. Claudia Tobin notes that ‘although ‘hum’[ming] evokes sonic vibrations, the term also highlights the inadequacy of language and hints at non-verbal communion’.[26] Steven Connor writes that humming is suggestive of ‘some utterance that lies beneath or to the side of speech’.[27] This evokes Woolf’s own challenge in articulating Jacob’s presence with language, a presence that is felt, resonant, but never fully expressed in words. Woolf does not attempt to completely capture Jacob, as she is aware language will fail, she instead enacts circular vibrating noises to signal vitality or movement in her subjects. The narration hovers around Jacob as though repeatedly casting a net towards him, briefly catching resonances and impressions before they slip away again. Woolf concludes this passage remarking how ‘what remains is mostly a matter of guess work. Yet over him we hang vibrating’ (p. 57). The verb ‘hang’ connotes suspension, recalling an insect vibrating midair. The narrative and the reader, like Jacob, remain poised in midair, never quite settling. The repetition of ‘vibrating’ reinforces the idea that we do not move away from Jacob, nor do we pin him down; instead, we oscillate around him, caught in the spiral of simultaneous proximity and inaccessibility.

In *Jacob’s Room*, Woolf suggests that people cannot be summed up; Jacob is never fully defined but is glimpsed obliquely, much like the insects that flutter, circle, and evade capture throughout the novel. His presence is felt rather than fixed, reinforcing Woolf’s belief that identity cannot be pinned down through language. This refusal to define Jacob is not a failure but an aesthetic principle, embracing the beauty found in what is partial, elusive, and always in motion. Woolf’s writing does not capture but nets: perceiving fragments while accepting loss, to circle rather than to enclose.

---

[26] Claudia Tobin, *ModernismandStillLife:Artists, Writers, Dancers* (Edinburgh University Press, 2020) pp. 50-51.

[27] Steven Connor, quoted by Tobin, p. 51.

The term ‘netting’ is therefore crucial because, unlike ‘capturing’ or ‘pinning’, it preserves and anticipates the possibility of escape. Woolf’s imagery of moths, butterflies, and vibrating movement reflects her own conception of language as a porous net, one that may momentarily gather impressions but can never wholly secure the living essence beneath them. Jacob continually slips through these narrative meshes, leaving behind only fleeting traces, partial impressions, and reverberations. The reader is held in the tension between glimpsing and losing Jacob, circling endlessly around a figure who both invites and resists definition. We do not flee from Jacob;we engage in a circular humming around him.

BIBLIOGRAPHY

Primary Works

Woolf, Virginia, *TheDiaryofVirginiaWoolf:Vol.3,1925-1930* , ed. by Anne Oliver Bell and Andrew McNeillie (Hogarth, 1980)

---------, *Jacob's Room*, ed. by Urmila Seshagiri (Oxford University Press, 2022)

---------, *TheLettersofVirginiaWoolf:Vol.2, 1912-1922,* ed. by Nigel Nicolson and Joanne Trautmann (Hogarth, 1976)

---------, *TheLettersofVirginiaWoolf:Vol.4, 1929-1931* , ed. by Nigel Nicolson and Joanne Trautmann (Hogarth, 1978)

---------, 'The Narrow Bridge of Art', in *Granite and Rainbow*, ed. by Leonard Woolf (Harcourt, 1958), pp. 11-23

---------, *APassionateApprentice:TheEarly Journals, 1897-1909* , ed. by Mitchell Alexander Leaska (Hogarth, 1990)

---------, *Selected Essays*, ed. by David Bradshaw (Oxford University Press, 2009)

---------, *VirginiaWoolf'sJacob'sRoom: The Holograph Draft* , ed. by Edward Bishop (Pace University Press, 2010)

Secondary Works

Alt, Christina, *Virginia Woolf and the Study of Nature* (Cambridge University Press, 2010)

Flint, Kate, Introduction, Virginia Woolf, *Jacob's Room* (Oxford University Press, 2008) pp. xii-xxviii

Froula, Christine, *VirginiaWoolfandtheBloomsburyAvant-Garde: War, Civilization, Modernity* (Columbia University Press, 2004)

Raitt, Suzanne, ‘Virginia Woolf’s Early Novels: Finding a Voice’, in *TheCambridge Companion to Virginia Woolf*, ed. by Susan Sellers (Cambridge University Press, 2010), pp. 29–48

Richter, Harvena, ‘Hunting the Moth: Virginia Woolf and the Creative Imagination’, in *Virginia Woolf: Revaluation and Continuity: A Collection of Essays*, ed. by Ralph Freedman (University of California Press, 1980) pp. 13-28

Sarsfield, Rachel, ‘Cassandra’s Worms: Unravelling the Threads of Virginia Woolf’s Lepidoptera Imagery’, *Hungarian Journal of English and American Studies*, 9.1 (2003), pp. 101–17

Swanson, Diana L., “The Real World’: Virginia Woolf and Ecofeminism’, in *Virginia Woolf and the Natural World*, ed. by Kristin Czarnecki and Carrie Rohman (Clemson University Press, 2011), pp. 24–34

Tobin, Claudia, *Modernismand Still Life: Artists, Writers, Dancers* (Edinburgh University Press, 2020)

Authors Bio

I am a recent graduate from the University of Bristol, holding an MA in English Literature and a BA in English and Philosophy. My research interests include modernism, experimental literature, the works of Virginia Woolf and literary engagements with the natural world and non-human animals.

How and why did the ‘long historiography’ of the Reformation in Bristol evolve between the sixteenth and the early twentieth centuries?

Author: Nathan Roberts
ORCID: 0009-0006-0423-5589
Programme: BA History
Affiliation: School of Humanities
Email: natrob789@gmail.com

Abstract

This paper addresses a lacuna in Bristol’s Reformation historiography by tracing the evolution of its interpretations from the sixteenth to the early twentieth century. While national revisionism has inspired recent scholars to illuminate the vitality of lay devotion to the Catholic Church in Bristol on the eve of the Reformation, less attention has been paid to the development of its interpretations amongst earlier generations of historical writers. This paper argues that the ‘long historiography’ of Bristol’s Reformation advanced in response to changing authorial purposes: from annalists recording events, to antiquarians integrating religious developments into civic histories, to semi-professional historians focusing on specialised ecclesiastical study. This process was facilitated by concurrent methodological shifts, as antiquarians turned to official municipal documents before ecclesiastical historians explored parish inventories and churchwarden accounts. Once confessional biases faded, new questions were asked of these sources enabling both national and regional revisionism of the triumphalist Protestant narrative. Bristol’s historiographical trajectory is thus situated within the national progression of Reformation scholarship and the broader evolution of historical writing.

Keywords:Historiography, Reformation, Methodology, Sources, Revisionism

Historiography

Most analysesof the English Reformation have focused on answering two long-standing questions: how fast was the process of Protestant conversion, and by what mechanism did it occur? More precisely, was the new faith ardently embraced by a despondent laity, or were they coerced into accepting it through government statutes? The traditional interpretation, posited by A. G. Dickens, held that the Reformation was a rapid, bottom-up process that saw England turn irreversibly Protestant by the end of Edward VI's reign.[1] The revisionist works of Eamon Duffy and Christopher Haigh contested this view. Duffy demonstrated the strength of lay devotion to the late medieval Church, while Haigh argued that the population became majority Protestant only by the latter half of Elizabeth I's reign.[2]

These evolving national debates have shaped the most recent scholarship on the Reformation in Bristol. Clive Burgess has mined surviving churchwarden accounts and inventories to illuminate, in line with Duffy's work, how richly endowed and supported one Bristol parish was on the eve of the Reformation.[3] Martha C. Skeeters has analysed how the religious upheaval completely transformed the relationship between the clergy and laity in Bristol.[4] Meanwhile, Joseph Bettey has examined popular responses to some of the Reformation's seminal events such as the suppression of the religious houses in 1547.[5] While these scholars have illuminated the lived experience of the Reformation in Bristol, they have largely overlooked how earlier commentators interpreted and examined events. Though Bettey and Skeeters have used Bristol's manuscript calendars and early civic histories in their works, these sources are yet to be analysed independently to understand how and why portrayals of the Reformation evolved. Examining this process will shed light on how the contemporary understanding of the Reformation in Bristol has been reached.

---

1 A.G Dickens, *The English Reformation*, 2nd edn (Pennsylvania State University Press, 1989).
2 Eamon Duffy, *The Stripping of the Altars: Traditional Religion in England, c.1400-c.1580* (Yale University Press, 1992); Christopher Haigh, 'Success and failure in the English Reformation', *Past & Present*, 173 (2001), pp.28-49.
3 Clive Burgess (ed.), *The Pre-Reformation Records of All Saints' Church*, 3 parts (1995-2004, Bristol Record Society).
4 Martha C. Skeeters, *Community and Clergy: Bristol and the Reformation c.1530-c.1570* (Oxford University Press, 1993).
5 Joseph Bettey, *The Suppression of the Religious Houses in Bristol* (Bristol Historical Association Pamphlets, No. 74, 1990).

Furthermore, this study could be expanded to compare how the historicisation of the Reformation evolved across other English towns and cities during the same period.

Argument

In this paper, I will argue that the evolution of interpretations regarding Bristol's Reformation was driven by the changing aims of writers – from chroniclers to antiquarians to historians. Secondly, I will suggest that these changing priorities shaped the sources that these writers consulted. Along with concurrent developments in historical methodology, this resulted in an increasingly detailed picture of the Reformation being drawn. However, this did not lead to a sea-change in how the process was understood. The triumphalist narrative of the Reformation was only overturned when religious prejudices faded, and the nature of the questions being asked of the sources changed. Since this was a late twentieth century development, the purpose of this paper is to chart the evolution of accounts of Bristol's Reformation prior to this time. Accordingly, my analysis will begin with the manuscript chronicles, before progressing to eighteenth century antiquarians and nineteenth century civic histories. Finally, I will examine the emergence of the field of ecclesiastical history in Bristol, which took root in the late nineteenth century and grew into the early twentieth, exploring its implications for analyses of the Reformation.

Chroniclers

The aim of early chroniclers such as Robert Ricart - the first author of Bristol's official calendar - was to record notable yearly events, whether religious or secular. Ricart's six organising principles omitted a specific religious focus in his work. His is the only extant chronicle with explicit authorial commentary on its purpose.[6] For him and the other chroniclers examined here - the anonymous *Bristol Annal*, William Adams and William Morgan - the inclusion of religious events in a given year seemed to be a matter of personal whim, leading to significant inconsistencies in their accounts.[7] Gaps

[6] Lucy Toulmin Smith (ed.), *The Maire of Bristowe is Kalendar by Robert Ricart, Town Clerk of Bristol 18 Edward IV* (Camden Society, 1872), pp.3-6.

[7] William Tyson (ed.), 'William Morgan's Annals of Bristol, 1246-1687', *The Bristol Memorialist* (Bristol, 1823); Francis F. Fox (ed.), *Adams's Chronicle of Bristol* (J. W. Arrowsmith, 1910); Evan T. Jones (ed.), 'Bristol Annal: Bristol Archives 09594/1' (version 30 September 2019).

of several years between entries, as well as discrepancies in how incidents were recorded, mean a set canon of events in the Bristol Reformation cannot be determined by cross-referencing the chroniclers' accounts. For example, Adams' Chronicle provides the first explicit mention of early religious upheaval, noting that 'the reverend Doctor Barnes wore a faggot on at his back in Bristow' in 1525.[8] Robert Barnes, an early religious reformer, faced a heresy charge in 1526 because of a sermon he delivered in Cambridge and was ordered to do penance by wearing a faggot in London. He is not known to have come to Bristol.[9] In contrast, Ricart first acknowledges the Reformation implicitly in 1538: 'this yere was laste thofferyng on Mighelmas day at S. Mighells.'[10] This likely refers to the closure of the Augustinian canonesses' house on St Michael's Hill, stipulated under the Suppression of Religious Houses Act, enacted in March 1536.

This mismatch in the periodisation of the Reformation across the chronicles is difficult to explain definitively. Adams' inclusion of the information on Barnes may be because he compiled information from at least '3 books of several men's writings'.[11] There are no footnotes in the other manuscript calendars analysed here making it uncertain what sources they used, if any. The uncertainty this causes is highlighted by an inconsistency in the *Bristol Annal*. The original 1534 entry in the 'Annal of Bristol 1245-1735' records the visit of 'Thomas, Arch Bishop of Canterbury'.[12] However, in the additional notes of the manuscript, a different hand provided a separate 1534 account detailing the 'Great Troubles all over the Kingdom about preaching particularly in Bristol where Mr Latimer preached and there preached against him Mr Hobberton and Dr Powell'.[13] These events, later termed the 'Battle of the Pulpits', broke out following Hugh Latimer's Easter sermons in Bristol in 1533 to which conservative preachers John Hilsey and William Hubberdyne delivered public refutations.[14] The addition of this account by a later hand suggests the use of a source unavailable to previous contributors. It was common for authors to borrow information from other accounts to supplement their own writings. Adams noted in his 1532 entry that 'in this

---

8 Fox, p.86.
9 Jones, p.31.
10 Smith, p.54.
11 Fox, p.49.
12 Jones, p.32.
13 Jones, p.17.
14 G.R. Elton, *PolicyandPolice:TheEnforcementof the Reformation in the Age of Thomas Cromwell* (Cambridge University Press, 1972), pp.114-115.

year Mr Latimer and Hubbardine disputed together.'[15] It is plausible the hand that added to the Bristol Annal could have included this information from Adams's Chronicle, albeit the two-year discrepancy may indicate that the information came from another contemporary source now lost. Alternatively, this discrepancy may result from different dating practices that the chronicles used, whereby some took the year to begin with the mayor's appointment in March whilst others followed the Gregorian calendar.

The uncertainty about the sources used by the sixteenth century and seventeenth century chroniclers does not diminish the value of their Reformation portrayals for two reasons. First, they illustrate the events perceived to be worth recording. For example, the visitation of Thomas Cranmer, the Dissolution of St Augustine's Abbey and Gaunt's Hospital, and the creation of the Diocese of Bristol are referenced in all four chronicles. Adams' account is by far the most detailed, offering a complete, if basic, outline of the key events in Bristol's Reformation. A notable point of contention between the chronicles arises in the discussion of the number of Protestant martyrs executed during Mary I's reign. Adams identifies four men drawn and quartered in April 1555: Morgan mentions a weaver and a cobbler 'burnt for Religion called heresie'; the *Bristol Annal* refers to 'William Shapman weaver and Several others were burned' while Ricart has no entries between 1550 and 1568. Later studies, despite access to a wider range of sources, remain uncertain of the number that perished.[16] What remained constant though was that all chronicles ended their Reformation-related entries with Elizabeth's accession. Indeed, Adams interpreted Mary's premature death as evidence of the 'loving countenance of God over the protestants'.[17] This emphasis on Elizabeth's reign as the endpoint of the Reformation reflects Adams' explicit belief that God favoured Protestantism.

Indeed, the ability to discern early attitudes towards the Reformation is the second reason these chronicles retain historical value. Alongside a conviction that God favoured Protestantism, it is evident that the chroniclers were also deeply concerned with the maintenance of civic order: these were their two primary preoccupations. While there is little emotional sentiment expressed in the *Bristol Annal* and Morgan's

[15] Fox, p.87.
[16] K.G. Powell, *The Marian Martyrs of Bristol* (Bristol Historical Association Pamphlets, No. 31, 1972).
[17] Fox, p.106.

Chronicle, the writer's views seep through more in Ricart's Chronicle. For example, the 1539 entry condemned the extreme reformer George Wishart who 'set furth in his lecture in S. Nicholas church of Bristowe, the moost blasphemous heresy that euer was herd'.[18] Though Wishart was a reformer, the radicalism inherent in his suggestion that 'Christ nother hath nor coulde merite for hym ne yett for vs' ensured that the town's authorities - who had been lenient to other reformers such as Latimer - forced him to carry a faggot.[19] In Adams's Chronicle, his personal views are discernible in almost every entry concerning the Reformation. His opinions permeate through his account, whether discussing the 'certain monks, friars […] of a devilish intent' who suggested that God 'was highly displeased with the King for the divorcement from the Ladie Katherine', or in his reaction to the re-introduction of Catholicism under Mary - 'now God's true religion was abolished'.[20] More than any other chronicler, Adams pioneered the triumphant Protestant narrative of the Reformation.

Antiquarians and Civic Historians

The triumphalist narrative of the Reformation was underscored by the eighteenth-century antiquarian William Barrett and the nineteenth century civic historians Samuel Seyer and George Pryce. These writers shared a common aim: to produce the authoritative account of Bristol's history. Understanding this intention is crucial, because it informed the sources they consulted and their subsequent portrayal of the Reformation. Rosemary Sweet has charted the rise of Urban Histories in Eighteenth- Century England, asserting that the growth of antiquarianism and concomitant preservation of civic records spurred the development of specialist scholarship on towns.[21] This emphasis on using documentary records to create an impartial history was a stated aim among all three writers. For example, Barrett describes how 'the original deeds and copies collected for this History have been procured with so much labour'.[22] Ironically, the veracity of his *History* was tarnished by his inclusion of historical transcriptions forged by Thomas Chatterton. Hence, Seyer took pains to

---

18 Smith, p.55.
19 Smith, p.55.
20 Fox, p.88; p.102.
21 Rosemary Sweet, *The Writing of Urban Histories in Eighteenth-CenturyEngland* (Oxford University Press, 1997), p.36.
22 William Barrett, *The History and Antiques of the City of Bristol* (W. Pine, 1789), p.vi.

distance and emphasise the credibility of his *Memoirs*.[23] Likewise, Pryce took aim at Barrett's 'wretched history' filled with the 'fictions of Chatterton', though also castigated Seyer for treating the manuscript calendars as authorities on many 'minute particulars'.[24] Thus, the signal shift in priorities from the chroniclers to the antiquarians lay in the heightened emphasis on the authenticity and reliability of their sources. This not only distinguished these writers from chroniclers but also reflected a broader methodological shift toward the use of empirical evidence in scholarly works. As such, this shift was a prerequisite for the emergence of more rigorous accounts of Bristol's Reformation in the late eighteenth and early nineteenth centuries.

The shift toward verifiable evidence led to the systematic use of documentary sources for the first time which also demanded the use of new historical techniques. Barrett, unlike Seyer and Pryce, does not address the Reformation directly. His parish-by-parish approach offers fragmented insights. He largely omits discussion of the effects of the suppression of the smaller religious houses, instead focusing on the dissolution of St Augustine's Abbey and the Gaunt's Hospital.[25] His account is primarily descriptive, though is distinguished by his transcription of an array of previously unpublished sixteenth century documents. In discussing the creation of the new diocese in Bristol in 1542, he transcribes a document from 'the old liber valorum' which details the number of new cathedral officials and wages that they were to be paid.[26] Similarly, his use of 'inquisitions' from the Crown's Exchequer records sheds light on the valuation of the possessions of Gaunt's Hospital during its dissolution.[27] Although Barrett's systematic use of primary sources marked a wholesale departure from the chroniclers' approach, his failure to interpret the evidence and synthesise it into a more complete account of the Reformation illustrates the limitations of antiquarianism.

On the other hand, Seyer's account, situated within his discussion of the reigns of the incumbent monarchs during the Reformation, provides a more coherent interpretation. From the outset, he makes clear his intention to posit a narrative of Protestantism's swift triumph over 'Popery': 'during the reign of Henry VIII the doctrines

---

[23] Samuel Seyer, *MemoirsHistorical And Topographical Of Bristol And Its Neighbourhood; From The EarliestPeriodDownToThePresent Time*, Vol. 1 (J.M. Gutch, 1823), p.vi.
[24] George Pryce, *APopularHistory of Bristol* (W. Mack, 1861), p.vi; Seyer, p.xii.
[25] Barrett, pp.270-271; p.372.
[26] Barrett, p.283.
[27] Barrett, p.373.

of the Reformation made great progress […] It is evident they were embraced in Bristol with as much eagerness, as in the rest of the kingdom'.[28] From then on, he details events through the use of two sources: first, the manuscript chronicles, which he uses to provide explanations and dates of events. Second, and more significantly, he transcribes several contemporaneous letters from the Cotton Library Collection.[29] The Cotton Library Collection held as many as 360 manuscripts gathered by Sir Robert Cotton (1571-1631), an early antiquarian. These were previously unpublished, hence the inclusion of them in Seyer's work broadened the knowledge of contemporaneous reactions to Latimer's preaching and the commission Cromwell appointed to monitor religious discord. Once more, this is indicative of how early civic historians such as Seyer consulted and employed a broader range of historical documents than the chroniclers in their discussion of the Bristol Reformation.

Nonetheless, between the seventeenth and mid-nineteenth centuries, the interpretation of the Bristol Reformation as the overthrow of a corrupt Catholic church remained entrenched. Despite the new emphasis among antiquarians and early civic historians on veracity and their exploration of untapped sources, such as the Cotton Library manuscripts, the triumphalist Protestant narrative remained unchallenged. Even Pryce's *History*, published in 1861, offered little new. Despite condemning Seyer and Barrett's accounts, he cited the manuscript chronicles repeatedly.[30] This continuity was reinforced by the antiquarian focus on compiling documents rather than critically interrogating them. Moreover, Seyer himself was a member of the Church of England. These factors combined to ensure that the overall narrative of the Bristol Reformation continued to remain unchallenged into the mid-nineteenth century.

Ecclesiastical Historians

The emergence of ecclesiastical history in the late nineteenth and early twentieth centuries heralded the development of a more complicated narrative of the Reformation in Bristol, and most importantly, sowed the seeds for Duffy's revisionism a century later. The first such history was produced by J.F. Nicholls and John Taylor, who set out to explore the 'original muniments of the churches' and 'churchwarden

[28] Seyer, p.212.
[29] Seyer, pp.216-223.
[30] Pryce, pp.79-83.

accounts' which hitherto, they saw as having been 'sparingly used.'[31] When describing the Reformation in Bristol however, they relied primarily upon Seyer's account and the chronicles. Nonetheless, they made several novel contributions. First, they emphasised the powerful Benedictine system of monachism that had pervaded ecclesiastical culture in England right up to the eve of the Reformation.[32] Second, they problematised the beginning point of the Reformation, highlighting how Seyer's assertion of Bristol as a focal point of late medieval Lollardy relied solely on his reading of the manuscript chronicles. Third, they included not only the sermons preached by Latimer, but also those of the conservative Roger Edgeworth, providing a more accurate representation of the disorderly progression of the religious changes in Bristol. Most consequentially, they evaluated the reasons for the lack of popular resistance to the ecclesiastical and doctrinal changes in Bristol. They suggested that many figured it was 'easier to conform than burn', and that the 'popular conscience seemed to be at the disposal of the master-will of a powerful government, to be moulded […] to any religious form.'[33] This argument prefigured Elton's thesis nearly a century later in his *Policy and Police* that the Reformation was implemented by government coercion, spearheaded by Cromwell as chief minister. However, despite their promise, the authors do not cite any information from churchwarden accounts in their analysis of the Reformation. It was this source that was central to late twentieth-century revisionist accounts that challenged the dominant Protestant narrative.[34]

It was left to E. G. Cuthbert F. Atchley to transcribe the extant probate inventories, vestry agreements and churchwarden accounts in the early twentieth century.[35] His primary motivation for doing these transcriptions appears to have been to serve the interests of members of the *St Paul's Ecclesiological Society*, which promoted 'the non-professional study of Ecclesiology.'[36] Atchley's primary occupation was as a surgeon, though he was also a member of the Alcuin Club - an Anglican

---

31 J.F. Nicholls and John Taylor, *Bristol Past & Present, Vol II – Ecclesiastical History* (J. R. Arrowsmith, 1882), p.4.

32 Nicholls and Taylor, p.4.

33 Nicholls and Taylor, pp.19-20.

34 Ronald Hutton, 'The Local Impact of the Tudor Reformations', *The English Reformation Revised*, ed. by Christopher Haigh (Cambridge University Press, 1987), pp.114-138.

35 E. G. Cuthbert F. Atchley, 'Medieval Parish-Clerks in Bristol', *Transactions of the St Paul's Ecclesiological Society*, Vol. V (1904), pp.107-116; 'Some inventories of the parish church of St Stephen, Bristol', *Transactions of the St Paul's Ecclesiological Society*, Vol. VI, Part III (1908), pp.161-184.

36 *Transactions of the St. Pauls Ecclesiology Society*, Vol. 1 (Alabaster, Passmore and Sons, 1884), p.iii.

organisation that encouraged the study of worship and liturgy. Therefore, his work on ecclesiastical records appears to have been more of a hobby than a scholarly pursuit. This may explain why he merely described the transcriptions and did not interpret his findings and relate them to existing historical literature. Indeed, Atchley makes cursory reference to the broader implications when he examines the contents of the inventories or accounts in the years that marked the high tide of religious reform. For instance, when analysing the inventory of St Stephen in 1550, he details extensively the goods sold rather than comparing the contents of the inventory to preceding entries - he could have analysed how Edward's statutes against idolatry impacted parish worship.[37] Clive Burgess would later examine similar records from All Saints' Church to demonstrate the vitality of Catholic devotion on the eve of the Reformation in Bristol. Hence by 1908, sources of the kind that were eventually used to critique the triumphalist Protestant narrative had already been published. Not until Duffy used them to demonstrate how unwelcome the Reformation was to most people that other historians, such as Burgess, turned to examine Bristol parish records once more.

Conclusion

This paper has demonstrated that the development of the analyses of the Reformation from the sixteenth to the early twentieth century was contingent upon three interlinked factors. Firstly, the evolution in the priorities of the writers, from chronicling the key yearly events, to writing civic histories, to examining ecclesiastical history. Therefore, the portrayals of the Reformation became more complete throughout the period. This was aided by the concurrent development in historical methodology, whereby the emphasis on rooting explanations in documents first led Barrett and Seyer to civil and ecclesiastical records. Following that, Nicholls and Taylor, and Atchley explored parish warden accounts and inventories which further nuanced the understanding of the Reformation in Bristol. Their work laid the foundation for the Bristol historians of the late twentieth and early twenty-first centuries who - free from any affiliation with the Anglican Church and influenced by the broader revisionist movement - began asking new questions of these sources.

[37] Atchley, 'Some inventories of the parish church of St Stephen', p.172.

Bibliography

Primary Sources

Atchley, E. G. Cuthbert F., ‘Medieval Parish-Clerks in Bristol’, *Transactions of the St Paul’s Ecclesiological Society*, Vol. V (1904), pp.107-116

Atchley, E. G. Cuthbert F., ‘Some inventories of the parish church of St Stephen, Bristol’, *Transactions of the St Paul’s Ecclesiological Society*, Vol. VI, Part III (1908), pp.161-184

Barrett, William, *The History and Antiques of the City of Bristol* (W. Pine, 1789)

Jones, Evan T. (ed.), ‘Bristol Annal: Bristol Archives 09594/1’ (version 30 September 2019)

Fox, Francis F. (ed.), *Adams’s Chronicle of Bristol* (J. W. Arrowsmith, 1910)

Tyson, William (ed.), ‘William Morgan’s Annals of Bristol, 1246-1687’, *The Bristol Memorialist* (Bristol, 1823)

Nicholls, J. F. and Taylor, John, *Bristol Past & Present, Vol II – Ecclesiastical History* (J. R. Arrowsmith, 1882)

Pryce, George, *A Popular History of Bristol* (W. Mack, 1861)

Seyer, Samuel, *Memoirs historical and topographical of Bristol and its neighbourhood; from the earliest period down to the present time*, Vol. 1 (J.M. Gutch, 1823)

Smith, Lucy Toulmin (ed.), *The Maire of Bristowe is Kalendar by Robert Ricart, Town Clerk of Bristol 18 Edward IV* (Camden Society, 1872)

Transactions of the St. Paul’s Ecclesiology Society, Vol. 1 (Alabaster, Passmore and Sons, 1884)

Secondary Literature:

Bettey, Joseph, *TheSuppressionofthe Religious Houses in Bristol* (Bristol Historical Association Pamphlets, No. 74, 1990)

Burgess, Clive (ed.), *ThePre-Reformation Records of All Saints’ Church* , 3 parts (1995-2004, Bristol Record Society)

Dickens, A. G., *The English Reformation*, 2nd edn (Pennsylvania State University Press, 1989)

Duffy, Eamon, *The Stripping of the Altars: Traditional Religion in England, c.1400-c.1580* (Yale University Press, 1992)

Elton, G.R., *Policy and Police: The Reformation in the Age of Thomas Cromwell* (Cambridge University Press, 1972)

Haigh, Christopher, 'Success and failure in the English Reformation', *Past & Present*, 173 (2001), pp.28-49

Hutton, Ronald, 'The Local Impact of the Tudor Reformations', *The English Reformation Revised*, by Christopher Haigh (Cambridge University Press, 1987), pp.114-138

Powell, K.G., *The Marian Martyrs of Bristol* (Bristol Historical Association Pamphlets, No. 31, 1972)

Skeeters, Martha C., *Community and Clergy: Bristol and the Reformation c.1530-c.1570* (Oxford University Press, 1993)

Sweet, Rosemary, *The Writing of Urban Histories in Eighteenth-Century England* (Oxford University Press, 1997)

Authors Bio

Nathan Roberts has recently completed a BA in History at the University of Bristol. Over the course of his degree, his research became increasingly focused on the early modern period. He particularly enjoyed the opportunity to study Bristol across political, economic, religious, and legal contexts. This allowed him to work on history within a local setting, which he felt made it more tangible. He hopes either to continue at Bristol to study an MPhil in History or to begin a law conversion course elsewhere.

Policy and Governance

Meet the Subject Editor

Anna Shahbazyan
Policy and Governance

Anna Shahbazyan is currently pursuing a Master's degree in Education Policy and International Development at the University of Bristol. With previous editorial experience across several undergraduate literary magazines, she was excited to join the Student Research Journal as an Education, Policy, and Arts Subject Editor. The role has provided her with the opportunity to engage with a wide range of interdisciplinary scholarship and gain insight into the diverse and innovative research being undertaken by students across disciplines. Her academic interests include multilingual education, refugee and migrant education, educational inclusion, and language policy implementation. She hopes to continue working in the field of education, contributing to the development of policies that promote equitable and inclusive learning opportunities for students around the world.

Political Hate Speech as anelectoral strategy in India

Author: Kongkita Konwar
ORCID: 0009-0005-5581-3863
Programme: MSc Public Policy
Affiliation: School for Policy Studies
Email: sz25310@bristol.ac.uk

Abstract

This paper uses the 2024-2025 election cycle as a case study to critically examine how political hate speech is used as a deliberate electoral strategy in India. The paper interprets the empirical data within broader theoretical debates of ethnic democracy, populist reason and dangerous speech. This paper identifies five hate speech patterns: conspiracy theories and dog whistles, construction of demographic threat, religious victimisation, punitive populism, and securitisation politics. It argues that hate speech in India is not a series of random incidents but is institutionally organised from the top down and deeply rooted within the majoritarian state project that Jaffrelot (2021) characterises as 'ethnic democracy.' Applying Mintrom's (2012) Implementation Analysis to three viable policy options: institutional and legal reform, civic peacebuilding, and curriculum development and community dialogue, the findings of the paper highlight that the central governance failure is one of political will rather than legislative absence. Furthermore, the paper contends that the analytical limitations of Implementation Analysis, particularly in its bureaucratic formalism and failure to account for power asymmetries, make it necessary to adopt structural and discourse-oriented perspectives. Finally, the paper concludes that any durable counter-hate speech framework must not only be grounded in institutional autonomy but also media accountability and long-term civic education, while continuing to survive the entrenched political interests that constrain each of these.

Keywords: political hate speech, electoral strategy, communal polarisation, India,

ethnic democracy

Introduction

India's 2024 general election, the largest exercise in history, was marked not only by its magnitude but also for the unusually high level of anti-minority language that dominated its campaign narrative. In March 2024, the United Nations Human Rights Council raised a formal warning of the ongoing attacks on religious and ethnic minorities and urged corrective action before polling began (UNHRC, 2024). Very unfortunately, that warning went unheeded. India Hate Lab (2025, p. 13) reports that in-person hate speech incidents against religious minority increased from 668 in 2023 to 1,165 in 2024 (74.4%) and reached 269 incidents in May during the campaign peak.

The analytical significance of this phenomenon lies not only in how widespread it is, but in its character. According to the argument presented, hate speech in contemporary India is not incidental to elections, but central to them, operating as a coordinated, top- down institutional strategy aimed at building majoritarian solidarity and mobilising votes by creating narratives of existential danger. The distinction between episodic bigotry and structured political violence is central to the (Benesch, 2012; 2016) literature on dangerous speech and to the comparative political sociology of Hindu nationalism (Jaffrelot, 2021; Chatterji, Hansen and Jaffrelot, 2019).

According to Andres and Slivko (2021), hate speech refers to any communication that is hostile or derogatory towards individuals based on gender, race, religion or political identity. Pukallus and Arthur (2024, p. 467) defined it as 'communicative driver of oppression.' But, in the Indian context, this general definition needs to be refined because what is at stake is not simply offensive speech but what the UNESCO framework identifies as its institutionalisation, the chronic process by which hate speech 'shifts from unconventional exchange to policy discourse' in settings where ruling parties exert dominance over media and regulatory spaces (Gagliardone et al., 2015).

The paper proceeds in the following way. Section 2 explains the methodology used in the study, followed by Section 3 which examines five patterns of political hate speech used as an electoral strategy in India during the 2024-2025 elections. Section 4 looks at possible policy responses using Mintrom's (2012) Implementation Analysis, while Section 5 reflects on the limits of this framework. The paper concludes in Section 6.

Methodology

The sources of this paper are derived from databases of hate speech, legislative records, and Election Commission rules. It also analyses the recorded hate speeches made by politicians. Furthermore, a critical approach is being used here, and it is understood through a specific theoretical framework. This study also builds on existing research about populism, hate speech and how certain issues are constructed to present as security threats. Furthermore, reports from India Hate Lab, Human Rights Watch, Amnesty International, South Asia Justice Campaign, Citizens for Justice and Peace, and the Mahanirban Calcutta Research Group are being reviewed.

Mintrom's (2012, pp. 292-295) Implementation Analysis is applied to the policy options, aseven-step structured approach through which the institutional requirements and threats to successful implementation of policy is systematically interrogated. The selection of this particular framework is due to its emphasis on implementation gaps, which is particularly apt for the Indian context, where the Supreme Court of India itself has noted that 'the root of the problem is not the absence of laws but a lack of effective execution' (PUCL, 2024). The paper, however, also critically interrogates the limitations of this framework, drawing on theory of populist reason by Laclau (2005) and dangerous speech framework by Benesch (2012; 2016) to supplement its bureaucratic logic with structural and discursive analysis.

Anote on scope and positionality:
Hate speech events recorded from January 2024 and early 2025, with special focus on the Lok Sabha elections is being covered in this analysis. The report doesn't provide an all-inclusive record of events; however, it does provide evidence that has been verified against external sources, and based upon the evidence provided, it highlights some larger structural trends. Furthermore, it acknowledges the contested nature of 'hate speech' as a category and addresses the tension between free expression and minority protection that runs through all regulatory debates in this domain.

Key Findings

According to research data collected in 2024, it appears that instances of hate speech were shaped by a clear logical framework rather than a random arrangement. The India Hate Lab (2025) further verified this by reporting that the BJP directly facilitated 340 gatherings, 29.2% of all documented hate speech events. This represents a 580% increase from 2023, when the party organised only 50 such events. Evidently, while the state-level politicians were determined as the primary instigators in 2023, the year 2024 saw national figures such as Prime Minister Narendra Modi, Home Minister Amit Shah, and Uttar Pradesh Chief Minister Yogi Adityanath stand out as the central actors of communal rhetoric (India Hate Lab, 2025). The analysis of this top-down structural shift is significant because it signals not only intensification of societal bigotry but also deliberate mainstreaming of hate speech as an official mode of campaign communication. As the Citizens for Justice and Peace (CJP, 2025) concluded in their structural analysis, the absence of enforcement does not fail to stop speech but actively incentivises it.

In this architecture of electoral hate speech as a political campaign strategy, five recurring patterns can be identified, each linked to a distinct theoretical framework.

Conspiracy Theories and Dog Whistles

The deployment of politically constructed conspiracy theories has served as the ideological bedrock of the BJP's electoral campaigns. The concept of 'love jihad,' a specious claim that Muslim men systematically ensnare Hindu women in marriage to convert them to Islam, functions as what Laclau (2005) would term an 'empty signifier': a term flexible enough to encapsulate a diverse chain of anxieties (demographic decline, sexual vulnerability, cultural contamination) into a single, mobilising narrative. Froystad (2021) similarly identifies these narratives as 'sound bites' whose rhetorical power lies in their ability to trigger affective responses without the burden of evidential support.

By 2024, the entire ecosystem had proliferated into an extensive repertoire: ‘land jihad,’ ‘vote jihad,’ ‘halal jihad,’ and others (India Hate Lab, 2025, pp. 9-10). Critically, in 2024, out of 1,165,
581 hate speech events were documented, in which almost exactly half incorporated such conspiracy theories (India Hate Lab, 2025). In a paradigmatic instance, the Uttar Pradesh Chief Minister, Yogi Adityanath, during his campaign in Jharkhand, alleged that mass Bangladeshi and Rohingya infiltration have led to ‘love jihad’ and ‘land jihad,’ posing a threat to women, jobs, and land (Hindutva Watch, 2024). What is notable here is not the content alone but the mechanism: the ‘jihad’ framing operates through Benesch’s (2012) concept of ‘accusation in a mirror,’ which frames the targeted group as the aggressor in order to rationalise defensive violence against them.

Demographic Threat

The securitisation of cross-border migration, where a social and administrative issue is framed as a security threat, conforms to the theoretical logic advanced by Buzan, Weaver and Wilde (1998, p. 26), who argue that any issue can be transformed into a security matter through ‘speech acts’ of authorised ‘security actors.’ In the Indian context, this mechanism has been systematically instrumentalised against Muslim migrants and Indian Muslims, who are constructed as proxies for Pakistani or Bangladeshi infiltrators.

Mahanirban Calcutta Research Group (2024, pp. 12-13) demonstrates how this securitisation has resulted in discriminatory and exclusionary practices in border regions. For instance, the Chief Minister of Assam, Himanta Biswa Sarma, at a campaign event in Jharkhand, stated that the election was about ‘driving out infiltrators from Jharkhand and saving Hindus’ and he warned that if infiltration went unchecked, the Hindu population in Jharkhand would fall to 50% within twenty years (The Indian Express, 2024). Jaffrelot (2021) argues that such demographic threats serve a dual function, which is that they construct a unified Hindu ‘people’ by invoking a logic of equivalence, in which all Hindus are framed as subject to a common threat, while concurrently dehumanising Muslims as existential others. Benesch (2016) identifies this exact structure as dangerous speech, in which the target group is portrayed as a moral

threat, making violence appear not merely legitimate but necessary.

Religious Victimisation

Construction of Hindu victimhood, reframing majority as a narrative of cultural persecution is one of the most strategically refined elements of BJP's rhetorical repertoire. Prime Minister Narendra Modi's speech at Jaipur on 23rd April 2024, demonstrates this further pattern. He alleged that Congress had protected those who 'pelted stones' at Ram Navami processions and framed the electoral victory of BJP as the guarantee of unrestricted Hindu religious practice. This speech was one of the seventeen by senior BJP officials that appeared to breach Section 123(3) of the Representation of the People Act, 1951, which prohibits seeking votes on religious grounds (MG, 2024). The South Asia Justice Campaign (2024) documented that between the official notification of the election (16 March) and the end of campaigning (29 May), they recorded 380 instances of hate speech by senior political and religious figures, of which 72 (19%) met the UN's threshold for 'top-level' hate speech. Despite multiple complaints, the Election Commission of India (ECI) failed to take meaningful action, and the Election Commission Chief later admitted that the body had 'deliberately decided' not to take action against the top leaders of either party, which reflects a statement amounting to institutional abdication (The Wire, 2024).

Punitive Populism

Punitive populism, as defined by Wood (2014), refers to the mobilisation of popular support for coercive state action against targeted groups, framed as a means of restoring social order. In the Indian context, this has taken its most visceral form, commonly known as 'bulldozer justice,' which involves the extrajudicial demolition of Muslim properties, residences and places of worship by state authorities, often within the immediate aftermath of communal incidents.

Amnesty International (2024) documented 128 targeted demolitions across Assam, Delhi, Gujarat, Madhya Pradesh, and Uttar Pradesh. A paradigmatic example occurred on 15 June 2024 in Mandla, Madhya Pradesh, where homes of eleven Muslim families

were demolished by authorities on the basis of alleged presence of beef in their refrigerators, while sixteen other allegedly illegal structures in the same neighbourhood, where no beef was found, were left intact (Anand Mohan J, 2024). While the Supreme Court subsequently issued guidelines to restrain discriminatory demolitions (Karpuram, 2024), the political symbolism of bulldozer had already been established as a campaign signifier, which in Laclau's (2005) terms, is a floating signifier, which is capable of condensing anxieties about crime, sovereignty, and Muslim presence into a single image of majoritarian authority. As the CJP (2025) observed, the bulldozer has become a'structured and institutionalised feature of Indian political life,' not an aberration.

Securitisation politics

The final and most dangerous pattern is expressed through the explicit dehumanisation of minority communities in the language of national security, a move that Benesch (2012) identifies as a key indicator of dangerous speech, associated historically with the escalation of mass violence. For instance, the Chief Minister of UP, Yogi Adityanath, stated that the opposition would 'distribute the wealth of the citizens to infiltrators, i.e., Bangladeshi, Pakistani or any other Muslim infiltrators' (India Hate Lab, 2025, p. 28). This is precisely what Benesch (2016) describes as a 'accusation in a mirror' logic, which represents the targeted group as the aggressor. This represents the advancement of what Jaffrelot (2021, p. 82) terms as 'civilisational majoritarianism', a systemic project aimed toward Hinduisation of India through exclusion and subordination of minority populations.

Evaluating Policy Options

Mintrom's (2012) Implementation Analysis is used to evaluate three policy options. The framework is used not as a technocratic exercise but as a structured lens for identifying where the gap between policy aspiration and political reality is more pronounced and why.

Strengthening Institutional Mechanisms and Legal Procedures

India's existing legal framework is not devoid of resources. Sections 153A and 295A of the Indian Penal Code criminalise communal incitement. Section 123(3) of the Representation of the People Act, 1951, prohibits religious appeals in electoral campaigns, and the ECI's Model Code of Conduct (MCC) sets out an explicit prohibition of communal campaigning. As the Supreme Court itself has acknowledged, the central failure is not the absence of law but paralysis in enforcement (PUCL, 2024). The ECI, once regarded as a global reference model for electoral administration, has, in The Wire (2024) assessment, become 'increasingly susceptible to executive influence,' with its Chief Commissioner publicly defending the non-imposition of sanctions on the Prime Minister for speech violations through the logic of institutional symmetry.

Howlett (2018) argues that effective policy instruments must 'fit' their operational context. In this context, the ECI cannot be an impartial enforcer when its independence is itself contested. The 2024 election saw thousands of citizen complaints lodged with the ECI regarding Modi's speeches with no meaningful impact (HRW, 2024). According to a PUCL (2024) cited survey, 58% of respondents expressed distrust in the ECI by 2024, compared with 78% who reported trust following the 2019 elections.

At a minimum, institutional reform would need to establish constitutionally guaranteed independence for ECI appointments, removing any executive discretion. A multilingual monitoring unit for hate speech, with algorithmic support akin to Indonesia's NLP-based BAWASLU system (Wijnarko et al., 2024); time-bound, mandatory investigation protocols for Model Code of Conduct (MCC) violations; and alongside graduated sanctions that apply even to 'star campaigners.' The Implementation Analysis highlights that the main threat to this option is political, as enforcement institutions are controlled by those whose behaviour the policy seeks to curb.

Civic Engagement for Peacebuilding

The structural problem with India's media landscape, which Gagliardone et al. (2015) identify as the capture of media houses by ruling parties, is exacerbated by ownership patterns, particularly the fact that several leading Hindi-language television networks are owned by individuals closely tied to the BJP government. As Pukallus and Arthur (2024) proposed, this concentration of media power means that the 'mixed advisory and regulatory space' faces a fundamental challenge, which suggests that any government- involved body is vulnerable to co-optation, while civil society options lack the power to enforce decisions.

Nonetheless, civil society initiatives represent a crucial second line of resistance. A glaring example could be derived from the Hate Speech Beda initiative in Karnataka (Boruah, 2024), which tracks, documents, and legally challenges hate speech, demonstrating that civil society can create accountability pressure even without institutional enforcement powers. Through training young leaders to map digital identity dynamics (Pukallus and Arthur, 2024, p. 478), the DMAPS programme presents a scalable model for strengthening media literacy and counter-narrative capacity. The Strong Cities Network (2024) provides an international framework for local government engagement.

Curriculum Development and Community Dialogue

According to UNESCO (2023), critical thinking, socio-emotional skills, and intercultural dialogue should be integrated into education systems as foundational commitments rather than supplementary initiatives. Education as a long-term resilience is the most compelling case, as Gagliardone et al. (2015) argues that societies with stronger traditions of civic education and historical literacy have shown greater resistance to the escalation of hate speech. However, in the context of Indian curriculum, a specific political challenge since 2014 is that the BJP government, through systematic textbook revisions, has reduced the prominence of content perceived as incompatible with Hindutva ideology, such as India's pluralist constitutional traditions, the cultural

contribution of Muslim rulers, and historical accounts of communal violence (Chatterji, Hansen and Jaffrelot, 2019).

Community-based programmes which are delivered through non-governmental organisations, self-help groups, and civil society organisations at the grassroots level face fewer risks of top-down curricular intervention than state-controlled education systems and are better positioned to address India's rural-urban education divide. A particularly relevant example here is The Strong Cities Network (2024) model, which equips local authorities with context-specific interventions against hate speech and polarisation. The Implementation Analysis highlights that implementation may be hindered by institutional inertia in state education boards, many of which are under BJP- aligned administrators.

Critical Reflection on Implementation Analysis:

Mintrom's (2012) Implementation Analysis is a useful diagnostic tool precisely because itrejects the idea that well-designed policies lead to well implementation of its own.

However, the limitations of the framework are significant and must be acknowledged. In the Indian context, where the ECI, the judiciary, and the police operate under varying degrees of political pressure from the executive, a framework that assumes these institutions as neutral agents of implementation misconstrues the political economy of implementation. As Fowler (2021) contends that ambiguity and uncertainty in implementation are not merely administrative problems but are often politically produced and politically functional, allowing powerful actors to selectively apply or ignore rules fitted to their interests. Second, the framework's emphasis on quantifiable outcomes (a 30% reduction in hate speech events or faster response times) risks reducing a multifaceted social and political phenomenon to a set of administrative metrics. Hate speech is not simply an outcome to be measured and reduced; it reflects deeper structures of ethnic majoritarianism whose roots demand structural political remedies, not just administrative. Laclau's (2005) theory of populist reason is illuminating here: BJP's construction of 'the Hindu people' against the Muslim

‘other’ is not a policy error to be corrected but a hegemonic project that has reshaped Indian political identity. No implementation plan addresses this at the level of its political logic. Third, the framework pays insufficient attention to multilingual complexity. India’s hate speech landscape is overwhelmingly conducted in Hindi and regional languages; the most dangerous speeches are often precisely those that do not translate into English and therefore escape the notice of national and international monitoring systems. This requires not merely regional language moderators but a fundamentally multilingual approach to monitoring, adjudication, and counter-speech.

Conclusion

Since 2014, under the BJP, political hate speech in India has undergone a qualitative transformation. What was once episodic has become normalized in a more structured, top-down strategy through the manufacture of Muslim threat. The 74.4% surge in verified hate speech events between 2023 and 2024, and the 580% increase in BJP-organised hate speech gatherings. From normalisation of hate speech to the institutionalisation of what Jaffrelot (2021) calls ethnic democracy and what Benesch (2012) identifies as the preconditions for dangerous speech to produce real-world violence.

The Implementation Analysis of three policy responses - institutional reform, civic peacebuilding, and education - reveals that none is sufficient in isolation, as all confront a common structural obstacle in which the political actors driving hate speech also control its regulators. ECI reform without genuine constitutional independence will remain cosmetic; media regulation without structural change in ownership patterns will be captured; curriculum reform without political independence will be reversed. As mentioned above, the Supreme Court’s own observation that the problem is not the absence of law, but the failure of execution implies that implementation is not a technical problem but a political one.

The paper has its own limitations. The reliance on English-language sources and documented incidents inevitably undercounts the true volume of hate speech in regional

language contexts, where the most virulent and locally effective rhetoric circulates. Implementation Analysis, as argued above, underestimates the structural and discursive dimensions of the problem. Future research should engage more deeply with the lived experiences of affected minority communities' perspectives that are systematically absent from policy-oriented frameworks, and with the comparative literature on hate speech regulation in analogous contexts of democratic backsliding, including Hungary, Brazil, and the United States. What India's trajectory makes clear, however, is that the defence of pluralist democracy against majoritarian populism is not primarily a legal or institutional problem, but a political and moral one, and that the international community's capacity to address it is constrained by the very norms of sovereignty and non-interference that authoritarian populists routinely exploit.

Authors Bio

Kongkita Konwar is a policy researcher at the University of Bristol, specialising in democratic governance, political communication, and qualitative research. Her academic work is grounded in three years of fieldwork experience across government, civil society, and electoral politics in India, where she developed a deep interest in how political narratives are constructed and weaponised to serve electoral ends. She believes that understanding hate speech not merely as a social harm but as a deliberate strategic instrument is essential to protecting democratic integrity in an era of accelerating disinformation. When not dissecting the mechanics of political manipulation, she restores her faith in humanity by reading, journalling, dancing, and watching trees and sunsets with a well-deserved glass of wine.

References:

ADR (no date) *Details of IPC Sections 153A, 295 & 295A*. Available at: https://adrindia.org/sites/default/files/Details%20of%20IPC%20Sections%20153A,%2029 5%20 &%20295A.pdf (Accessed: 22 November 2025).

Amnesty International (2024) *Stop bulldozer injustice in India*, 7 February. Available at: https://www.amnesty.org/en/latest/news/2024/02/india-authorities-must-immediately-stop-unjust-targeted-demolition-of-muslim-properties/ (Accessed: 5 December 2025).

Anand Mohan, J. (2024) 'We demolished homes where beef was found: inside the MP village where bulldozers came a day after police', *The Indian Express*, 19 June. Available at: https://indianexpress.com/article/india/we-demolished-homes-where-beef-was-found-inside-the-mp-village-where-bulldozers-came-a-day-after-police-9400789/ (Accessed: 5 December 2025).

Andres, R. and Slivko, O. (2021) 'Combating online hate speech: the impact of legislation on Twitter', *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.4013662.

Benesch, S. (2012) *Dangerous speech: a proposal to prevent group violence*. Washington, DC: Dangerous Speech Project. Available at: https://dangerousspeech.org/wp-content/uploads/2018/01/Dangerous-Speech-Guidelines-2013. pdf (Accessed: 10 January 2026).

Benesch, S. (2016) *Dangerous speech: a practical guide*. Washington, DC: Dangerous Speech Project. Available at: https://dangerousspeech.org/a-practical-guide-to-dangerous-speech/ (Accessed: 10 January 2026).

Boruah, M. (2024) "Serious job': how a collective in election-bound India fights hate speech', *Al Jazeera*, 7 May. Available at: https://www.aljazeera.com/news/2024/5/7/how-a-collective-in-indias-karnataka-is-standing-up-to-hate-ahead-of-polls (Accessed: 19 November 2025).

Buzan, B., Waever, O. and Wilde, J. de (1998) *Security: a new framework for analysis*. Boulder, CO: Rienner.Chatterji, A.P., Hansen, T.B. and Jaffrelot, C. (eds.) (2019)

Majoritarian state: how Hindu nationalism is changing India. London: Hurst. Available at: https://doi.org/10.1093/oso/9780190078171.001.0001.

Chowdhury, R. (2025) 'Indifferent to alarming surge in hate speech, India outrages over comic's crass comment on YouTube show', *The Diplomat*, February. Available at: https://thediplomat.com/2025/02/indifferent-to-alarming-surge-in-hate-speech-india-outrages-ov er-comics-crass-comment-on-youtube-show/ (Accessed: 5 December 2025).

CitizensforJusticeandPeace(CJP)(2025) *The architecture of polarisation: a structural analysis of communal hate speech as a core electoral strategy in India (2024–2025)*.Available at:https://cjp.org.in/the-architecture-of-polarisation(Accessed:3 January 2026).

Fowler, L. (2021) 'How to implement policy: coping with ambiguity and uncertainty', *Public Administration*, 99(3), pp. 581–597. Available at: https://doi.org/10.1111/padm.12702.

Froystad, K. (2021) 'Sound biting conspiracy: from India with "love jihad"', *Religions*, 12(12), p. 1064. Available at: https://doi.org/10.3390/rel12121064.

Gagliardone, I., Gal, D., Alves, T. and Martinez, G. (2015) *Countering online hate speech*. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000233231 (Accessed: 5 December 2025).

Hate Speech Beda (2023) *Campaign against hate speech*. Available at: https://hatespeechbeda.wordpress.com/ (Accessed: 6 December 2025).

Hindutva Watch (2024) 'Uttar Pradesh Chief Minister Yogi Adityanath delivers hate speech in Nala, Dhumka', *Hindutva Watch*, 18 November. Available at: https://www.hindutvawatch.org/uttar-pradesh-chief-minister-yogi-adityanath-delivers-hate-speec h-in-nala-dhumka/ (Accessed: 5 December 2025).

Howlett, M. (2018) 'The criteria for effective policy design: character and context in policy instrument choice', *Journal of Asian Public Policy*, 11(3), pp. 245–266. Available at:

https://doi.org/10.1080/17516234.2017.1412284.

Human Rights Watch (2024) *India: hate speech fueled Modi's election campaign*, 14 August. Available at: https://www.hrw.org/news/2024/08/14/india-hate-speech-fueled-modis-election-campaign (Accessed: 11 November 2025).

India Hate Lab (2025) *Hate speech events in India: report 2024*. Available at: https://www.csohate.org/wp-content/uploads/2025/02/Hate-Speech-Events-in-India_Report_2024.pdf (Accessed: 5 December 2025). Also available at: https://ssrn.com/abstract=5130832.

Jaffrelot, C. (2021) *Modi's India: Hindu nationalism and the rise of ethnic democracy*. Princeton, NJ: Princeton University Press.

Karpuram, A. (2024) 'SC issues guidelines to curb illegal demolitions: what it said, what do state laws say?', *The Indian Express*, 13 November. Available at: https://indianexpress.com/article/explained/explained-law/sc-guidelines-illegal-demolitions-9667 414/ (Accessed: 5 December 2025).

Laclau, E. (2005) *On populist reason*. London: Verso.

Mahanirban Calcutta Research Group (2024) *Human security and migration in South Asia: analysing India's response to the Global Compact for Safe and Orderly Migration*. Available at: http://www.mcrg.ac.in/PP168.pdf (Accessed: 5 December 2025).

Mintrom, M. (2012) *Contemporary policy analysis*. New York: Oxford University Press.

People's Union for Civil Liberties (PUCL) (2024) *Open letter to Election Commission of India on concerns regarding the conduct of the Lok Sabha Elections 2024*. Available at: https://pucl.org/manage-writings/open-letter-to-election-commission-of-india-on-concerns-regarding-the-conduct-of-the-lok-sabha-elections-2024/ (Accessed: 5 January 2026).

Pukallus, S. and Arthur, C. (2024) 'Combating hate speech on social media: applying targeted regulation, developing civil-communicative skills and utilising local evidence-based anti-hate speech interventions', *Journalism and Media*, 5(2), pp. 467–484.

Available at: https://doi.org/10.3390/journalmedia5020031.

South Asia Justice Campaign (2024) *India hate speech monitor*. Available at: https://southasiajusticecampaign.org/hate-speech-monitor/ (Accessed: 10 January 2026).

Strong Cities Network (2024) *Countering hate speech in times of crisis: considerations for mayors and local governments*. Available at: https://strongcitiesnetwork.org/resource/countering-hate-speech-in-times-of-crisis-considerations-for-mayors-and-local-governments/ (Accessed: 7 December 2025).

The Indian Express (2024) 'Jharkhand polls for driving out infiltrators, saving Hindus, claims Himanta', *The Indian Express*, 2 November. Available at: https://indianexpress.com/article/india/jharkhand-polls-for-driving-out-infiltrators-himanta-biswa-s arma-9650162/ (Accessed: 5 December 2025).

The Milli Gazette (2024) 'Hate speeches and spiral of hatred in the build-up to 2024 general elections in India', *The Milli Gazette*, 2024. Available at: https://www.milligazette.com/news/7-analysis/34108-hate-speeches-and-spiral-of-hatred-in-the-build-up-to-2024-general-elections-in-india/ (Accessed: 5 December 2025).

The Wire (2024) 'How the Election Commission is failing the hate speech test', *The Wire*, 12 May. Available at: https://m.thewire.in/article/government/election-commission-hate-speech (Accessed: 5 January 2026).

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2023) *Addressing hate speech through education: a guide for policymakers*. New York: United Nations. Available at: https://www.un.org/en/genocideprevention/documents/publications-and-resources/Addressing_hate_speech_through_education_A_guide_for_policy_makers.pdf (Accessed: 5 December 2025).

United Nations Human Rights Council (UNHRC) (2024) *India:UNexpertsurge corrective action to protect human rights and end attacks against minorities in lead up*

to elections, 7 March.

Available at: https://www.ohchr.org/en/press-releases/2024/03/india-un-experts-urge-corrective-action-protect-human-rights-and-end-attacks (Accessed: 10 January 2026).

Virani, S. (2024) 'Hate speeches at poll rallies: citizens challenge ECI to "grow a spine or resign"', *Hindustan Times*, 12 May. Available at: https://www.hindustantimes.com/cities/mumbai-news/hate-speeches-at-poll-rallies-citizens-chall enge-eci-to-grow-a-spine-or-resign-101715454513758.html (Accessed: 5 December 2025).

Wijanarko, M.I., Septiandri, A.A., Muis, I., Buntoro, G.A. and Santoso, H.B. (2024) 'Monitoring hate speech in Indonesia: an NLP-based classification of social media texts', in Hernandez

Farias, D.I., Hope, T. and Li, M. (eds.) *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: system demonstrations*.Miami, FL: Association for Computational Linguistics, pp. 142–152. Available at: https://doi.org/10.18653/v1/2024.emnlp-demo.15.

Wood, W.R. (2014) 'Punitive populism', in Miller, J.M. (ed.) *The encyclopedia of theoretical criminology*. 1st edn. Hoboken, NJ: Wiley, pp. 1–4. Available at: https://doi.org/10.1002/9781118517390.wbetc140

Paradigm Shifts and Policy Reform: The Role of New Ideas in UK Asylum Policy

Author: Foday Kabia

ORCID: 0009-0003-1369-2026

Programme: MSc in Public Policy

Affiliation: School of Social Policy

Email: qn25910@bristol.ac.uk

Abstract

This essay examines how new ideas drive fundamental transformation in public policy, using the evolution of UK asylum policy as a case study. Drawing on Hall's (1993) theory of policy paradigms, it argues that the shift from a humanitarian protection model to a deterrence-oriented control paradigm represents a third-order, paradigmatic change. The analysis shows how asylum, one grounded in the moral and legal obligations of the 1951 Refugee Convention, has been reframed since the 1990s as a challenge of border management, welfare pressure, national security. Schmidt's (2010) discursive institutionalism explains how political elites and bureaucratic actors legitimised this shift through narratives portraying asylum seekers as "illegal migrants" or security risks, reshaping public discourse and administrative priorities. Pierson's (1997) concepts of increasing returns and path dependence further illuminate how early reforms such as dispersal, detention, and accelerated procedures became self-reinforcing, embedding deterrence within institutional routines. Through analysis of key legislation, including the Immigration and Asylum Acts (1990, 2002) and the Nationality and Borders Act (2022), the essay demonstrates how ideas redefined policy goals, instruments and implementation practices. It concludes that UK asylum policy has undergone a paradigmatic reorientation sustained by political incentives, institutional inaction, and discursive reframing, illustrating the powerful role of ideas in shaping policy reform.

I would like to express my sincere gratitude to my module tutor Noemi Lendvai-Bainton and seminar leaders for their guidance, constructive feedback, and intellectual support throughout the development of this essay. Their support strengthened my understanding of policy paradigms and ideational change. I am also grateful to my peers for their thoughtful discussions, which helped refine my arguments and deepen my engagement with the literature on asylum policy. Finally, I extend my appreciation to the School of Social Policy for providing an enriching academic environment that has shaped my learning and research.

Keywords: policy innovation, asylum reform, paradigm shift, advocacy networks, public discourse.

INTRODUCTION

This essay is critically examining and arguing how ideas can lead to a fundamental transformation of policy by pinpointing on the United Kingdom Asylum Policy as a case study. The concept of asylum has long stood at the crossroads of humanitarian obligation, international law, and political discretion (Gil-Bazo, 2015). Associated with the principles of protecting individuals fleeing persecution, the term asylum encompasses a range of legal, ethical, and social dimensions that vary across national and regional contexts (Convention Relating to the Status of Refugee 1951). To begin with, the essay will examine the meaning of ideas and fundamental change and their implications towards public policy. Then, it applies the relevant theoretical framework by drawing on Hall's Paradigm from the social constructive perspective which features how shared ideas and interpretive frameworks guide shifts in policy thinking. This essay situates asylum policy within evolving moral and legal narratives looking at moral and legal duties towards asylum seekers. Finally, trace the ideational shift in UK asylum system from humanitarian duty to border and deterrence control and assess its limitations by building on examples.

DEFINITION OF KEY TERMS

Arguably, the terms "new idea" and "fundamental change" have been crucial in policy, settings, debates. Political theorist like Schmidt (2010) explicitly argues that public policy does not change as result of the problem that came into existence, it changes when "new ideas" rethink existing problems, reorganize political priorities, and reframe legitimate responses.

New Ideas

The concept of "new ideas" implies to clear ways of interpreting social Phenomena (Hall,1993). Also, Campbell (2002) defines new ideas as cognitive paradigms, normative frameworks, frame, and programmatic ideas that shape how policymakers perceive, interpret, and respond to policy problem. He stresses that ideas can constrain or enable action depending on how they are included in the discourse, institutions and actor processes. In other words, shapes how problems are

comprehended and the type of remedies considered legal. According to Cairney (2020b) ideas can be understood in general sense as collective beliefs, perspectives, or modes of thought. When we move past this broad understanding and seek a more precise definition, different, approaches emerge (Cairney, 2020b). Some view ideas as normative beliefs, which concern judgements about what is right or wrong, while others see them as empirical beliefs which focus on explanations of what influences change or causes certain outcome.

Fundamental change

The concept of “Fundamental change” according to Hall (1993) connotes third other or paradigmatic change in the policy rather than incremental change. In other words, a large, rapid shift in policy agendas, institutions, or public attention (Baumgartner and Jones 1993). These are moments when issues that had been laying quiet, suddenly make it way into political discussions, with media, public are all involved. For instance, political discourses surrounding asylum policy are often activated by specific events such as high-profile Channel crossings, humanitarian crises like the Syrian refugee influx, or intense media coverage of border enforcement which reframe public narratives and influences policy responses (Mayblin et al, 2024).

Prior to this, the UK asylum system was characterized by humanitarian obligation associated with the 1951 Geneva Convention (Migration Observatory 2024). Building on this conceptual framing, the discussion turns to the 1951 Refugee Convention, whose outcome established a binding humanitarian obligation for signatory states to protect individuals fleeing persecution, affirming asylum as a legal right grounded in moral duty rather than discretionary charity. Conversely, from the 1990s till now, a different lens has emerged projecting asylum not as humanitarian obligation but as threat to border integrity and national welfare (Castles, 2003). Innes (2010) states that asylum seekers constantly appear in political discourse as “illegal migrants”, “welfare dependents” or potential security threats. As a result of this policies such as the Immigration and Asylum Act 1999, the Nationality, Immigration and Asylum Act 2002, and the Nationality and Borders Acts 2022 that has stringent instruments such as detention, dispersal and offshoring has been legislated (UK Government 1999; UK Government 2002; UK Government 2022).

Considering Hall's (1993) framework, these changes constitute a third-order change in UK asylum policy. The dominant idea of asylum has been redefined from humanitarian principle to protection of borders and identity. The essay will now set out the relevant theoretical lens, trace the ideational shift in UK asylum system and assess its limitations by drawing on examples.

METHDOLOGY

Hall' policy paradigm coherently depicts how shifts in ideas underpin major changes in asylum policy direction. Building on the works of Kunh's theory, developed in the Structure of Scientific Revolutions (1962), explains how dominant frameworks of thought guide inquiry until differences prompt a paradigm shift a fundamental change in underlying assumptions. Hall (1993) adapts this concept to social policy, arguing that policy paradigms similarly structure how problems are defined and solutions conceived, with shifts reflecting broader ideological transformations rather than purely technical adjustments. He argues that policymaker's function within a "policy paradigm". This is evident in the lines:

> Policy makers customarily work within a framework of ideas and standards that specifies not only the goals of the policy and the kind of instruments that can be used to attain them, but also the very nature of the problems they are meant to be addressing. Like a Gestalt, this frame is embedded in the very terminology through which policymakers communicate about their work, and it is influential precisely because so much of it is taken for granted and unamenable to scrutiny as a whole. I am going to call this interpretive framework a policy paradigm (p. 279).

The above extract paints a picture that paradigms are not just technical toolkits, they also influence how problem are seen, what alternatives are seen as legitimate and how actors understand their own roles. It is a worldview that guides decision making unconsciously. Hall's paradigm identifies and distinguishes between three orders of policy change which are discussed below.

Firstly, the "first-order change", refers to incremental adjustments to the settings of existing policy instruments without altering broader goals or tools of policy (Hall 1993).

First-order change involves routine, technical, and frequent adjustments to the settings of existing policy instruments. For instance, a central bank slightly changing interest rates with an established monetary policy framework (Hall 1993). In relation to asylum, an example will include increasing or decreasing number of caseworkers to reduce backlog or adjusting the waiting time targets for asylum interviews (UK Government, 2023). Note that the asylum system' s goals such protection and processing of claims are still intact, only the settings are adjusted.

Secondly, the second order- change involves changing the instruments themselves while keeping the broder goals unchanged (Hall, 1993). In other words, the second-order change connotes changing the actual policy instruments used but does so within the same overall policy goals of the existing paradigm (Hall,1993). A typical illustration will be, by adopting new forms of surveillance or new administrative mechanism. Again, linking it to asylum policy, the broder goal like providing support stays the same but tools are changed. For example, introducing digital case management systems instead of paper files to speed up decisions. Moving from cash allowances to prepaid cards for subsistence support (Home Office, 2025).

Finally, the most dramatic or intense form of change, the third-order change takes place when goals, instrument and settings all shift simultaneously leading to a paradigm shift. Third order of change is arguing that only the most radical kind constitutes a paradigm shift. Hall (1993) proves that a major change in ideas (third order) are not exclusively technical but involves public debate and media influence, contest for authority (for instance between experts and politicians) and electoral competition. In his work, he cited the example of Thatcher's adoption of monetarism (stabilizing inflation) from Keynesianism (reducing unemployment through fiscal activism) (Hall, 1993). In the case of asylum, it not just a technical adjustment but a fundamental change on what asylum policy is. Notably moving from right or humanitarian paradigm to deterrence (Home Office, 2023). In other words, from protection under international law to asylum as a migration issue. Also, from policies of integration and protection to safe third country (Home Office, 2023). This changes the goal, instruments and the settings as well.

Compared to incrementalist theories like Multiple Streams Analysis (Cairney, 2020b; Kingdon's, 1984), which perceived policymaking as a process of gradual adaptation.

Hall's (1993) third- order of change model constructively explains period of radical and transformational change. It entails a transformation in the ideas and goals that define the purpose of the policy. It supports with constructivist and ideational approaches to policymaking (Beland, 2019; Schmidt 2010), which argue that political actors interpret the world through shared cognitive and normative frameworks rather than objective evidence alone.

Hall's third-order change is also powerful because it reflects in how UK asylum goals shifted from protection to deterrence (Hall,1993). In post-war period, asylum was framed within a humanitarian paradigm guided by the 1951 Refugee Convention and the moral imperative of protection after World War II (Castles, 2003; Betts, 2014). During this era the policy goals centred on safeguarding vulnerable individuals from persecution, with legitimacy drawn from international law and moral perspectives. International obligations were fully respected and taken into consideration. Conversely, from the 1990s onwards, this humanitarian logic was radically replaced to control paradigm that redefined asylum as a problem of immigration management, welfare dependency, and security (Boswell, 2007). Legislative reforms such as the Immigration and Asylum Acts (1999, 2002) and the Nationality and Borders Act (2022) exemplifies this reorientation. The central objective became to deter "unfounded" claims, restrict access to welfare, and externalise asylum processing.

In nexus to Hall's (1993) model, this transition reflects a third-order change. The goals were redefined from protection to prevention. His framework underscores how this not merely a practical adjustment but cognitive and normative shift in what policymakers viewed as legitimate and achievable.

However, while Hall's Paradigm (1993) remains influential, it also faced a lot of several criticism. To begin with, scholars argue that Hall underplays the political and institutional processes through which idea are contested and adopted (Beland and Cox, 2011). Paradigm shifts rarely occur in the clean, revolutionary fashion, instead they evolve through layering, bricolage, and translation across political arenas. In the UK context, policy change often unfolds through layering, where new institutional or ideological elements are added to existing frameworks, bricolage, the creative recombination of established policy tools to address emerging challenges and translation, the adaptation of ideas across political arenas to fit differing institutional

and discursive contexts (Schmidt, 2010). In the UK asylum case, for instance, the humanitarian and control paradigms coexist and overlap. Humanitarian rhetoric is often invoked even as deterrent measures intensify. This rhetoric is frequently invoked by political actors opposing expansive asylum protections, often framing migration as a challenge to national sovereignty or public security (Huysmans, 2006). For instance, UK government discourse surrounding the 2022 Nationality and Borders Act employed such rhetoric to justify deterrence- based measures, portraying asylum seekers arriving via irregular routes as undermining legal order and fairness. (Sandberg, Schultz, Syppli Kohl, 2025)

Secondly, according to Baumgartner and Jones (1993) it has the potential to risk elite bias. It emphasises policymakers' cognitive learning but neglects the role of civil society, media, and advocacy in shaping dominant ideas (Baumgartner and Jones, 1993). Asylum policy is deeply politicised, and paradigm change is as much about discursive struggle (the contestation of ideas among political actors seeking to shape policy meaning) and public framing as bureaucratic learning (the process through which institution reinterpret these discourses to adjust administrative practices and policy implementation). (Schmidt 2010).

Regardless of these criticism, Hall's paradigm theory (1993) remains invaluable for explaining why policy direction changes at all and why governments reorient from one set of core assumptions to another. While its limitations especially, political contestation and risks elite bias (Beland and Cox, 2011), it remains a cornerstone of critically analysing how the UK's asylum policy direction has moved from humanitarian protection to deterrence and control. These limitations may affect the strength of analysis by restricting the generalisability of findings and potentially reinforcing the very elite perspectives the essay critiques. Recognising this helps situate the argument as interpretive rather than exhaustive, clarifying that its insights apply primarily within the UK's Political and media context.

KEY RESULTS

The trajectory of United Kingdom asylum since the 1990s clearly shows a sustained change in policy direction. That is from a post war, protection-oriented approach

towards a control and deterrence paradigm (The Immigration and Asylum Act 1999, The Nationality, Immigration and Asylum Act 2002, The Nationality and Borders Act 2022) (Gammeltoft and Tan, 2018). This was not a single legislative act, but an extended process of reframing what asylum was understood to be from a humanitarian obligation under the 1951 Refugee Convention (Gil-bazo, 2015) to a problem of immigration control, fiscal burden and national security (Innes, 2010). Hall's (1993) concept of third order (paradigmatic) change throw lights on how new dominant ideas redefined policy goals and legitimised new instruments.

In the immediate post-war decades and into the late twentieth century, asylum in the UK was principally conceptualised within a humanitarian framework tied to international legal commitments (Convention Relating to the Status of Refugee 1951). This framing emphasized right-based protection for those fleeing persecution and located legitimacy in the moral and legal duty to admit refugees. Scholarly accounts emphasise how refugee protection occupied a normative centre in policy debates for much for twentieth century (Nantermoz, 2020, McClean 2025).

However, two contextual pressures contributed to ideological change. Firstly, sharp rising in asylum applications in the 1990s and the visible media coverage of arrivals which triggered political salience and public anxiety (Anderson 2017). Second, growing concerns about welfare costs, administrative backlogs and perceived "abuse" of the system provided fertile ground for alternative problem definitions that emphasized control and gatekeeping (Ramachandran, 2024). Official responses began to prioritise speed and deterrence to deal with a caseload that government presented as unsustainable (Anderson 2017). Parliamentary and government material from this period show a policy turn towards faster decision-making and tougher administrative arrangements (Laird 2008).

In the same trend, from the late 1990s, academic observers like (Crawley 2005: Schmidt 2010) charted a discursive shift. Asylum seekers were increasingly framed in a political rhetoric as "illegal migrants", "bogus claimants" (Innes, 2010) or potential security risks rather than primarily as victims entitled to protection (Boswell; 2007). This discursive shift reframed the goal of asylum policy from safeguarding refugees to managing migration flows and protecting borders thereby legitimising new, coercive instruments.

As a result of this, several legislative steps institutionalised the control paradigm. The Immigration and Asylum Act (1999) introduced mechanisms intended to speed up case processing and manage claimant accommodation (dispersal) which is an important administrative turning point. It makes it cumbersome for asylum seekers to work legally while waiting for a decision, for many into poverty and control. It expanded the use of detentions for asylum seekers including those who claims were pending or being appealed, often without clear limits. It makes the asylum process uncertain for claimants.

Next, the Nationality, Immigration and Asylum Act (2002) expanded administrative powers, tightened appeal procedures and created new offences related to immigration facilitation, reflecting growing emphasis on deterrence and enforcement. Equally so, tougher detention powers, including before their claims were fully processed or while awaiting removal (Immigration and Asylum Act, 2002). Removal or withdrawal of support was also introduced. Also, people from certain countries labelled “safe” could be removed before their appeal was heard (non-suspensive appeal) (Immigration and Asylum Act, 2002). The shift from humanitarian duty to border enforcement reflects a gradual removal or withdrawal of support for asylum seekers as state actors reframe protection as conditional and contingent on compliance with increasingly restrictive criteria.

Under the UK’s non-appeal process, asylum seekers from officially designated as safe can be removed from the UK before their appeal is heard. This means that, if their claim is rejected, they may be deported while their appeal proceeds from outside the country. The policy assumes that individuals from these countries are unlikely to face persecution, thereby limiting their access to in-country appeal rights. Conversely, those from unsafe countries are granted in country appeal rights reflecting an uneven system of protection that privileges geopolitical assumptions over individual circumstances.

Most consequentially, the Nationality and Borders Act (2022) further embedded deterrence principles by changing the legal treatment of people arriving irregularly and expanding powers for returns and offshore processing. It makes it a criminal offence to arrive in the UK without permission, even to claim asylum. People can face up to 4 years in prison simply for entering irregularly. This permission is typically granted temporarily during asylum process, allowing applicants access to limited support and protection from removal. Those who do not apply or whose claims are refused, may

lose their permission and become subject to detention or deportation under immigration enforcement procedures. The act further criminalises the act of entering irregularly for example, crossing the channel in small boats or arriving without valid documents or visas and expanding the state's power to detain and remove asylum seekers.

DISCUSSION

Hall's (1993) paradigm concept explains this transformation. As the dominant idea of what asylum is changed, so too did the criteria for legitimate action enabling measures task would previously have been seen as unacceptable (e.g., wider use of detention curtailed appeal rights externalisation of processing). The recurring reforms to UK asylum policy such as successive legislative amendments and procedural tightening were not simply administrative adjustments but reflected a redefinition of broader policy goals, shaping the instruments chosen to manage migration and border control. A clear point to note here is that the Nationality and Borders Act of (2022) was designed to deter illegal entry and strengthen border control. The act explicitly sought to disincentivise unauthorised routes and introduced new thresholds and restrictions that critics argue curtail protection (and which UNHCR and NGOs have publicly challenged by condemning it as inconsistent with the 1951 convention (Amnesty International 2022). These Acts collectively mark institutional codification of the control paradigm. Critics like Morris, (2022), Webber, (2023) and Haynes (2023) argue that the 2022 Acts represents the fullest expression of the deterrence paradigm. Protection obligations are subordinated to the logic of border enforcement. This corresponds to Hall's (1993) highest order of change a complete reorientation of policy goals and moral justification.

By 2022, detention, dispersal, and restricted support for asylum seekers were not exceptions but standard administrative practices for asylum seekers. Pierson's (1997) analysis of increasing returns and path dependence explains why initial instruments changes (faster procedures, dispersal, new offences generate self-reinforcing effects like administrative routines vested interests, budgetary allocations and internal

expertise that make reversal difficult). Once a particular set of tools such as deterrence-based measure or restrictive asylum procedures is adopted, the cost of

reversal increase, leading institutions to continue along the same trajectory even when alternative approaches might be more effective. The 1999 and 2002 Acts and subsequent measures reconfigured incentives and capacities within the Home Office and local delivery systems. Over time these reforms hardened the control paradigm into institutional practice (Slaven, Casella Colombeau and Badenhoop, 2020). This Hall's third-order change is realised through many second-order and first-order moves that cumulatively, create a new status quo.

Again, electoral competition encouraged increasingly tough rhetoric on asylum (Geddes and Scholten, 2016). Research indicates that restrictive asylum policies are closely linked to populist appeals to sovereignty and national identity (Geddes and Scholten, 2016). Political incentives helped sustain restrictive approaches across party lines, making asylum restriction as durable policy with the broader landscape of UK migration politics.

Despite the dominance of control paradigm, it faces legal and moral challenges (Kent, 2023). UK courts have occasionally constrained executive power, for example in rulings on detention length and procedural fairness (Kent, 2023). NGOs, legal professionals, and local authorities often operate under a residual humanitarian ethos, maintaining aspects of the older paradigm (Mayblin, 2019).

Furthermore, evidence suggests deterrence has not reduced asylum demand but has increased administrative backlogs and costs (Home Office, 2023). Hotel accommodation expenses exceeding £8 million per day in 2023 (NAO, 2023), indicate that restrictive measure may undermine efficiency while exacerbating humanitarian concerns. Thus, the asylum paradigm continues not because it is effective but because it is ideologically and institutionally kept firmly across party politics (Hall, 1993). This conclusion is consistent with Hall's (1993) insight that ideas, once institutionalised, structure policy through practice.

CONCLUSION

Conclusively, the UK asylum policy case study shows how ideas can lead to paradigmatic change (Hall, 1993). The humanitarian paradigm anchored on international law, and moral duty (Gil-bazo, 2015) was radically displaced by a deterrence paradigm emphasising sovereignty, efficiency, and control (Gammeltoft and Tan, 2018). Hall's (1993) framework reveals that policy direction change not simply due to administrative necessity but because policymakers redefined what counted as legitimate governance. When asylum came to be seen as a problem of control rather care, new instruments and institutional routines followed naturally. The Nationality and Borders Act (2022) marks the culmination of this ideational trajectory a clear example of paradigm shift

REFERENCES

Anderson, B. J.(2017) 'The Politics of Pests: Immigration and the Invasive Other. Social Research: *An International Quarterly*, 84(1), pp. 7- 28. Available at: https://muse.jhu.edu/article/659220 (Accessed: 10 November 2025).

Amnesty International UK (2022) *Nationality and Borders Bill: House of Lords Report, 28 February & 2 & 8 March 2022*. London: Amnesty International UK. Available at: https://www.amnesty.org.uk/resources/nationality-and-borders-bill-session-2021-2022 (Accessed: 21 November 2025).

Baumgartner, F. R. and Jones, B. D. (1993). *Agendas and Instability in America Politics*. Chicago: University of Chicago Press.

Béland, D. (2019) *How Ideas and Institutions Shape the Politics of Public Policy*. Cambridge: Cambridge University Press (Elements in Public Policy).

Béland, Daniel, and Robert Henry Cox (eds), (2011) *Ideas and Politics in Social Science Research* (2010; online edn, Oxford Academic, Available at: https://doi.org/10.1093/acprof:oso/9780199736430.001.0001, (Accessed: 12 November 2025).

Betts, A., (2014) The global governance of crisis migration. In *Humanitarian crises and migration* pp. 349-367. Routledge.

Boswell, C., (2007) 'Migration control in Europe after 9/11: Explaining the absence of securitization'. *JCMS: Journal of Common Market Studies*, *45*(3), pp.589-610.

Cairney, P. (2020b) 'Chapter 11 Ideas and Multiple Streams Analysis', in *Understanding public policy: theories and issues*. 2nd edition. London: Red Globe Press. Available at: https://bibliu.com/users/saml/samlUniversityOfBristol?RelayState=eyJjdXN0b21fbGF1bmNoX3VybCI6IiMvdmlldy9ib29rcy85NzgxMTM3NTQ1MjEzL2VwdWIvT0VCUFMvMDdfdG9jLXRpdGxlLmh0bWwifQ%3D%3D. (Accessed: 7 November 2025).

Campbell, J. L. (2002). Ideas, Politics, and Public Policy. *Annual Review of Sociology*, *2John L. Campbell (2002) 'Ideas, Politics, and Public Policy', Annual Review of Sociology, 28 (2002), pp. 21–38. Available at: https://www-jstor-org.bris.idm.oclc.org/stable/3069233.8*, 21–38. http://www.jstor.org/stable/3069233 (Accessed:9 November 2025).

Castles, S. (2003) 'Towards a Sociology of Forced Migration and Social Transformation', *Sociology*, *37*(1), 13-34. https://doi.org/10.1177/0038038503037001384.

Crawley, H. (2005), *Evidence on Attitudes to Asylum and Immigration: What We Know, Don't Know and Need to Know*. COMPAS Working Papers, University of Oxford.https://www.compas.ox.ac.uk/2005/wp-2005-023-crawley_attitudes_asylum_immigration/ (Accessed: 5 November 2025).

Gammeltoft-Hansen, T., & Tan, N. F. (2018). The End of the Deterrence Paradigm? Future Directions for Global Refugee Policy. *Journal on Migration and Human Security*, *5*(1), 28-56. https://doi.org/10.1177/233150241700500103 (Original work published 2017)

Gil-Bazo., M. T. (2015) 'Asylum as a General Principle of International Law, *International Journal of Refugee Law*, 27, (1), pp 3–28, Available at: https://doi.org/10.1093/ijrl/eeu062

Geddes, A. and Scholten, P., 2016. The Politics of Migration and Immigration in Europe. Second Edition55 City Road, London: SAGE Publications Ltd. Available at: https://doi.org/10.4135/9781473982703 (Accessed: 21 November 2025).

Hall, P. A (1993). Policy paradigms, social learning and the states: The case of economic policymaking in Britian, Comparative politics 25(3) pp. 275-296

Haynes, J. (2023). Human Trafficking: Iconic Victims, Folk Devils and the Nationality and Borders Act 2022. *Modern Law Review*, *86*(5), 1232-1264. Available at: https://doi.org/10.1111/1468-2230.12814.

Home Office (2023). Immigration Statistics Quarterly Release, March 2023. London: UK Government. Available at:

https://www.gov.uk/government/collections/immigration-statistics-quarterly-release (Accessed: 7 November 2025).

Home Office (2025) *Asylum support (asylum instructions)*. GOV.UK. Available at: https://www.gov.uk/government/collections/asylum-support-asylum-instructions (Accessed: 17 November 2025).

Home Office (2023) *UK-Rwanda treaty: provision of an asylum partnership*. GOV.UK. Available at: https://www.gov.uk/government/publications/uk-rwanda-treaty-provision-of-an-asylum-partnership (Accessed: 22 November 2025).

Home Office (2025) *Asylum claims in detention: caseworker guidance*. GOV.UK. Available at: https://www.gov.uk/government/publications/asylum-claims-in-detention. (Accessed: 7 November 2025).

Huysmans, J. (2006) *The politics of insecurity: fear, migration and asylum in the EU*. London: Routledge.

Innes, A. J. (2010). When the Threatened Become the Threat: The Construction of Asylum Seekers in British Media Narratives. *International Relations*, *24*(4), 456-477. https://doi.org/10.1177/0047117810385882 .

Kent, J. (2023). Liberal democracies and asylum: legal transformation and implementation challenges. *International Relations*, *39*(4), 511-538. https://doi.org/10.1177/00471178231197757 (Original work published 2025)

King, A. (1985) 'John W. Kingdon, Agendas, Alternatives, and Public Policies, Boston: Little, Brown, 1984, xi + 240 pp., $9.95.', *Journal of Public Policy*, 5(2), pp. 281–283. Available at: doi:10.1017/S0143814X00003068.

Kay, Adrian. (2005). A Critique of the Use of Path Dependency in Policy Studies. Public Administration. 83. 553 - 571. Available at:10.1111/j.0033-3298.2005.00462.x.

Laird, S. E. (2008). Racism and Ethnic Minorities. In: *Anti-Oppressive Social Work: A Guide for Developing Cultural Competence*. London: SAGE Publications Ltd. pp. 1-19 Available at: https://doi.org/10.4135/9781446269473.n1 (Accessed: 9 Nov 2025).

Mayblin, L. *et al.* (2024) ''Bringing order to the border': liberal and illiberal fantasies of border control in the English Channel', *Journal of Ethnic and Migration Studies*, 50(16), pp. 3894–3912. doi: 10.1080/1369183X.2024.2349691

McClean., A. Y. (2024) ‘Preserving Law and Order: How Institutions Implementing International Norms on Refugee Protection Can Restrict Asylum Outcomes’, *International Political Sociology*, 18 (1) pp 1-21. Available at: https://doi.org/10.1093/ips/olae001

Migration Observatory (2024) UK Public attitudes toward immigration. Oxford: The Migration Observatory.

Mahoney, J. and Thelen, K. (eds.) (2009) *Explaining Institutional Change: Ambiguity, Agency, and Power*. Cambridge: Cambridge University Press.

National Audit Office (NAO) (2023). Asylum and Accommodation: Value for Money Report. London: NAO. Available at: https://www.nao.org.uk/reports/investigation-into-asylum-accommodation/ (Accessed: 9 November 2025).

Nantermoz, O. (2020) ‘International refugee protection and the primary institutions of international society’, *Review of International Studies*, 46(2), pp. 256–277. Available at: doi:10.1017/S0260210520000029.

Pierson, P. (1997) *Increasing returns, path dependence and the study of Politics*. Jean Monnet Chair Papers, 320. Florence: European University Institute, Robert Schuman Centre

Ramachandran, N. (2024) ‘The Enforced Destitution of Asylum Seekers in the UK’. *J. Hum. Rights Soc. Work* 9, 139–153 Available at: https://doi.org/10.1007/s41134-024-00300-8

Ria Kapoor, Asylum After Empire: Colonial Legacies in the Politics of Seeking Asylum. By Lucy Mayblin, *Journal of Refugee Studies*, Volume 32, (2), Pages 342–345, Available at: https://doi.org/10.1093/jrs/fey067

Schmidt, V.A. (2010) ‘Taking ideas and discourse seriously: explaining change through discursive institutionalism as the fourth “new institutionalism”’, *European Political Science Review*, 2(1), pp. 1–25. Available at: doi:10.1017/S175577390999021X.

Sandberg, M., Schultz, J. and Syppli Kohl, K. (2025) ‘The temporary turn in asylum: a new agenda for researching the politics of deterrence in practice’, *Journal of Ethnic*

and Migration Studies, 51(8), pp. 1997–2014. Available at: doi: 10.1080/1369183X.2024.2441594.

Slaven, M., Casella Colombeau, S., & Badenhoop, E. (2020). What Drives the Immigration-Welfare Policy Link? Comparing Germany, France and the United Kingdom. *Comparative Political Studies*, *54*(5), 855-888. https://doi.org/10.1177/0010414020957674.

UN General Assembly, (1951) '1951 Convention Relating to the Status of Refugees, United Nations, Treaty Series', 189,(28) pp 137, available at: https://www.refworld.org/legal/agreements/unga/1951/en/39821 (Accessed: 09 November 2025)

UK Government (1999) Immigration and Asylum Act 1999. Available at: https://www.legislation.gov.uk › ukpga › 1999 › contents. (Accessed 23 November 2025)

UK Government (2002) Nationality Immigration and Asylum Act 2002. Available at: https://www.legislation.gov.uk › ukpga › 2002 › contents (Accessed: 23 November 2025)

UK Government (2022) Nationality and Borders Act 2022. Available at: https://www.legislation.gov.uk › ukpga › 2022 (Accessed :23 November 2025).

UK Government (2023) *The Asylum Support (Amendment) Regulations 2023*. London: The Stationery Office. Available at: https://www.legislation.gov.uk/uksi/2023/1372/pdfs/uksiem_20231372_en_001.pdf (Accessed: 20 November 2025).

Webber, F. (2023). *Borderline Justice: The Fight for Refugee and Migrant Rights*. London. Pluto Press.

Authors Bio

My name is Foday Kabia, and I am currently completing a Master of Science in Public Policy at the University of Bristol, alongside a Master of Science in Human Rights and Global Ethics at the University of Leicester. Throughout my academic and professional journey, I have become deeply committed to understanding how public policy, human rights, and social inequalities intersect particularly in the lives of migrants, asylum seekers, and marginalised communities. My work with Bristol Mind and Aid Box Community has strengthened this commitment, giving me first-hand insight into how political narratives and policy decisions shape people's lived experiences. I believe that rigorous, justice focused research is essential for building fairer and more inclusive societies. Exploring issues of migration, race, and policy is not only intellectually meaningful to me, it aligns with my long standing aspiration to become a researcher, lecturer, and professor dedicated to producing knowledge that drives positive social change.

The Effect of Women's Political Empowerment on Environmental Performance

Author: Theadora Gallie

ORCID: 0009-0000-3669-4946

Programme: BScSocial Policy and Politics

Affiliation: School for Policy Studies

Email: thea.gallie.2023@bristol.ac.uk

Abstract

Calls for environmental policy solutions are continually increasing, with academics branching away from economic and institutional factors to consider alternative factors driving environmental success. This paper contributes to the growing literature supporting the role of Women's Political Empowerment (WPE) for advancing global environmental performance, investigating the relationship between these measures across countries globally. Secondary data sourced from the Quality of Government 2025 Standard dataset is utilised to test for a positive causal relationship between Women's Political Empowerment and environmental performance. Using multiple linear regression, logistic regression, and interactive regression models, the complex and nonlinear effects of WPE on a global level are highlighted. Linear regressions identify a nonlinear relationship, with significant positive effect exerted only when WPE is high. Also, interactive models display that the effect of WPE is dependent on high levels of GDP per capita. Lastly, multivariate logistic regression affirms that WPE elevates the probability of high climate change mitigation. The findings demonstrate that high levels of WPE, when GDP per capita is not a limiting factor, do promote environmental performance. This paper supports suggestions that policy makers must consider the role of WPE in global environmental action.

Keywords: Women's Political Empowerment, Environmental Performance, Climate Change Mitigation, Climate Crisis

Introduction

This paper investigates the relationship between Women's Political Empowerment (WPE) and environmental performance globally. The rapidly developing environmental crisis has sparked widespread discussion concerning how to reduce the negative impacts of environmental damage (Winkler *et al*., 2007; Sowah and Kirikkaleli, 2022; Zhang *et al*., 2022). Subsequently, the main hypothesis that increased WPE has a positive causal effect on environmental performance, is tested here. Linear regressions, used to estimate how much one variable (environmental performance) changes with a one-unit increase in another (WPE), identify a nonlinear relationship, with significant positive effect exerted only when WPE is high. Interactive models, which explore if the effect of this independent variable (WPE) is influenced by the value of another (GDP per capita), aid evaluation of the second hypothesis that the effect of WPE is dependent on high levels of GDP per capita. The analysis confirms that low levels of WPE or GDP per capita undermine the positive impact of WPE. Lastly, the third hypothesis that increased WPE elevates the probability of high climate change mitigation (one facet of environmental performance) is affirmed using multivariate logistic regression (used to predict the likelihood of a binary outcome based on multiple factors). Throughout, potential confounders including socio-economic and institutional factors are controlled for to ensure results are not misleading. The analysis builds on an increasing literature focused on the factors aiding environmental performance and global climate action (Gallego-Alvarez *et al.*, 2014; Wendling *et al.*, 2022; Handoyo, 2024). It is also prompted by discussion over the ranging policy impacts of WPE (De Siano and Chiariello, 2021; Asongu, Messono and Guttemberg, 2022; Dahlum, Knutsen and Mechkova, 2022). This paper supports suggestions that policy makers must consider the role of WPE in global environmental action.

Theory and literature

Despite the expandingliterature, the lack of a clear consistent definition of WPE hinders the progressionof empirical research. Alexander, Bolzendahl and Jalalzai (2016) illustrate this issue, and conceptualise WPE as enabling women's abilities, achievements and resources to promote political equality and influence. Through this lens WPE incorporates the complex formal and informal pathways of political influence

(Alexander, Bolzendahl and Jalalzai, 2016). This perspective compliments that of Sundstrom *et al*. (2017), who present three categories concerning the civil liberties, civil society participation, and political representation of women to account for WPE's multifaceted nature. This conceptualisation of WPE underpins the WPE index presented by Sundstrom *et al*. (2017).

The previous literature examining potential relationships between women's representation or leadership, and environmental policy and performance point to the disproportionate negative impacts of climate change and pollution on women. Repeatedly, evidence depicts this in relation to women's health both mental and physical (Stone, Blinn and Spencer, 2022; Sinaga and Harahap, 2023; Pratama *et al*., 2024; Anastasia and Nabilla, 2025), whilst also signifying that gendered impacts arise from pre-existing inequalities and patriarchal social norms (Nguyen, 2019; Fruttero *et al*., 2024).

Additionally, ecofeminist perspectives underpin this research, signifying the link between the domination of women and that of nature (Mies and Shiva, 1993; Chen, 2014). Mies and Shiva (1993) argue that the liberation of women and the environmental movement are interconnected and form part of a greater consensus to fight to protect life on earth. By examining the complex and connected facets of patriarchy and the environmental crisis, the need for further research investigating the potential for WPE to strengthen environmental policy is made clear.

Scholars have examined many possible explanatory factors for the variation in environmental performance. Wendling *et al*. (2022) conducts multivariate regressions of environmental performance on 11 variables whilst controlling for GDP per capita, demonstrating the positive significant impact of government effectiveness, control of corruption, rule of law and regulation quality. Yet similarly to many other studies (Gallego-Alvarez *et al*., 2014; Boleti *et al*., 2021; Handoyo, 2024), variables concerning WPE remain excluded. These studies do however provide guidance for the selection of control variables when assessing the effects of WPE.

A few relatively recent studies investigate the effects of WPE on environmental issues, providing evidence of a causal relationship. However, these studies often focus on impacts upon CO^2 emissions per capita (Lv and Deng, 2019; Ganda, 2024). Lv and Deng (2019) and Ganda (2024) illustrate the potential of WPE as a tool for improving

CO^2 emissions regionally and globally. Asongu, Ndour and Ngoungou (2024) extend analysis to assess the impact of WPE on environmental performance more broadly, perhaps initialising a shift away from CO^2 emissions as the dominant outcome measure. However, the sample is limited to African countries, limiting the generalisability of results beyond this region. This paper builds on the evidence presented by Asongu, Ndour and Ngoungou (2024), by testing whether the significant causal effect of WPE on environmental performance remain evident with a globally representative sample.

Methods

This research utilises the cross-sectional Quality of Government 2025 Standard dataset, which includes data from the year 2021 or surrounding years up to three years difference, with country as the unit of analysis (Teorell *et al*., 2025). The initial sample size for simple linear regression is 166. Table 1 presents summary statistics of the relevant variables.

Table 1. Summary Statistics

Variable Name	Code	Variable Type	Count Nonmissing	Mean/Median/ Mode	Standard dev	Range
Environmental performance	epi_epi	Scale	179	46.989	11.470	24.6 - 75.7
Climate change mitigation	epi_cchbinary	Binary	179	High	-	-
Women's political empowerment	vdem_gender	Scale	172	0.730	0.190	0.065 - 0.959
Logarithmic transformation of Population size	Lnpop	Scale	192	15.636	2.235	9.324 - 21.069
Urban population size	wdi_popurb	Scale	192	59.744	23.050	13.456 - 100
Logarithmic transformation of GDP per capita	Lngdpcap	Scale	187	8.688	1.412	5.577 - 12.229
Democracy status	bmr_dem	Nominal binary	193	Country is democratic	-	-
Corruption indicator	bci_bci	Scale	188	45.006	18.717	0.050 - 78.942
Government effectiveness	wbgi_gee	Scale	193	-0.058	0.996	-2.396 - 2.248
Average years of schooling	opri_yearschoolt	Scale	140	9.224	3.282	1.246 - 14.26
Conflict intensity	bti_ci	Scale	135	4.867	2.205	1 - 10

Note. For continuous variables the mean is shown, for nominal variables the mode is shown.

The Environmental Performance Index (EPI) (Block *et al*., 2024) forms the dependent variable for the linear regressions, chosen due to the multifaceted incorporation of environmental health (25%), ecosystem vitality (45%), and climate change (30%). The 0-100 scale represents 58 indicators related to issues such as mitigation, air quality, biodiversity and air pollution.

A binary measure of climate change mitigation forms the dependent variable for logistic regressions. This variable was generated from a continuous measure (epi_cch) which constitutes 30% of the EPI (Block *et al*., 2024). It indicates the effectiveness of a country's mitigation efforts in relation to historical emissions, encompassing annual emission growth rates, target proximity, net carbon fluxes, and projected emissions (Block *et al*., 2024, pp. 24-30). A binary measure was generated by recoding the variable into two levels to note high or low climate change mitigation with the median (41.8) as the benchmark.

The 0-1 scale WPE index (Coppedge *et al*., 2023a; Pemstein *et al*., 2023) constitutes the main independent variable based on the previously discussed conceptualisation of WPE. The index incorporates women's civil liberties, civil society participation, and political participation (Coppedge *et al*., 2023b). This choice reflects that women's political influence over EPI is not limited to formal methods such as parliamentary representation. The square of WPE (WPE^2) is also incorporated due to nonlinearity between WPE and EPI discovered when generating a two-way scatterplot.

Scholars indicate that socio-economic factors influence environmental performance (Boleti *et al*., 2021; Wendling *et al*., 2022). Therefore, continuous measures of population size, urban population percentage, and GDP per capita are included as controls (World Bank, 2024). Both population size and GDP per capita underwent logarithmic transformation due to high positive skewness. Additionally, the influence of institutional factors is evident (Wendling *et al*., 2022; Handoyo, 2024). A binary measure of democracy (Miller, Boix and Rosato, 2022) accounts for the predicted positive effect of democracy on both WPE and EPI. A continuous corruption measure is incorporated due to possible impacts on environmental policy efficiency and accountability (Standaert, 2015). A continuous government effectiveness score controls for impact on the success of policy development and implementation (Kaufmann and Kraay, 2024).

Two additional controls, a continuous variable for average schooling years (UNESCO, 2024) and a categorical index for conflict intensity (bti_ci) (Donner, Hartmann and Steinkamp, 2024) are treated cautiously due to limited sample sizes. As conflict intensity is measured on a 1-

10 scale, it is treated as continuous for ease of interpretation. Multiple imputations were conducted on average schooling years, conflict intensity, corruption, and log of GDP per capita to counteract sample size limitations. Yet the generalisability of results must be assessed carefully. The predictors selected were continuous measures of government effectiveness, political stability (wbgi_pve) (Kaufmann and Kraay, 2024) and percentage of urban population (World Bank, 2024), and a categorical measure of region (ht_region) (Hadenius and Teorell, 2007; Wahman, Teorell and Hadenius, 2013).

The impacts of WPE on EPI are examined using simple and multiple linear regression models with incrementally introduced controls and WPE^2 to account for nonlinearity. No heteroskedasticity is found when plotting residuals against fitted values and requesting variance inflation factors confirms no multicollinearity (VIF <10 for all variables). To combat missing data multiple imputations are conducted, and estimations are repeated. Interactive models are generated to assess the moderating effect of GDP per capita on WPE, with predictions that WPE may have a greater effect when GDP per capita is higher due to a country's ability to invest in infrastructure, technology and other policy solutions. Marginal effects are generated to aid interpretation of the models. Lastly, bivariate and multivariate logistic regressions are conducted to explore the impact of WPE on the probability of high climate change mitigation.

Results

Table 2 presents the results of all regression models. Overall, the results of Models 1-6 suggest that WPE has a significant positive causal effect on environmental performance at higher levels of WPE, with controls accounted for. Model 1 indicates that WPE has a strong significant positive effect on environmental performance. However, Model 2 is used to reveal that there is a U-shaped relationship between WPE and environmental performance rather than a linear one. This is confirmed through creating a scatterplot with a quadratic line of best fit.

For further investigation, the marginal effects of WPE at the 10th, 25th, 50th, 75th, and 90th percentiles were produced to estimate how the impact of increasing WPE on environmental performance may differ depending on the initial level of WPE (see Appendix A). The results depict that WPE has no significant marginal effect at low levels until it reaches .637 (B = 40.701, p <.001), after which it has an increasingly significant impact. . This reaffirms that the

influence of WPE on environmental performance is nonlinear. However, the exact magnitude at higher levels of WPE cannot be estimated due to overlapping confidence intervals.

When control variables are added in Models 3 and 4, a nonlinear relationship is confirmed, and GDP per capita, corruption, and government effectiveness prove significant influences on environmental performance. Estimating marginal effects using Model 4 suggests that the influence of WPE above .417 WPE is increasingly strong, yet this moderately large and overlapping confidence intervals again undermine reliability. All things considered, at .637 WPE and above, a positive causal effect is indicated, even with fixed controls, but the exact magnitude and change in effect between levels cannot be determined.

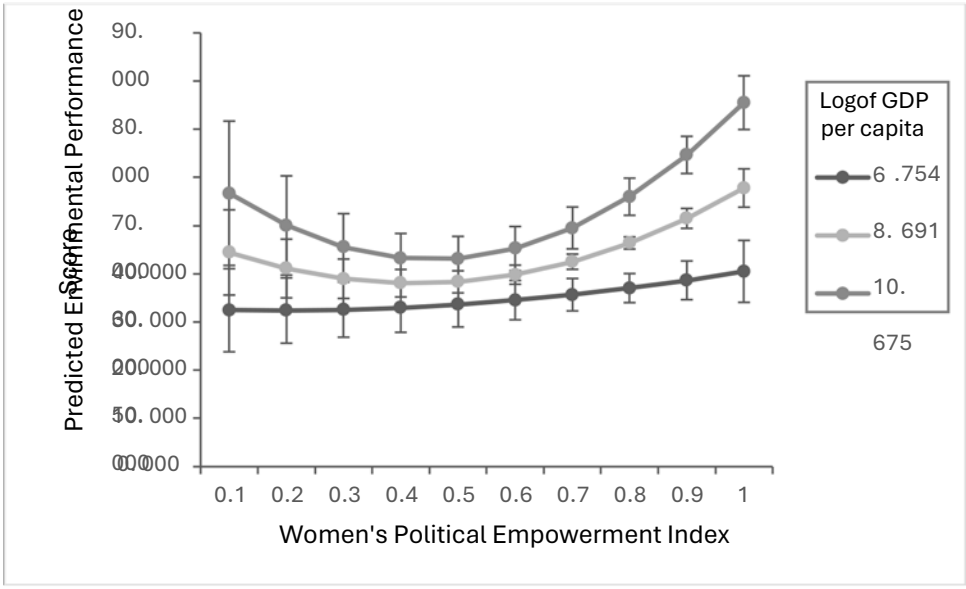
Table 2. Results of Linear Regressions

	M1	M2	M3	M4	M5	M6	M7	M8	M9
Women's Political Empowerment	40.851 (<.001)	-101.640 (<.001)	24.534 (<.001)	-50.452 (.003)	-46.195 (.012)	-36.548 (.071)	165.810 (.028)	160.324 (.035)	-47.628 (.003)
Women's Political Empowerment^2		111.728 (<.001)	-	62.777 (<.001)	58.453 (<.001)	48.883 (.005)	-159.112 (.011)	-156.277 (.014)	60.465 (<.001)
Log of population size			-0.249 (.416)	-0.170 (.555)	-0.245 (.488)	-0.587 (.258)	-	-0.353 (.205)	0.047 (.889)
Percentage of Urban population			0.001 (.988)	0.005 (.889)	-0.002 (.965)	0.005 (.907)	-	0.036 (.253)	0.014 (.670)
Log of GDP per capita			5.798 (<.001)	5.266 (<.001)	4.199 (<.001)	3.382 (.020)	8.606 (.003)	8.380 (.005)	4.153 (<.001)
The country is democratic			0.474 (.737)	-1.229 (.372)	-1.321 (.396)	-1.748 (.297)	-	-0.771 (.557)	-1.431 (.303)
Corruption indicator			-0.210 (.001)	-0.189 (.002)	-0.234 (.001)	-0.183 (.037)	-	-0.129 (.028)	-0.194 (.003)
Government effectiveness			-5.030 (.002)	-4.984 (.001)	-4.620 (.008)	-4.565 (.029)	-	-3.992 (.007)	-5.041 (.002)
Average years of schooling					0.483 (.127)	0.527 (.137)	-	-	0.480 (.113)
Conflict intensity						-0.381 (.378)	-	-	-0.275 (.437)
Interaction: WPE and log of GDP per capita							-25.708 (.008)	-24.358 (.012)	-
Interaction: WPE^2 and log of GDP per capita							25.444 (.001)	24.909 (.002)	
Intercept	16.518 (<.001)	57.215 (<.001)	-8.770 (.371)	13.572 (.192)	21.382 (.068)	30.595 (.027)	-22.318 (.310)	-14.392 (.541)	15.578 (.133)
N	166	166	160	160	125	100	163	160	165
R-squared	0.372	0.527	0.719	0.754	0.776	0.654	0.767	0.782	-

Note: P-values in parenthese

Additionally, Models 5 and 6 introduce average schooling years and conflict intensity as control variables, yet both controls remain nonsignificant. The effect of WPE seemingly decreases when conflict intensity is controlled for, however this is likely as it also reduces the sample size which prevents generalisability. Model 9 presents the same multiple linear regression after multiple imputation was conducted on the log of GDP per capita, corruption average years of schooling and conflict intensity in attempts to combat these missing data limitations. The results reaffirm that the missing data likely biased the results of Model 6 as the coefficient of both WPE (and WPE^2 increase, with WPE gaining statistical significance and WPE^2 increasing in significance. Following this, as average schooling years and conflict intensity remain nonsignificant, Model 4 is deemed the best fit for estimating the causal effects of WPE.

Figure 1. Predicted Levels of Environmental Performance by Women's Political Empowerment at Representative Values of Log of GDP per Capita.



Note: Marginal effects presented at the 10th, 50th and 90th percentilesof Log of GDP per capita. Confidence intervals displayed.

To explore if the effect of WPE is influenced by that of GDP per capita, interactive effects were estimates with and without control variables. The results suggest that both higher GDP per capita and WPE need to coincide for WPE to exert a positive effect on EPI. Figure 1 displays the estimated marginal effects of WPE at varying levels of both WPE and log of GDP per capita. In line with previous models, low levels of WPE fail to exert significant marginal effects. WPE is shown to have a substantive positive effect on EPI only when both WPE and log GDP per capita are at the median level or above. This significant effect ($p < .001$) remains consistent when control variables are included, adding further validity to the results. Figure 1 illustrates how a log GDP per capita at the median or 90th percentile level leads WPE above the median to have significant marginal effects, supporting the second hypothesis. The wide and often overlapping confidence intervals do however limit certainty of the magnitude of this effect and therefore it cannot be confirmed if this varies as GDP per capita rises. There is relatively substantive evidence presented that when both log of GDP per capita and WPE are significantly increased (at the 90th percentile) the marginal effect is larger than when both remain at the median (B = 39.456 vs B = 99.576) yet the extent of this increase remains somewhat uncertain when assessing the wide confidence intervals. Figure 1 indicates that the nonlinearity of WPE's effect is dependent on the level of GDP per capita as demonstrated by the varying curvature. Notably, the larger confidence intervals towards the lower end of WPE and at the highest level are likely due to the low number of countries with scores within this range. The interactive effects

presented suggest that the positive effects of WPE may be obstructed by limited ability of a country to invest in the infrastructure and technology necessary for efficient and progressive solutions.

Table 3. Bivariate and Multivariate Logistic Regression Results with Climate Change Mitigation as the Dependent Variable

	M10	**M11**
Women's Political Empowerment (WPE)	8.010	6.280
	(<.001)	(<.001)
Log of GDP per capita		0.927
		(<.001)
Log of population size		-0.247
		(.038)
Corruption indicator		0.034
		(.063)
Intercept	-6.098	-10.287
	(<.001)	(.002)
N	166	160
R-squared	0.200	0.296

Note:*P* -value in parentheses.

To test the third hypothesis bivariate and multivariate logistic regressions were conducted and the results presented in Table 3. Model 10 depicts the results of the bivariate regression, suggesting that as WPE increases, the probability that a country has high levels of climate change mitigation also increases (B = 8.010, P <.001). To test the meaningfulness of these results, Model 11 presents a multivariate logistic regression with log of GDP per capita, log of population size, and corruption as control variables. The coefficient of WPE decreases by approximately ¼ (B = 6.280, p <.001), suggesting some effect was originally misallocated. As log of population size remains statistically insignificant and corruption is seemingly nonsignificant as well (p = .063), this reduction is likely due to the confounding effects of log GDP per capita present in Model 10, which are removed by controlling for this. Therefore, Model 11 provides more transparent and robust estimates of the impact of WPE. These margins were used to predict high climate change mitigation at varying levels of WPE (see Appendix B). To estimate probabilities using Model 11, control variables were kept at their mean values. When WPE is low (.05, .25), the predictability of high climate change mitigation is uncertain and insignificant. This is likely due to the low number of observations with WPE levels of .25 and below (p10 = .417). Contrastingly, when WPE increases to .50 and above, the probability of high climate change mitigation increases significantly. When WPE is at .50, with control variables fixed, the probability of a high level is lower than 20% whereas with WPE at .95 this probability rises to 80%. Evidently, the hypothesis that higher levels of WPE make a country more likely to achieve high levels of climate change mitigation is supported by this analysis.

Discussion and conclusions

Overall, the analysis presented in this paper has indicated that WPE at high levels does have a causal effect on a country's environmental performance, yet that this effect is nonlinear and dependent on additional factors such as GDP per capita. The funding constraints faced by states with lower GDP per capita likely dampens the benefits of WPE for effective policy. However, as states with high GDP per capital and WPE are estimated to gain positive environmental performance benefits from raising WPE, this aligns with ecofeminist perspectives that the journey towards advancing women's rights and the environmental movement remain interconnected (Mies and Shiva, 1993). Future research utilising qualitative methods to investigate the gendered motivations behind environmental policies, assessing individuals at a leadership level and also drivers of public opinion, could help to illuminate if the relationship indicated in these results is impacted by gendered social norms around caring responsibilities, cooperation and risk-averse decision making.

WPE is also found to increase the probability of high climate change mitigation. As harmful impacts of climate change and environmental degradation disproportionately effect women (Chitiga-Mabugu et al., 2023; Zavala et al., 2024), women are more likely to value and support environmental performance. Therefore, increasing WPE is likely to expand public pressure for environmental policy improvement. This research supports that of Asongu, Ndour, and Ngoungou (2024), illustrating how the positive impacts of WPE hold global potential. Policy makers should direct further attention to the role WPE may play in combatting global environmental crisis, and shape policy and strategy to account for this. Lastly, this paper does face some statistical limitations due to nonlinearity violating the assumptions of linear regression. Further research with additional nonlinear modelling methods, perhaps utilising longitudinal data, would be preferable to affirm findings.

Reference List:

Alexander, A.C., Bolzendahl, C. and Jalalzai, F. (2016) 'Defining Women's Global Political Empowerment: Theories and Evidence', *Sociology Compass*, 10(6), pp. 432–441. Available at: https://doi.org/10.1111/soc4.12375

Anastasia, A. and Nabilla, F. (2025) 'Forest fire disasters and ecological crisis: Impacts on women', *ASEAN Natural Disaster Mitigation and Education Journal*, 2(2), pp. 164-178. Available at: https://doi.org/10.61511/andmej.v2i2.2025.1623

Asongu, S.A., Messono, O.O. and Guttemberg, K.T.J. (2022) 'Women political empowerment and vulnerability to climate change: evidence from 169 countries', *Climatic Change*, 174(3-4), article number 30. Available at: https://doi.org/10.1007/s10584-022-03451-7

Asongu, S.A., Ndour, C.T. and Ngoungou, J.C.M. (2024) 'The effects of gender political inclusion and democracy on environmental performance: Evidence from the method of moments by quantile regression', *Politics & Policy*, 52(1), pp. 118-139. Available at: https://doi.org/10.1111/polp.12576

Block, S., Emerson, J.W., Esty, D.C., de Sherbinin, A., Wendling, Z.A., *et al.* (2024) *2024 Environmental Performance Index*. New Haven, CT: Yale Center for Environmental Law & Policy. Available at: epi.yale.edu (Accessed: 20 February 2025)

Boix, C., Miller, M.K. and Rosato, S. (2013) 'A Complete Data Set of Political Regimes, 1800-2007', Comparative Political Studies, 46(12), pp. 1523-1554. Available at: https://doi-org.bris.idm.oclc.org/10.1177/0010414012463905

Boleti, E., Garas, A., Kyriakou, A. and Lapatinas, A. (2021) 'Economic Complexity and Environmental Performance: Evidence from a World Sample', *Environmental Modeling & Assessment*, 26(3), pp. 251-270. Available at: https://doi.org/10.1007/s10666-021-09750-0

Chen, L. (2014) 'The Manifestations and Symptoms of Ecofeminism', *Studies in Sociology of Science*, 5(3), pp. 95-100. Available at: http://dx.doi.org/10.3968/5229

Chitiga-Mabugu, M., Henseler, M., Maisonnave, H. and Mabugu, R. (2023) 'Climate Change and Women - Impacts and Adaptation', *International Review of Environmental and Resource Economics*, 17(1), pp. 99-152. Available at: http://dx.doi.org/10.1561/101.00000151

Coppedge, M., Gerring, J., Knutsen, C.H., Lindberg, S.I., Teorell, J., Altman, D., Bernhard, M., Cornell, A., Fish, M.S., Gastaldi, L., Gjerlow, H., Glynn, A., God, A.G., Grahn, S., Hicken, A., Kinzelbach, K., Krusell, J., Marquardt, K.L., McMann, K., Mechkova, V., Medzihorsky, J., Natsika, N., Neundorf, A., Paxton, P., Pemstein, D., Pernes, J., Ryden, O., Von Romer, J., Seim, B., Sigman, R., Skaaning, S.-E., Staton, J., Sundstrom, A., Tzelgov, E., Wang, Y.-T., Wig, T., Wilson, S. and Ziblatt, D. (2023a) 'V-Dem Country-Year Dataset v13', Varieties of Democracy (V-Dem) Project. Available at: https://doi.org/10.23696/vdemds23

Coppedge, M., Gerring, J., Knutsen, C.H., Lindberg, S.I., Teorell, J., Altman, D., Bernhard, M., Cornell, A., Fish, M.S., Gastaldi, L., Gjerlow, H., Glynn, A., Grahn, S., Hicken, A., Kinzelbach, K., Marquardt, K.L., McMann, K., Mechkova, V., Neundorf, N., Paxton, P., Pemstein, D., Ryden, O., Von Romer, J., Seim, B., Sigman, R., Skaaning, S.-E., Staton, J., Sundstrom, A., Tzelgov, E., Uberti, L., Wang, Y.-T., Wig, T. and Ziblatt, D. (2023b) 'V-Dem Codebook v13', Varieties of Democracy (V-Dem) Project. Available at: https://doi.org/10.23696/vdemds23

Dahlum, S., Knutsen, C.H. and Mechkova, V. (2022) 'Women's political empowerment and economic growth', *World Development*, 156, article number 105822. Available at: https://doi.org/10.1016/j.worlddev.2022.105822

De Siano, R. and Chiariello, V. (2021) 'Women's political empowerment and welfare policy decisions: a spatial analysis of European countries', *Spatial Economic Analysis*, 17(1), pp. 101–126. Available at: https://doi.org/10.1080/17421772.2021.1905173

Donner, S., Hartmann, H. and Steinkamp, S. (2024) 'The Transformation Index BTI 2024: Governance in international comparison'. Bertelsmann Stiftung. Available at: http://www.bti-project.org

Fruttero, A., Halim, D., Broccolini, C., Coelho, B., Gninafon, H. and Muller, N. (2024) 'Gendered impacts of climate change: evidence from weather shocks',

Environmental Research: Climate, 3, 045018. Available at: https://doi.org/10.1088/2752-5295/ad8025

Gallego-Alvarez, I., Vicente-Galindo, M.P., Galindo-Villardón, M.P. and Rodríguez-Rosa, M. (2014) 'Environmental Performance in Countries Worldwide: Determinant Factors and Multivariate Analysis', *Sustainability*, 6(11), pp. 7807-7832. Available at: https://doi.org/10.3390/su6117807

Ganda, F. (2024) 'Gender empowerment and environmental impact: A quantile regression analysis in BRICS economies', *Heliyon*, 10(22), article number 40056. Available at: https://doi.org/10.1016/j.heliyon.2024.e40056

Hadenius, A. and Teorell, J. (2007) 'Pathways from Authoritarianism', *Journal of Democracy*, 18(1), pp. 143-156.

Handoyo, S. (2024) 'Public governance and national environmental performance nexus: Evidence from cross-country studies', *Heliyon*, 10(23), article number 40637. Available at: https://doi.org/10.1016/j.heliyon.2024.e40637

Inglehart, R. (1990) *Culture shift in advanced industrial society*. Princeton, NJ: Princeton University Press.

Inglehart, R. and Norris, P. (2003) *Rising tide: Gender equality and cultural change around the world*. Cambridge, UK: Cambridge University Press.

Kaufmann, D. and Kraay, A. (2024) 'Worldwide governance indicators, 2024 update'. Available at: http://www.govindicators.org (Accessed: 19 November 2024).

Lv, Z. and Deng, C. (2019) 'Does women's political empowerment matter for improving the environment? A heterogeneous dynamic panel analysis', *Sustainable Development*, 27(4), pp. 603-612. Available at: https://doi.org/10.1002/sd.1926

Mies, M. and Shiva, V. (1993) *Ecofeminism*. London, England: Zed Books.

Miller, M., Boix, C. and Rosato, S. (2022) 'Boix-Miller-Rosato Dichotomous Coding of Democracy, 1800-2020', *Harvard Dataverse*, V1. Available at: https://doi.org/10.7910/DVN/FENWWR

Nguyen, H.T. (2019) 'Gendered Vulnerabilities in Times of Natural Disasters: Male-to-Female Violence in the Philippines in the Aftermath of Super Typhoon Haiyan',

Violence Against Women, 25(4), pp. 421-440. Available at: https://doi-org.bris.idm.oclc.org/10.1177/1077801218790701

Pemstein, D., Marquardt, K.L., Tzelgov, E., Wang, Y.-T., Medzihorsky, J., Krusell, J., Miri, F. and Von Romer, J. (2023) *The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data*. V-Dem Working Paper No. 21. 8th edition. University of Gothenburg: Varieties of Democracy Institute.

Pratama, Y.A., Kadir, M.Y.A., Rivaldi, A., Mulya, I.C., Amirah, S., Iqhrammullah, M. (2024) 'Bibliometric analysis on the impacts of environmental degradation on women and the importance of women's representation', *Global Journal of Environmental Science and Management*, 10(3), pp. 939-954. Available at: https://doi.org/10.22034/gjesm.2024.03.01

Sinaga, C.A. and Harahap, R.H. (2023) 'Ecofeminist perspective: Climate change and its impact on women', *Journal of Sumatera Sociological Indicators*, 2(2), pp. 193-199. Available at: https://doi.org/10.32734/jssi.v2i02.14002

Sowah, J.K., Kirikkaleli, D. (2022) 'Investigating factors affecting global environmental sustainability: evidence from nonlinear ARDL bounds test', *Environmental Science and Pollution Research*, 29, pp. 80502–80519. Available at: https://doi.org/10.1007/s11356-022-21399-0

Standaert, S. (2015) 'Divining the level of corruption: A Bayesian state-space approach', *Journal of Comparative Economics*, 43(3), pp. 782–803. Available at: https://doi.org/10.1016/j.jce.2014.05.007

Stone, K., Blinn, N. and Spencer, R. (2022) 'Mental Health Impacts of Climate Change on Women: a Scoping Review', *Current Environmental Health Reports*, 9, pp. 228-243. Available at: https://doi.org/10.1007/s40572-022-00346-8

Sundstrom, A., Paxton, P., Wang, Y.-T. and Lindberg, S.I. (2017) 'Women's Political Empowerment: A New Global Index, 1900-2012', *World Development*, 94, pp. 321-335. Available at: https://doi.org/10.1016/j.worlddev.2017.01.016

Teorell, J., Sundstrom, A., Holmberg, S., Rothstein, B., Pachon, N.A., Dalli, C.M., Valverde, R.L., Phiri, V.S. and Gerber, L. (2025*) The Quality of Government*

Standard Dataset, Version Jan25. University of Gothenburg: The Quality of Government Institute. Available at: https://doi.org/10.18157/QoGStdJan25

UNESCO. (2024) 'UNESCO Institute for Statistics: Other Policy Relevant Indicators (OPRI)'. Available at: http://data.uis.unesco.org/

Wahman, M., Teorell, J. and Hadenius, A. (2013) 'Authoritarian Regime Types Revisited: Updated Data in Comparative Perspective', *Contemporary Politics*, 19(1), pp. 19-34.

Wendling, Z.A., Jacob, M., Esty, D.C., and Emerson, J.W. (2022) 'Explaining environmental performance: Insights for progress on sustainability', *Environmental Development*, 44, article number 100741. Available at: https://doi.org/10.1016/j.envdev.2022.100741

Winkler, H., Baumert, K., Blanchard, O., Burch, S. and Robinson, J. (2007) 'What factors influence mitigative capacity?', *Energy Policy*, 35(1), pp. 692-703. Available at: https://doi.org/10.1016/j.enpol.2006.01.009

World Bank. (2024) 'World Development Indicators'. Available at: https://databank.worldbank.org/source/world-development-indicators

Zavala, M. D., Cejas, C., Rubinstein, A. and Lopez, A. (2024) 'Gender Inequalities in the Impact of Climate Change on Health: A Scoping Review', *International Journal of Environmental Research and Public Health*, 21(8), pp. 1-20. Available at: https://doi.org/10.3390/ijerph21081093

Zhang, C., Khan, I., Dagar, V., Saeed, A. and Zafar, M.W. (2022) 'Environmental impact of information and communication technology: Unveiling the role of education in developing countries', *Technological Forecasting and Social Change*, 178, article number 121570. Available at: https://doi.org/10.1016/j.techfore.2022.121570

Appendices

Appendix A:

The Marginal Effects at Representative Values of WPE on EPI

Women's Political Empowerment index	M2		M4		M5		M6	
	Marginal effects of WPE	95% CI	Marginal effects of WPE	95% CI	Marginal effects of WPE	95% CI	Marginal effects of WPE	95% CI
.417	-8.459 (.271)	[-23.570, 6.651]	1.904 (.765)	[-10.657, 14.465]	2.555 (.713)	[-11.162, 16.271]	4.221 (.5630	[-10.212, 18.653]
.637	40.701 (<.001)	[33.582, 47.819]	29.526 (<.001)	[21.187, 37.865]	28.274 (<.001)	[18.637, 37.912]	25.729 (<.001)	[14.364, 37.095]
.782	73.102 (<.001)	[61.847, 84.356]	47.732 (<.001)	[34.984, 60.479]	45.226 (<.001)	[30.695, 59.756]	39.905 (<.001)	[22.014, 57.797]
.883	95.671 (<.001)	[79.232, 112.110]	60.413 (<.001)	[43.144, 77.681]	57.034 (<.001)	[37.578, 76.488]	49.780 (<.001)	[25.963, 73.596]
.925	105.056 (<.001)	[86.298, 123.814]	65.686 (<.001)	[46.404, 84.968]	61.943 (<.001)	[40.296, 83.591]	53.886 (<.001)	[27.474, 80.297]

Note: P-value in parentheses. CI stands for Confident Intervals. Marginal effects presented at the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of WPE.

Appendix B:

Predicted Probability of High Climate Change Mitigation at Varying Levels of Women's Political Empowerment

Women's Political Empowerment Index	Log of GDP per capita	Probability of high climate change mitigation	Log of population size	Corruption indicator
.05	-	-.003 (.363)	-	-
.05	8.688	.014 (.342)	15.636	45.006
.25	-	-.016 (.209)	-	-
.25	8.688	.047 (.117)	15.636	45.006
.50	-	-.110 (.013)	-	-
.50	8.688	.192 (.004)	15.636	45.006
.75	-	-.478 (<.001)	-	-
.75	8.688	.533 (<.001)	15.636	45.006
.95	-	-.819 (<.001)	-	-
.95	8.688	.800 (<.001)	15.636	45.006

Note: P-value in parentheses. Log of GDP per capita, Log of population size and Corruption indicator fixed at mean values.

Authors Bio

Theadora Gallie is a 2026 graduate of the University of Bristol's School for Policy Studies with a BSc in Social Policy and Politics. She is interested in research related to health policy, health equity, gender equality, disability and intersectionality. Through her undergraduate studies she has refined her skills in both quantitative and qualitative research methods and aims to continue to develop and use these skills in future research with the aims of supporting intersectional health equity. Outside of academics, she also enjoys taking part in various advisory and voluntary roles, using lived experience to influence health and disability research and policy.

Psychology andNeuroscience

Meet the Subject Editor

Emily Birt
Psychology and Neuroscience

Emily Birt is a second year PhD student in the School of Physiology, Pharmacology and Neuroscience, where she is investigating the circadian rhythms of pain. Having gained experience as a peer reviewer in the 2025 issue of the BILT Student Research Journal, she was excited to change roles and work as a Subject Editor this year. She has enjoyed reading all of the exciting work submitted by students and looks forward to seeing the finished product.

Serum Hydrogen Sulphide as a Biochemical Correlate of Major Depressive Disorder: A Cross-Sectional Case-Control Study from a Tertiary Care Centre in Sub-Himalayan India

Authors: Atrim Das

ORCID: 0009-0000-6441-3769

Programme: Molecular Neuroscience MSc

Affiliation: SchoolofPhysiology,Pharmacology and Neuroscience

Email: qf25081@bristol.ac.uk

Abstract

Hydrogen sulphide (H_2S) acts as a vital signalling molecule in the brain, regulating neuroinflammation and long-term potentiation (LTP), the sustained strengthening of synaptic connections necessary for memory. Since these neural pathways are often disrupted in Major Depressive Disorder (MDD), this study investigated whether measurable serum H_2S concentrations differ between depressed patients and healthy individuals. This cross-sectional study evaluated 50 drug-naive MDD patients and 50 healthy controls. Depression severity was assessed using the Hamilton Depression Rating Scale (HAM-D-17), and serum H_2S was quantified via a spectrophotometric methylene blue method. Mean serum H_2S was lower in the MDD group (0.068 ± 0.044 μmol/L) than in healthy controls (0.541 ± 0.272 μmol/L), representing an approximately 8-fold difference. HAM-D-17 scores confirmed very severe depression in the MDD cohort (28.94 ± 12.78 versus 3.96 ± 2.31). No significant correlation between serum H_2S and symptom severity was found within the MDD group. These findings suggest that substantial H_2S depletion is a core biochemical feature of untreated MDD, supporting the neuroprotective role of H_2S and highlighting the CBS/CSE–H_2S axis as a candidate biomarker and therapeutic target warranting further investigation.

Keywords hydrogen sulphide; major depressive disorder; gasotransmitter; biomarker; neuroinflammation

1. Introduction

Major Depressive Disorder (MDD) is a globally debilitating condition affecting approximately 280million people, ranking asaleading causeofyears lived with disability.[1] The World Health Organisation defines MDD as a persistent depressed mood or loss of pleasure lasting at least two weeks, paired with physical and cognitive disruptions that severely interfere with daily life.[1] Despite decades of research, the exact mechanisms remain partially obscured, and approximately one-third of patients do not respond to standard first-line medications.[2] For over sixty years, the monoamine hypothesis has dominated antidepressant research, attributing depression to deficits in serotonin, norepinephrine, and dopamine. While influential, this model struggles to explain the delayed onset of antidepressant action, high rates of treatment resistance, and the heterogeneity of depressive symptoms.[3] Scientific consensus is shifting toward a broader framework encompassing disrupted synaptic plasticity, dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis, neuroinflammation, and oxidative stress.[4,5] Hydrogen sulphide (H_2S) represents a promising candidate in this search. Known alongside nitric oxide and carbon monoxide as the third member of the gasotransmitter family, H_2S was once viewed strictly as an environmental toxin[6] but is now recognised as a naturally occurring signalling molecule. In the brain, it is primarily synthesised by cystathionine-beta-synthase (CBS), with cystathionine-gamma-lyase (CSE) operating in peripheral tissues and 3-mercaptopyruvate sulphurtransferase (3-MST) functioning within neurons.[7] CBS is highly concentrated in the hippocampus and cerebellum, predominantly in astrocytes, and is activated by calcium and calmodulin. At physiological concentrations, H_2S directly supports hippocampal long-term potentiation (LTP) by enhancing NMDA receptor activity, a process essential for memory and learning.[8] Subsequent research has expanded our understanding of H_2S, demonstrating that it suppresses neuroinflammation by modulating microglial and astrocytic activity, mitigates oxidative stress by restoring superoxide dismutase activity, modulates the HPA axis by suppressing corticotropin-releasing hormone, promotes neurogenesis via BDNF-related pathways, and activates the mTORC1 signalling pathway similarly to rapid-acting antidepressants.[9,10,11] In animal models subjected to chronic unpredictable stress, hippocampal H_2S is consistently depleted, and administering an H_2S donor reliably reverses depressive behaviours.[11] Despite robust preclinical findings, direct measurements of circulating H_2S in human MDD patients remain scarce. Yang et al. (2021) reported significantly lower plasma H_2S in depressed patients, with a negative correlation with symptom severity scores.[12] Dutta et al. (2024) found reduced serum H_2S in an MDD cohort from North Bengal, identifying a potential diagnostic threshold of less than 78.5 μmol/L with 96% sensitivity and 88% specificity.[13] A primary limitation

of earlier studies is the inclusion of patients already receiving pharmacological treatment, which introduces confounding due to the antioxidant effects of antidepressants. India faces a severe burden of untreated depression. The National Mental Health Survey (2015–2016) indicated that one in twenty Indians experiences depression, yet over 75% remain untreated.[14] Developing accessible and cost-effective biological markers could improve early detection in resource-limited settings. This study was therefore designed to compare serum H_2S concentrations between strictly drug-naive MDD patients and healthy controls at a tertiary care centre in the sub-Himalayan foothills of West Bengal, using the Hamilton Depression Rating Scale (HAM-D-17) for clinical assessment and the spectrophotometric methylene blue method for biochemical quantification.

2. Methods

2.1 Study Design and Setting

A hospital-based, cross-sectional case-control study was conducted over twelve months (January to December 2024) at North Bengal Medical College and Hospital (NBMCH), Sushrutanagar, Darjeeling, a 1,000-bed tertiary referral centre serving the sub-Himalayan region. The research protocol received formal ethical approval from the NBMCH Institutional Ethics Committee, and all participants provided written informed consentintheirpreferred language.

2.2 Participants

Fifty consecutive patients attending the Psychiatry outpatient department who met DSM-5 criteria for MDD were enrolled. All were classified as drug-naive, defined as having received no psychotropic medications for at least six weeks before enrolment. A control group of fifty healthy volunteers with no personal or close family history of psychiatric illness was recruited. All participants were aged 18 to 65 years and capable of providing informed consent. Individuals with chronic physical illness, co-occurring psychiatric conditions, substance use disorders, or bipolar depression were excluded. Sample size was calculated based on previously published standard deviations for serum H_2S in depressed and healthy populations, yielding a minimum of 47 participants per group (rounded to 50).

2.3 Clinical Assessment

A standardised questionnaire collected demographic data, including age, sex, residence, language, religion, marital status, education, occupation, and family structure. Socioeconomic status was assessed using the Modified B.G. Prasad Scale.

Depression severity was quantified by trained psychiatrists using the Hamilton Depression Rating Scale (HAM-D-17), a validated 17-item structured interview scored 0–52, with established thresholds: no depression (0–7), mild (8–12), moderate (13–17), severe (18–24), and very severe (≥25).[15]

2.4 Biochemical Analysis

A 5 mL venous blood sample was collected from each participant into a standard red-cap tube and processed immediately. Serum H_2S concentration was quantified using the validated spectrophotometric methylene blue method. Briefly, serum was mixed with phosphate-buffered saline and trichloroacetic acid, centrifuged at 3,000 rpm for 30 minutes, and the supernatant was treated with zinc acetate to precipitate sulphide. The precipitate was dissolved in N, N-dimethyl-p-phenylenediamine sulphate and ferric chloride, followed by sodium hydroxide. After 10 minutes of incubation, absorbance was measured at 650 nm using a double-beam UV/Visible spectrophotometer. Concentrations were calculated against a sodium hydrosulfide standard curve (R^2 = 0.99) and expressed in µmol/L.

2.5 Statistical Analysis

Data were organised in Microsoft Excel and analysed using IBM SPSS version 20. Continuous variables are presented as mean ± standard deviation, and categorical variables as counts and percentages. Group differences in continuous variables (serum H_2S, optical density, HAM-D-17 score) were assessed using independent- samples t-tests after confirming approximate normality by histogram and Q-Q plot inspection. Associations between categorical sociodemographic variables and serum H_2S were evaluated using chi-square tests. The relationship between serum H_2S and HAM-D-17 score was examined by linear regression within each group and across the combined sample. Statistical significance was set at $p < 0.05$.

3. Results

3.1 Sociodemographic Characteristics

The study enrolled 100 participants (50 MDD, 50 controls). Both groups shared highly similar demographic profiles with no statistically significant differences across any variable (all $p > 0.05$, chi-square test; Table 1). Most participants in both cohorts were young adults aged 21–30 years. Females accounted for 74% of the MDD group and 72% of controls, consistent with the well-documented higher prevalence of depression in women. The majority of participants across both groups were from rural areas, spoke Bengali, and fell into lower socioeconomic tiers (Classes IV and V,

Modified B.G. Prasad Scale). Patients with MDD were more likely to be single (62% vs 46%) and had higherrates of illiteracy (50% vs 32%). Mean illness duration in the MDD cohort was 19.26±10.63 months.

Variable	Category	MDD (n=50) n (%)	Controls (n=50) n (%)	p-value
Age (years)	<20	7 (14.0)	2 (4.0)	
	21–30	23 (46.0)	22 (44.0)	
	31–40	12 (24.0)	15 (30.0)	
	41–50	4 (8.0)	3 (6.0)	
	51–60	4 (8.0)	8 (16.0)	>0.05
Sex	Female	37 (74.0)	36 (72.0)	
	Male	13 (26.0)	14 (28.0)	>0.05
Residence	Rural	39 (78.0)	35 (70.0)	
	Urban	11 (22.0)	15 (30.0)	>0.05
Marital status	Single	31 (62.0)	23 (46.0)	
	Married	10 (20.0)	22 (44.0)	
	Widowed	9 (18.0)	5 (10.0)	>0.05
Education	Illiterate	25 (50.0)	16 (32.0)	
	Primary	10 (20.0)	23 (46.0)	
	High School	5 (10.0)	7 (14.0)	
	Graduate+	10 (20.0)	4 (8.0)	>0.05
Family type	Nuclear	34 (68.0)	31 (62.0)	
	Joint	16 (32.0)	19 (38.0)	>0.05
SES (Prasad)	Class I–II	13 (26.0)	15 (30.0)	

	Class III	5 (10.0)	4 (8.0)	
	Class IV–V	32 (64.0)	31 (62.0)	>0.05
Duration of illness (months)	Mean ± SD	19.26 ± 10.63	18.32 ± 10.32	>0.05

Table 1. Sociodemographic profile of MDD cases and healthy controls (n = 100) SES = socioeconomic status; Prasad = Modified B.G. Prasad Scale 2024. All p-values from the chi-square test.

3.2 Biochemical and Clinical Parameters

Mean serum H_2S was lower in the MDDgroup (0.068 ± 0.044 µmol/L) than in healthy controls (0.541 ± 0.272 µmol/L) [Table2], representing an approximately 8-fold difference ($p < 0.0001$, independent-samples t-test; Figure 1A). This difference was directly reflected in the raw optical densityreadings at 650 nm, which averaged 1.44 ± 1.18 in the MDD group versus 81.32±57.85 in controls ($p < 0.0001$; Figure 1C). HAM-D-17 scores confirmed very severedepression in the MDD cohort (28.94 ± 12.78) compared with the normal rangeincontrols (3.96 ± 2.31; $p < 0.0001$; Figure 1B).

Parameter	MDD Group (n=50) Mean ± SD	Control Group (n=50) Mean ± SD	Difference	p-value
Serum H_2S (µmol/L)	0.068 ± 0.044	0.541 ± 0.272	~8-fold lower in MDD	< 0.0001
Optical Density (650 nm)	1.44 ± 1.18	81.32 ± 57.85	~56-fold lower in MDD	< 0.0001
HAM-D-17 Score	28.94 ± 12.78	3.96 ± 2.31	Very severe vs Normal	< 0.0001
Duration of illness (months)	19.26 ± 10.63	18.32 ± 10.32	Comparable	> 0.05

Table 2. Key biochemical and clinical parameters in MDD cases and healthy controls (n = 100). H_2S = hydrogen sulphide; HAM-D = Hamilton Depression Rating Scale; SD = standard deviation. All group comparisons by an independent-samples t-test after confirming approximate normality by histogram and Q-Q plot inspection.

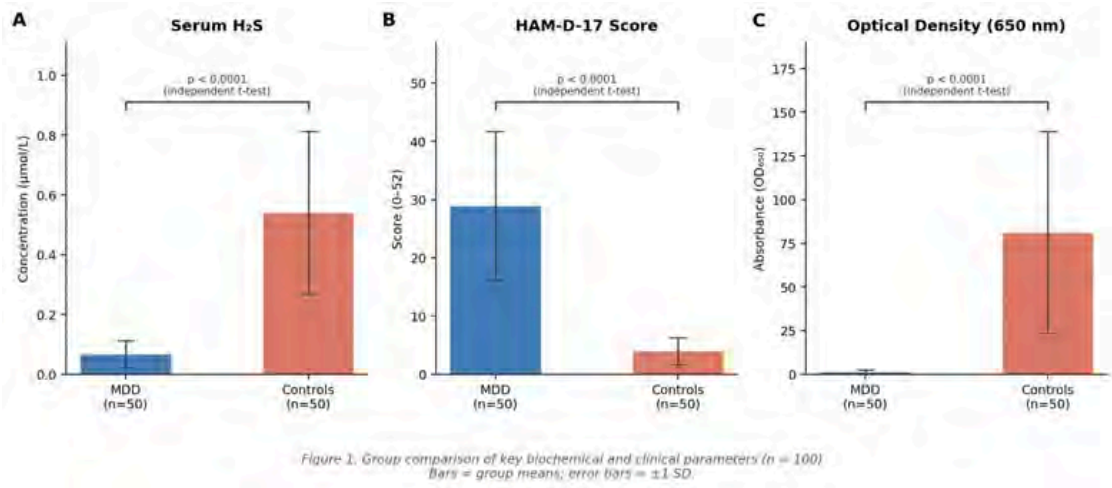


Figure 1. Group comparison of key biochemical and clinical parameters (n = 100). Bars = group means; error bars = ±1 SD. All comparisons by independent-samples t-test, p < 0.0001. (A) Serum H_2S is approximately 8-fold lower in MDD. (B) HAM-D-17 scores confirm very severe depression in MDD vs. the normal range in controls. (C) Optical density at 650 nm mirrors the H_2S difference.

3.3 Correlation Between H_2S and Depression Severity

Within the MDD group (Figure 2A), no significant correlation was observed (R^2 = 0.061, p = n.s.), with data points widely scattered without a clear trajectory. This indicates that while H_2S is substantially lower in clinical depression, the degree of depletion does not act as a linear predictor of symptom severity once the disorder is established. This pattern is consistent with a threshold-like relationship, in which H_2S may need to fall below a critical level for MDD to manifest, with secondary factors such as genetics and psychosocial stressors then determining symptom severity, though this interpretation remains speculative and requires prospective testing. In contrast, the control group displayed a strong positive correlation (R^2 = 0.772, p < 0.001; Figure 2B), reflecting the natural co-variation of H_2S and mood-related measures in the absence of psychiatric illness. The combined dataset showed a moderate negative correlation (R^2 = 0.287; Figure 2C), confirming that higher H_2S concentrations generally correspond to lower depressive symptom burden across the full sample.

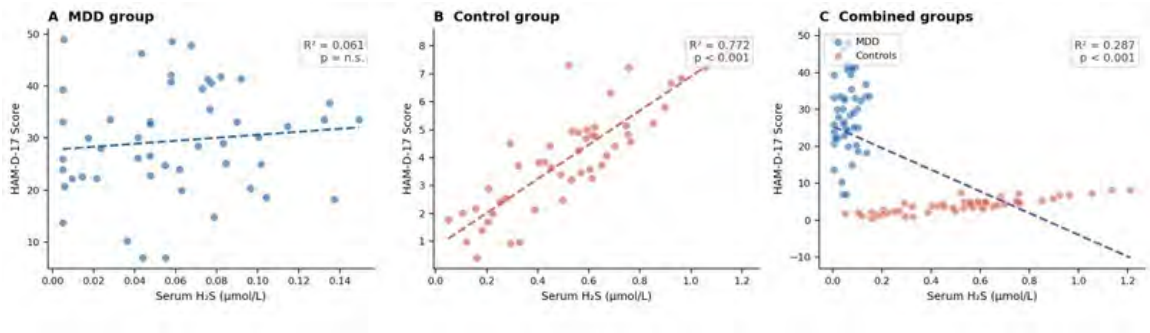


Figure 2. Linear regression analysis: serum H_2S vs HAM-D-17 score. (A) MDD group: no significant correlation (R^2 = 0.061, p = n.s.), consistent with a threshold-like relationship. (B) Control group: strong positive correlation (R^2 = 0.772, p < 0.001). (C) Combined groups: moderate negative correlation (R^2 = 0.287)

3.4 Associations with Sociodemographic Variables

Chi-square analyses in both the MDD and controlgroups found no statistically significant associations between serum H_2S concentration and any sociodemographic variable, including age, sex, residence, language, religion, marital status, education, occupation, socioeconomic status, family type, or duration of illness (all p > 0.05). This suggests that the observed H_2S depletion in MDD reflects a disease-specific biochemical phenomenon rather than demographic or social determinants.

4. Discussion

This study provides direct clinical evidence from a drug-naive South Asian population that serum H_2S is substantially lower in untreated MDD than in healthy controls. An approximately 8-fold lower mean serum H_2S in MDD patients compared with healthy controls is consistent with the hypothesis that H_2S deficiency is a biological correlate of MDD, though the cross-sectional design precludes causal inference. By focusing on a drug-naive cohort, this research offers a clearer view of the disease's baseline biochemistry, unconfounded by antidepressant effects. The near-complete suppression of circulating H_2S (mean 0.068 µmol/L) aligns with earlier findings by Yang et al. (2021)12 and Dutta et al. (2024).13 Biologically, this is coherent: H_2S is essential for activating NMDA receptors to stimulate LTP, the process of strengthening neural connections critical for emotional regulation and memory.8,4 Its depletion removes several layers of neurological defence. Normally, H_2S acts as a potent antioxidant and animal models show that replenishing it restores chemical balance and reverses depressive behaviours.11,5 It also modulates the brain's immune response, lowering inflammatory markers frequently elevated in depressed patients,16 buffers against stress hormones,17, encourages neurogenesis,11 and activates rapid-

acting neural pathwayssimilartoketamine.9 Theabsenceofasignificantwithin-MDD correlation betweenserumH_2SandHAM-D-17score(R^2=0.061,p=n.s.)suggests that serum H_2S distinguished MDD patients from controls more clearly than it tracked symptom severitywithintheMDDgroup.Thispatterniscompatiblewithathreshold- like relationship,inwhichH_2SmayfallbelowacriticalrangeinMDD,whileother factors suchasgenetics,inflammation,sleepdisruption,illnessduration,and psychosocialstressshapeclinicalseverity.However,thisinterpretationremains speculativeandrequiresprospectivetesting.Thecontrol-groupassociationshould also beinterpretedcautiouslybecauseHAM-D-17scoresinhealthyparticipants occupyanarrownon-clinicalrange.Similarly,thecombined-cohortcorrelationlikely reflectscase-controlseparationratherthanacontinuousseverityrelationshipwithin MDD.Thestrongpositivecorrelationinhealthyindividuals(R^2=0.772)confirmsthat H_2S naturallyfluctuatesalongsidehealthymoodregulation.Crucially,socialfactors like whetherapatientwasrural,single,orfacingeducationalbarriershadnosignificant impact onH_2Sconcentrations,confirmingthattheH$_2$Sdeficitisadistinctbiochemical feature ofthediseaseratherthanareflectionofsociodemographicdisadvantage. These results carry implications for psychiatric care. The spectrophotometric methylenebluemethodislow-costandrequiresonlystandardlaboratoryequipment, making H_2Smeasurementfeasibleinresource-limitedsettings.Beyonddiagnostics, this studypointstowardtherapeuticinnovation:drugsdesignedtosafelyreleaseH_2S have alreadyshownefficacyinreversingdepression-likesymptomsinanimalmodels, offering a potential new avenue for antidepressant development.18 This study has severallimitations.Thecross-sectionaldesigncapturesasingletimepointandcannot establishwhetherH_2Sdepletionprecedes,accompanies,orfollowstheonsetofMDD. The samplewasdrawnfromasingletertiarycentre,whichmaylimitgeneralisability. Formalnormalitytestingwasnotperformedbeforeparametricanalysis,andfuture studiesshouldconsidernon-parametricalternativesgiventhenear-floordistribution of H_2S intheMDDgroup.Multivariateadjustmentforpotentialconfounderssuchasdiet, physicalactivity,andgutmicrobiomecompositionwasnotperformed.Future directionsincludelongitudinaltrackingtodeterminewhetherH_2Sdropsbefore a depressiveepisode,cerebrospinalfluidanalysistomapH_2Sdirectlywithinthecentral nervoussystem,andexplorationofdietaryandgeneticinfluencesonthisprotective gasotransmitter.

5. Conclusion

This study demonstrates that serum H_2S is substantially and specifically lower in drug-naive MDD patients compared with healthy controls, independent of sociodemographic factors. The finding is biologically coherent with H_2S's established

roles in hippocampal synaptic plasticity, neuroinflammation, oxidative stress defence, HPA axis regulation, and neurogenesis, all pathways disrupted in MDD. The CBS/CSE–H_2S axis warrants further investigation as a candidate biomarker and therapeutic target in major depressive disorder.

References

[1] World HealthOrganization. Depressive disorder (depression). Geneva: WHO; 2023.

[2] Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder. Lancet. 2018;391(10128):1357–66.

[3] Filatova EV, Shadrina MI, Slominsky PA. Major Depression: One Brain, One Disease, One Set of Intertwined Processes. Cells. 2021;10(6):1283.

[4] Belleau EL, Treadway MT, Pizzagalli DA. The Impact of Stress and Major Depressive Disorder on Hippocampal and Medial Prefrontal Cortex Morphology. Biol Psychiatry. 2019;85(6):443–53.

[5] Juszczyk G, Mikulska J, Kasperek K, et al. Chronic Stress and Oxidative Stress as Common Factors of the Pathogenesis of Depression and Alzheimer's Disease. Antioxidants (Basel). 2021;10(9):1439.

[6] Furne J, Saeed A, Levitt MD. Whole tissue hydrogen sulfide concentrations are orders of magnitude lower than presently accepted values. Am J Physiol Regul Integr Comp Physiol. 2008;295(5):R1479–85.

[7] Tan BH, Wong PTH, Bian JS. Hydrogen sulfide: A novel signaling molecule in the central nervous system. Neurochem Int. 2010;56(1):3–10.

[8] Abe K, Kimura H. The possible role of hydrogen sulfide as an endogenous neuromodulator. J Neurosci. 1996;16(3):1066–71.

[9] Li N, Lee B, Liu RJ, et al. mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. Science. 2010;329(5994):959–64.

[10] Yang T, Nie Z, Shu H, et al. The Role of BDNF on Neural Plasticity in Depression. Front Cell Neurosci. 2020;14:82.

[11] Hou XY, Hu ZL, Zhang DZ, et al. Rapid Antidepressant Effect of Hydrogen Sulfide: Evidence for Activation of mTORC1-TrkB-AMPA Receptor Pathways. Antioxid Redox Signal. 2017;27(8):472–88.

[12] Yang YJ, Chen CN, Zhan JQ, et al. Decreased Plasma Hydrogen Sulfide Level Is Associated With the Severity of Depression in Patients With Depressive Disorder. Front Psychiatry. 2021;12:765664.

[13] Dutta SS, Dasgupta S, Banerjee AK, et al. Exploring the Role of Serum Hydrogen Sulphide (H_2S) Levels in Manic Depressive Psychosis. Cureus. 2024;16(3):e56857.

[14] Murthy RS. National Mental Health Survey of India 2015–16. Indian J Psychiatry. 2017;59(1):21–6.

[15] Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23(1):56–62.

[16] Kouba BR, de Araujo Borba L, Borges de Souza P, et al. Role of Inflammatory Mechanisms in Major Depressive Disorder. Cells. 2024;13(5):423.

[17] Kimura H. Hydrogen sulfide: its production and functions. Exp Physiol. 2011;96(9):833–5.

[18] Szabo C. Hydrogen sulphide and its therapeutic potential. Nat Rev Drug Discov. 2007;6(11):917–35.

[19] Cirino G, Szabo C, Papapetropoulos A. Physiological roles of hydrogen sulfide in mammalian cells, tissues, and organs. Physiol Rev. 2023;103(1):31–276.

Authors Bio:

My name is Atrim Das, and I am an MSc Molecular Neuroscience student at the University of Bristol, with a background in Biotechnology. My academic interests lie at the intersection of neuroscience, biotechnology, artificial intelligence, and translational medicine. Through my research experiences, I have explored areas ranging from psychiatric biomarkers and clinical biochemistry to gene therapy, molecular cloning, and AI-driven healthcare tools. I am particularly interested in how molecular and computational approaches can help us understand brain disorders and develop more accessible, precise, and patient-centred solutions. I hope to contribute to research that bridges laboratory science, clinical insight, and real-world impact.

Neuroticism, Conscientiousness, and Disease-Relevant Social Stress Processes

Author: Natalie Kah Xuan Koh

ORCID: 0009-0001-4720-5308

Programme: BSc Psychology

Affiliation: SchoolofPsychological Science

Email: pm23217@bristol.ac.uk

Abstract

Personality plays a critical role in shaping how individuals perceive and cope with stress within their social environments, influencing both physical and mental health. This essay discusses how individual differences in personality relate to disease-relevant social psychological processes, focusing on stress appraisal and coping. Based on Lazarus and Folkman's (1984) Transactional Model of Stress, it highlights how neuroticism and conscientiousness relate to differences in stress evaluation, coping strategies, and physiological regulation. Higher neuroticism is associated with heightened perceptions of social threat and greater stress reactivity, whereas conscientiousness is consistently linked to more adaptive, problem-focused coping, and social stability. When these traits interact, healthy neuroticism can emerge, where worry may motivate constructive, health-promoting behaviours. Together, these findings suggest that personality traits may contribute to disease vulnerability through their effects on social cognitive and emotional processes involved in stress regulation and health-related outcomes. These mechanisms emphasise that health results from the interaction between biological and social processes, shaped by individual differences in personality.

Keywords: Neuroticism, Conscientiousness, Stress appraisal, Social cognition, Disease vulnerability

Introduction

Personality plays a central role in shaping how individuals perceive and cope with stressors in their social environment. Disease-relevant social psychological processes are the cognitive and emotional mechanisms linking social interaction to health (Uchino, 2006). While much research has examined disgust sensitivity as a disease- protective mechanism, psychosocial models highlight stress appraisal and coping as equally critical for understanding disease vulnerability. This is because they are associated with changes in physiological systems, such as the hypothalamic– pituitary–adrenal (HPA) axis and immune function (Karin et al., 2020), which are linked to health outcomes. The HPA axis is a central neuroendocrine stress system (Dunlavey, 2018) involving interactions between the hypothalamus, pituitary gland, and adrenal glands, which regulate cortisol release in response to stress (DeMorrow, 2018). Activation of this system is adaptive in the short term, but prolonged or repeated activation may contribute to physiological dysregulation, including impaired immune functioning and increased disease vulnerability (Sheng et al., 2021). Understanding how personality traits shape these stress processes may therefore provide insight into why individuals differ in vulnerability to both mental and physical illness. This essay draws on research from personality, social, and health psychology to examine how personality influences stress appraisal, coping, and disease vulnerability. This essay focuses on stress appraisal and coping, a framework proposed by Lazarus and Folkman (1984), explaining how individuals evaluate and react to stressors in social contexts. Although Lazarus and Folkman's (1984) framework remains foundational within stress research, more recent studies have expanded these concepts by examining physiological regulation and long-term health outcomes (Graham et al., 2020; Obbarius et al., 2021). Primary appraisal involves evaluating whether a situation poses a threat, and secondary appraisal assesses the individual's coping capacity (Anderson, 2004). Neuroticism, characterised by emotional instability and sensitivity to threat, tends to be associated with heightened stress appraisal and promotes emotion-focused coping. In contrast, conscientiousness, associated with organisation, self-control, and practicality, is linked to adaptive, problem-focused coping (Costa & McCrae, 1992). The interaction between these two traits, known as healthy neuroticism, demonstrates how personality trait combinations can further shape health outcomes. This essay argues that individual differences in personality may shape

disease vulnerability through effects on stress appraisal and coping within social contexts, highlighting that the social evaluation of stress is a key pathway linking personality to health.

Neuroticism and Social Stress Appraisal

A key pathway linking personality to health is the appraisal of social stressors. Neuroticism, characterised by emotional instability, worry, and high sensitivity to threat, is associated with differences in how individuals perceive and cope with social threat. According to the Transactional Model of Stress (Lazarus & Folkman, 1984), the way an individual construes a situation influences their emotional and behavioural response. Stress arises when individuals perceive a situation as threatening (primary appraisal) and doubt their ability to cope (secondary appraisal). Individuals higher in neuroticism tend to exhibit heightened sensitivity to ambiguous social cues, often interpreting them negatively, and increased reactivity to social threats, such as exclusion and criticism. They also tend to underestimate their coping abilities and resources (Hopkins et al., 2021), although much of this evidence relies on self-report personality measures that may inflate associations due to shared method variance. This may skew their primary stress appraisal towards perceiving interpersonal interactions as more threatening and stressful.

This heightened effect on stress appraisal is especially pronounced in social- evaluative contexts, such as when one's social standing or self-esteem is at stake. Dickerson and Kemeny's (2004) meta-analysis found that socially evaluative tasks, such as being recorded or judged by peers, induced the strongest cortisol response, although most studies relied on controlled laboratory stress paradigms, which may not fully capture the complexity of real-world social stress. While earlier work focused primarily on subjective stress appraisal, more recent research has incorporated neuroendocrine markers such as cortisol regulation and inflammatory responses to better examine the psychological effects of chronic social stress (Knezevic et al., 2023). Consistent with this, Bosch et al. (2009) found that stressors without the aspect of social-evaluative threat likely failed to elicit a cortisol response due to a lack of HPA axis activation, and that HPA activation is a physiological response to social-evaluative threat. The repeated activation of this appraisal pattern sustains HPA axis activity and

cortisol release (Karin et al., 2020). As the HPA axis regulates adrenal secretion of hormones (Herman et al., 2016), prolonged exposure to such activation has a measurable impact on health. Over time, the dysregulation of the HPA axis from chronic stress contributes to higher vulnerability to mental health problems such as depression and anxiety (Kirkbride et al., 2024; Malouff et al., 2005), as well as physical illness such as immune suppression, and increased risk of cardiovascular disease (Rosmond & Björntorp, 2000).

However, this pathway is not deterministic as many studies rely on self-reports of neuroticism, which may be influenced by negative affectivity, a characteristic common in neurotic individuals (Miller et al., 2009). This tendency toward intense negative emotions may inflate observed associations between neuroticism and stress reactivity. However, the consistency of findings across physiological and behavioural studies suggests that the relationship between neuroticism and stress reactivity is unlikely to reflect self-report bias alone. Furthermore, the effects of neuroticism depend on social context, as perceived social support can buffer stress reactivity in neurotic individuals (Hostinar & Gunnar, 2015; Ozbay et al., 2007), emphasising that greater social support results in lower physiological responses to interpersonal stress. This suggests that social resources can buffer trait vulnerability and emphasises that neuroticism's impact on health comes from the interaction between personality and the social environment. Therefore, neuroticism heightens social stress appraisal and reactivity, especially in environments that are social-evaluative or low in support. These socially mediated stress responses link personality differences to disease vulnerability, illustrating how individual traits affect both physical and mental health through social psychological pathways.

Conscientiousness and Social Coping Processes

In contrast to neuroticism's tendency to amplify stress, conscientiousness promotes adaptive coping and social stability, protecting health. Individuals higher in conscientiousness, characterised by organisation, discipline, and self-control (Costa & McCrae, 1992), tend to approach interpersonal stressors proactively rather than reactively. Lazarus & Folkman's (1984) Transactional Model of Stress highlights that stress responses are not only dependent on how threatening the stressor is perceived

as during primary appraisal, but also one's perceived coping capacity during secondary appraisal. Therefore, individuals higher in conscientiousness tend to perceive social stressors as manageable and believe they can cope effectively, reducing perceived threat and emotional reactivity.

Conscientiousness promotes problem-focused coping, such as by planning, seeking solutions, and negotiating conflict constructively (Afshar et al., 2015). In social contexts, these coping strategies maintain stability and predictability in relationships. Consistent and responsible behaviour fosters trust and reciprocity, which can strengthen social bonds and perceived support (Canevello & Crocker, 2010). Strong social networks can buffer stress responses, lowering HPA axis activation and cortisol activity (Cohen & Wills, 1985). Over time, such socially mediated regulation of stress can reduce the risk of immunosuppression and cardiovascular strain (House et al., 1988). Empirically, conscientious individuals show greater positive affect, life satisfaction, and confidence in coping (Bartley & Roesch, 2011; Hayes & Joseph, 2003). These are factors which may reinforce adaptive stress responses. Meta-analytic findings also concluded that high levels of conscientiousness were associated with lower risk of mortality (Graham et al., 2017), although these findings are primarily based on observational cohort studies, which limits conclusions about causal mechanisms.

However, the adaptiveness of conscientiousness is not universal. Much of the current literature derives evidence from Western, educated, industrialised, rich, and democratic (WEIRD) populations (Gurven et al., 2013), limiting generalisability across cultural contexts. In collectivist cultures, for instance, assertive problem-solving may conflict with social harmony norms, as indirect or collaborative approaches tend to be favoured (Liu et al., 2021).

Moreover, excessively high conscientiousness may also undermine these protective pathways when self-regulation becomes excessively rigid or overcontrolled. In such cases, excessive planning, behavioural control, and concern with responsibility may increase interpersonal strain and stress rather than reduce it, becoming maladaptive (Samuel & Widiger, 2011), and potentially limiting the stress-buffering effects usually associated with conscientiousness. In extreme cases, maladaptive conscientiousness has also been linked to obsessive-compulsive tendencies and inflexible coping behaviours (Coleman et al., 2023). Moreover, the direction of causality is unclear, as

good health and stable routines may reinforce conscientious habits rather than result from them. In summary, conscientiousness is linked to better health outcomes by shaping social coping and relationship maintenance, thereby reducing stress reactivity. Through fostering problem-focused coping, cooperative social networks, and physiological regulation, conscientiousness protects against disease through socially mediated pathways linking personality to stress and health.

Healthy Neuroticism and Trait Interaction

Personality traits rarely operate in isolation,and the interaction between neuroticism and conscientiousness provides insight intohow combinations of traits can shape health outcomes. Healthy neuroticism describes individuals high in both neuroticism and conscientiousness, where heightened sensitivity to threat is moderated by self-regulation, organisation, and goal-directed coping (Friedman, 2000; Stieger et al., 2020). In this interaction, anxious vigilance may be channelled into precautionary and health-promoting behaviours through greater planning, behavioural monitoring, and active coping. Individuals high in healthy neuroticism may therefore display greater health vigilance, cautiousness, and self-monitoring, while still maintaining the organisational control needed to respond adaptively to stress (Weston & Jackson, 2018).

On its own, high neuroticism is typically linked to maladaptive stress appraisal and avoidance. However, when paired with conscientiousness, the anxiety and self-monitoring associated with neuroticism may be channelled into organised, preventive coping (Friedman, 2000). Conscientious self-regulation allows anxiety to motivate constructive actions and solutions, such as planning, seeking social support, and adhering to medical advice (Turiano et al., 2015). These behaviours may reduce physiological burdens of stress by facilitating problem-focused coping and improving adherence to socially reinforced health norms. In social-evaluative contexts, where reputation and self-esteem are at stake, this trait interaction fosters perceived control and stability. Individuals high in both traits are more likely to actively address interpersonal conflict or stressors through constructive problem-solving rather than withdrawal (Quek et al., 2025). These types of social coping behaviours maintain social support networks which can buffer stress and HPA activation, reducing health

risk (Cohen & Wills, 1985). A meta-analysis by Graham et al. (2020) supports this synergistic effect, showing that individuals high in both traits engaged more in disease-preventive actions, medical compliance, and were less likely to engage in unhealthy behaviours that ultimately contribute to earlier mortality.

Nevertheless, evidence for healthy neuroticism remains mixed. Some studies fail to replicate this interaction, possibly due to individual differences in conscientiousness and stress appraisal. Early conceptualisations of healthy neuroticism were largely theoretical, whereas more recent studies have examined its behavioural and health-related implications empirically. This inconsistency may reflect differences in how healthy neuroticism is conceptualised and measured across studies, with some research focusing on behavioural outcomes while others examine physiological or self-reported health indicators. For example, Weston et al. (2019) found little evidence of healthy neuroticism having a positive impact on health. Nevertheless, the broader literature generally supports the idea that conscientious self-regulation may moderate some maladaptive effects typically associated with neuroticism. Additionally, the concept may reflect situational variability rather than actual trait interaction, as stress and coping are context dependent (DeLongis & Holtzman, 2005). Therefore, more research is required to establish if this interaction between traits truly alters stress responses or if it is simply a reflection of self-report tendencies. Despite these inconsistencies, healthy neuroticism illustrates how personality configurations influence disease vulnerability through social psychological processes, including stress appraisal, social coping, and norm adherence. Together, these processes influence both stress regulation and broader health-related outcomes. When worry and stress are regulated through conscientious control, neurotic vigilance potentially becomes protective rather than harmful. Taken together, these findings suggest that the health effects of neuroticism depend less on the trait itself and more on the regulatory processes that accompany it.

Conclusion

Research consistently links higher neuroticism with heightened perceptions of social threat and prolonged physiological stress activation, which may contribute to greater disease vulnerability. In contrast, conscientiousness is associated with adaptive

coping and more effective stress regulation, which may be linked to lower health risks. When these two traits interact, conscientiousness can channel neurotic vigilance into constructive self-regulation, known as healthy neuroticism. Taken together, these findings illustrate that personality shapes health through social cognitive processes that determine how people interpret and manage social stress. Although the broader literature widely supports associations between personality and health, this essay highlights the importance of examining the social-cognitive mechanisms through which these effects operate. In particular, stress appraisal and coping processes appear to represent key pathways linking personality traits to both physiological regulation and disease vulnerability. Perceptions of appraisal, control, and support represent important psychological connections between personality and disease. Overall, this demonstrates that disease-relevant social psychological processes are shaped by individual differences in personality. Together, these findings suggest that personality traits influence health not only through individual behavioural tendencies, but also through social cognitive processes that shape how stress is perceived, managed, and biologically regulated. These mechanisms highlight that health is both a biological and social phenomenon, shaped and influenced by individual differences in personality. Future longitudinal and cross-cultural research incorporating behavioural and physiological measures may help clarify how social stress processes and personality traits interact to increase or protect against disease vulnerability.

References

Afshar, H., Roohafza, H. R., Keshteli, A. H., Mazaheri, M., Feizi, A., & Adibi, P. (2015). The association of personality traits and coping styles according to stress level. *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences*, *20*(4), 353–358. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4468450/

Anderson, N. (2004). *Stress, appraisal, and coping*. SAGE Publications, Inc. https://doi.org/10.4135/9781412952576.n198

Bartley, C. E., & Roesch, S. C. (2011). Coping with daily stress: The role of conscientiousness. *Personality and Individual Differences*, *50*(1), 79–83. https://doi.org/10.1016/j.paid.2010.08.027

Bosch, J. A., De Geus, E. J. C., Carroll, D., Goedhart, A. D., Anane, L. A., Veldhuizen Van Zanten, J. J., Helmerhorst, E. J., & Edwards, K. M. (2009). A general enhancement of autonomic and cortisol responses during social evaluative threat. *Psychosomatic Medicine*, *71*(8), 877–885. https://doi.org/10.1097/PSY.0b013e3181baef05

Canevello, A., & Crocker, J. (2010). Creating good relationships: Responsiveness, relationship quality, and interpersonal goals. *Journal of Personality and Social Psychology*, *99*(1), 78–106. https://doi.org/10.1037/a0018186

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, *98*(2), 310–357. https://doi.org/10.1037/0033-2909.98.2.310

Coleman, G., Furnham, A., & Treglown, L. (2023). Exploring the Dark side of conscientiousness. The relationship between conscientiousness and its

potential derailers: Perfectionism and narcissism. *Current Psychology*, *42*(31), 27744–27757. https://doi.org/10.1007/s12144-022-03828-y

Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

DeLongis, A., & Holtzman, S. (2005). Coping in context: The role of stress, social support, and personality in coping. *Journal of Personality*, *73*(6), 1633–1656. https://doi.org/10.1111/j.1467-6494.2005.00361.x

DeMorrow, S. (2018). Role of the hypothalamic–pituitary–adrenal axis in health and disease. *International Journal of Molecular Sciences*, *19*(4), 986. https://doi.org/10.3390/ijms19040986

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*(3), 355–391. https://doi.org/10.1037/0033-2909.130.3.355

Dunlavey, C. J. (2018). Introduction to the hypothalamic-pituitary-adrenal axis: Healthy and dysregulated stress responses, developmental stress and neurodegeneration. *Journal of Undergraduate Neuroscience Education*, *16*(2), R59–R60. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6057754/

Friedman, H. S. (2000). Long-term relations of personality and health: Dynamisms, mechanisms, tropisms. *Journal of Personality*, *68*(6), 1089–1107. https://doi.org/10.1111/1467-6494.00127

Graham, E. K., Rutsohn, J. P., Turiano, N. A., Bendayan, R., Batterham, P. J., Gerstorf, D., Katz, M. J., Reynolds, C. A., Sharp, E. S., Yoneda, T. B., Bastarache, E. D., Elleman, L. G., Zelinski, E. M., Johansson, B., Kuh, D., Barnes, L. L., Bennett, D. A., Deeg, D. J. H., Lipton, R. B., … Mroczek, D. K.

(2017). Personality predicts mortality risk: An integrative data analysis of 15 international longitudinal studies. *Journal of Research in Personality*, *70*, 174–186. https://doi.org/10.1016/j.jrp.2017.07.005

Graham, E. K., Weston, S. J., Turiano, N. A., Aschwanden, D., Booth, T., Harrison, F., James, B. D., Lewis, N. A., Makkar, S. R., Mueller, S., Wisniewski, K. M., Yoneda, T., Zhaoyang, R., Spiro, A., Willis, S., Schaie, K. W., Sliwinski, M., Lipton, R. A., Katz, M. J., … Mroczek, D. K. (2020). Is healthy neuroticism associated with health behaviors? A coordinated integrative data analysis. *Collabra: Psychology*, *6*(1), 32. https://doi.org/10.1525/collabra.266

Gurven, M., Von Rueden, C., Massenkoff, M., Kaplan, H., & Lero Vie, M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian Amazon. *Journal of Personality and Social Psychology*, *104*(2), 354–370. https://doi.org/10.1037/a0030841

Hayes, N., & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality and Individual Differences*, *34*(4), 723–727. https://doi.org/10.1016/S0191-8869(02)00057-0

Herman, J. P., McKlveen, J. M., Ghosal, S., Kopp, B., Wulsin, A., Makinson, R., Scheimann, J., & Myers, B. (2016). Regulation of the hypothalamic-pituitary-adrenocortical stress response. In Y. S. Prakash (Ed.), *Comprehensive Physiology* (1st ed., pp. 603–621). Wiley. https://doi.org/10.1002/cphy.c150015

Hopkins, J., Lang, C., Glenn, D. E., & Michalska, K. J. (2021). Neuroticism predicts increased sensitivity in identifying negative facial affect in young adults. *UC Riverside Undergraduate Research Journal*, *15*(1). https://doi.org/10.5070/RJ515355194

Hostinar, C. E., & Gunnar, M. R. (2015). Social support can buffer against stress and shape brain activity. *AJOB Neuroscience*, *6*(3), 34–42. https://doi.org/10.1080/21507740.2015.1047054

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science*, *241*(4865), 540–545. https://doi.org/10.1126/science.3399889

Karin, O., Raz, M., Tendler, A., Bar, A., Korem Kohanim, Y., Milo, T., & Alon, U. (2020). A new model for the HPA axis explains dysregulation of stress hormones on the timescale of weeks. *Molecular Systems Biology*, *16*(7), e9510. https://doi.org/10.15252/msb.20209510

Kirkbride, J. B., Anglin, D. M., Colman, I., Dykxhoorn, J., Jones, P. B., Patalay, P., Pitman, A., Soneson, E., Steare, T., Wright, T., & Griffiths, S. L. (2024). The social determinants of mental health and disorder: Evidence, prevention and recommendations. *World Psychiatry*, *23*(1), 58–90. https://doi.org/10.1002/wps.21160

Knezevic, E., Nenic, K., Milanovic, V., & Knezevic, N. N. (2023). The role of cortisol in chronic stress, neurodegenerative diseases, and psychological disorders. *Cells*, *12*(23), 2726. https://doi.org/10.3390/cells12232726

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer Publishing Company.

Liu, S. S., Shteynberg, G., Morris, M. W., Yang, Q., & Galinsky, A. D. (2021). How does collectivism affect social interactions? A test of two competing accounts. *Personality and Social Psychology Bulletin*, *47*(3), 362–376. https://doi.org/10.1177/0146167220923230

Malouff, J. M., Thorsteinsson, E. B., & Schutte, N. S. (2005). The relationship between the five-factor model of personality and symptoms of clinical disorders:

A meta-analysis. *Journal of Psychopathology and Behavioral Assessment*, *27*(2), 101–114. https://doi.org/10.1007/s10862-005-5384-y

Miller, D. J., Vachon, D. D., & Lynam, D. R. (2009). Neuroticism, negative affect, and negative affect instability: Establishing convergent and discriminant validity using ecological momentary assessment. *Personality and Individual Differences*, *47*(8), 873–877. https://doi.org/10.1016/j.paid.2009.07.007

Obbarius, N., Fischer, F., Liegl, G., Obbarius, A., & Rose, M. (2021). A modified version of the transactional stress concept according to lazarus and folkman was confirmed in a psychosomatic inpatient sample. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.584333

Ozbay, F., Johnson, D. C., Dimoulas, E., Morgan, C. A., Charney, D., & Southwick, S. (2007). Social support and resilience to stress. *Psychiatry (Edgmont)*, *4*(5), 35–40. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2921311/

Quek, F. Y. X., Yang, H., & Hartanto, A. (2025). Neuroticism predicts constructive conflict resolution styles in conscientious individuals: Evidence from self-reports and observer ratings. *Journal of Research in Personality*, *117*, 104625. https://doi.org/10.1016/j.jrp.2025.104625

Rosmond, R., & Björntorp, P. (2000). The hypothalamic–pituitary–adrenal axis activity as a predictor of cardiovascular disease, type 2 diabetes and stroke. *Journal of Internal Medicine*, *247*(2), 188–197. https://doi.org/10.1046/j.1365-2796.2000.00603.x

Samuel, D. B., & Widiger, T. A. (2011). Conscientiousness and obsessive-compulsive personality disorder. *Personality Disorders: Theory, Research, and Treatment*, *2*(3), 161–174. https://doi.org/10.1037/a0021216

Sheng, J. A., Bales, N. J., Myers, S. A., Bautista, A. I., Roueinfar, M., Hale, T. M., & Handa, R. J. (2021). The hypothalamic-pituitary-adrenal axis: Development, programming actions of hormones, and maternal-fetal interactions. *Frontiers in Behavioral Neuroscience*, *14*, 601939. https://doi.org/10.3389/fnbeh.2020.601939

Stieger, M., Robinson, S. A., Bisson, A. N., & Lachman, M. E. (2020). The relationship of personality and behavior change in a physical activity intervention: The role of conscientiousness and healthy neuroticism. *Personality and Individual Differences*, *166*, 110224. https://doi.org/10.1016/j.paid.2020.110224

Turiano, N. A., Chapman, B. P., Gruenewald, T. L., & Mroczek, D. K. (2015). Personality and the leading behavioral contributors of mortality. *Health Psychology*, *34*(1), 51–60. https://doi.org/10.1037/hea0000038

Uchino, B. N. (2006). Social support and health: A review of physiological processes potentially underlying links to disease outcomes. *Journal of Behavioral Medicine*, *29*(4), 377–387. https://doi.org/10.1007/s10865-006-9056-5

Weston, S. J., Hill, P. L., Edmonds, G. W., Mroczek, D. K., & Hampson, S. E. (2019). No evidence of “healthy neuroticism” in the hawaii personality and health cohort. *Annals of Behavioral Medicine*, *53*(5), 426–441. https://doi.org/10.1093/abm/kay055

Weston, S. J., & Jackson, J. J. (2018). The role of vigilance in the relationship between neuroticism and health: A registered report. *Journal of Research in Personality*, *73*, 27–34. https://doi.org/10.1016/j.jrp.2017.10.005

Authors Bio:

My name is Natalie Koh and I am a BSc Psychology student at the University of Bristol. I am interested in personality and mental health, and in understanding why people can experience the same situation in very different ways. Throughout my studies, I have become curious about how individual differences shape the way people think, feel, and respond to challenges, influencing health and wellbeing. These interests are reflected in this paper, which explores the relationship between personality, stress, coping, and health outcomes. Beyond psychology, I enjoy travelling, crafts, and experimenting with recipes from different cuisines.

Functional EEG Effects of Short-Term Focused Attention Meditation: Do They Mirror Long-Term Structural Adaptations?

Author: Lucy Haun Eade

ORCID: 0009-0003-6706-691X

Programme: BSc Neuroscience

Affiliation: SchoolofPsychology and Neuroscience

Email: fp24485@bristol.ac.uk

Abstract

Focusedattention (FA) meditation involves sustaining attention on a specific object while disengaging from distractions when they arise. Long-term practice has been linked to improved attentional regulation and structural brain changes, reflecting neuroplasticity. This study investigated whether a brief FA meditation could induce measurable functional changes in neural activity related to attentional control, potentially mirroring benefits of long-term practice. Novice participants completed two conditions: a 5-minute guided FA meditation and a 5-minute silent control, with EEG recorded during each session. Immediately afterward, participants performed a 30-second Stroop task. We hypothesized that meditation would improve Stroop performance and modulate EEG activity. Results showed a directional trend supporting this, but no statistically significant differences were observed in Stroop performance or EEG measures. Our findings suggest that a brief FA meditation may be insufficient to produce measurable cognitive or neural changes comparable to those reported in long-term meditation practitioners. The absence of EEG differences further suggests that the session may not have induced a sufficiently strong meditation-related cognitive state. Overall, we suggest that while brief meditation may produce some transient functional effects, more stable attentional benefits are likely to depend on sustained meditation practice.

Keywords: EEG, meditation, attention, cognition, executive function

Introduction

Meditationisafamily of complex emotional and attentional regulatory training regimes (Lutz et al,2008). Long-term meditation practice has been linked to enhanced attentionalregulation (Van-Leeuwen et al, 2012). These effects are associated with structuralbrainchanges such as thicker cortical matter in brain regions associated with attentionand sensory processing that arise from sustained practice (Lazar et al, 2005). Short-term meditation programs are increasingly promoted in the workplace as a means of producing greater focus and concentration (Corporate Wellness Magazine, 2026). However, it remains unclear how well short-term meditation can functionally mirror the attentional benefits associated with structural changes in the brain in the absence of long-term practice outside structured sessions. EEG records synchronous electrical activity occuring in the cortex. Certain patterns of activity are linked to sustained attention, and previous research has linked beta2 power to focused executive processing and the processing of bottom-up sensory information (Razumnikova, 2007). The present study investigates a type of meditation known as focused attention (FA) which entails the voluntary focusing of attention on a chosen object (Lutz et al, 2008). During FA meditation, participants repeatedly practice monitoring the focus of attention, detecting distractions, disengaging from the source of distraction, and redirecting attention back to the intended object. Because FA meditation requires maintaining attention on a specific object while monitoring the broader field of awareness, increases in beta2 power have been associated with this sustained attentional engagement (Travis et al., 2010).

This study investigated whether a short, focused attention (FA) meditation could induce EEG-measured neural changes, specifically in beta power, resembling improved attentional control associated with long-term meditation practice. These functional changes were expected to be reflected in improved behavioural performance on an attention control task.

By comparing short-term functional effects with evidence from the existing literature on long-term structural changes, this study aims to clarify whether transient state-dependent modulation can produce meaningful cognitive benefits worthwhile promoting, or whether it is more important to emphasise sustained practice.

Participants with little or no prior meditation experience were recruited to examine whethera brief, five-minute FA meditation could elicit hypothesised changes. In additionto behavioural measures, neural correlates were recorded using EEG.

Method

Participants

Nine self-reported novice meditators (4 male, 5 female) were recruited from the University of Bristol (mean age = 20.56 years, SD = 1.13) to participate in the study. All participants reported being well rested the previous night, with a mean sleep duration of 6.75 hours.

Equipment and design

Raw EEG data were collected using MUSE MU-02 headbands (Interaxon Inc., Toronto, Canada) The Muse system records EEG activity using four electrodes- AF7, AF8, TP9, and TP10, which measure electrical activity from the frontal and temporal lobes of the brain. Participants also wore noise-cancelling headphones during both experimental conditions to minimise external auditory distractions.

All participants completed an identical experimental protocol across two experimental conditions:meditation and silence (control). A within-subjects design was used, with the order ofconditions counterbalanced across participants to control for order effects. Participantswere anonymised. During each condition, EEG activity was recorded continuouslyfor 5 minutes. Immediately following each EEG recording period, participantscompleted a digital replication of the Stroop test lasting 30 seconds. The Stroop testmeasures ability to selectively redirect attention so was chosen for this study. Taskperformance was quantified as the number of correct responses within the allotted timeperiod (Stroop, 1935).

Procedure

Participantswere seated in a quiet room and wore noise-cancelling headphones. A MUSE EEGheadband was fitted to the participant’s head and calibrated to ensure that the sensors maintained good contact with the forehead for accurate signal acquisition. The headband was connected via Bluetooth to a tablet running the Mind Monitor EEG recording software. Participants were instructed to close their eyes and either listened to a 5-minute guided meditation recording where they were instructed

to focus on their breath, (Hands-on Meditation, 2023) or sat in silence for five minutes. Immediately following the five-minute period, EEG recording was stopped and participants removed the headphones. They then completed a Stroop task. After 30 seconds, the task was stopped and the total number of correct responses was recorded. Participants then undertook the alternate condition and repeated the same procedure.

Data processing and analysis

Raw EEG data collected usingthe MUSE headband were uploaded to the Mind Monitor software for analysis.The Mind Monitor software calculated average power (in Bels) for 5 EEG frequencybands (Alpha, Beta, Theta, Delta, Gamma) across the 5-minute recording period foreach participant. Mean power values were compared across experimental conditionsfor each participant.

Data was statistically analysedusing R coding software (R Core Team, 2025) Due to the small sample size (n = 9) and the absence of a strong assumption of normality, a one-tailed Wilcoxon signed-rank test was conducted to compare Stroop test scores between the two conditions. Additionally, two-tailed Wilcoxon signed-rank tests were performed on EEG data for each frequency band to assess whether significant differences were observed between conditions.

Ethics

All participants provided informed consent before taking part in the study. To protect participant confidentiality, identifying information was removed from the dataset and participants were assigned numerical codes, meaning that individuals could not be identified from the reported data. Participants were also informed that their participation was voluntary and that they could withdraw from the study at any time without consequence.

Results

Across all frequency bands, individual responses varied in direction, indicating no consistent effect of meditation on EEG power relative to silence.

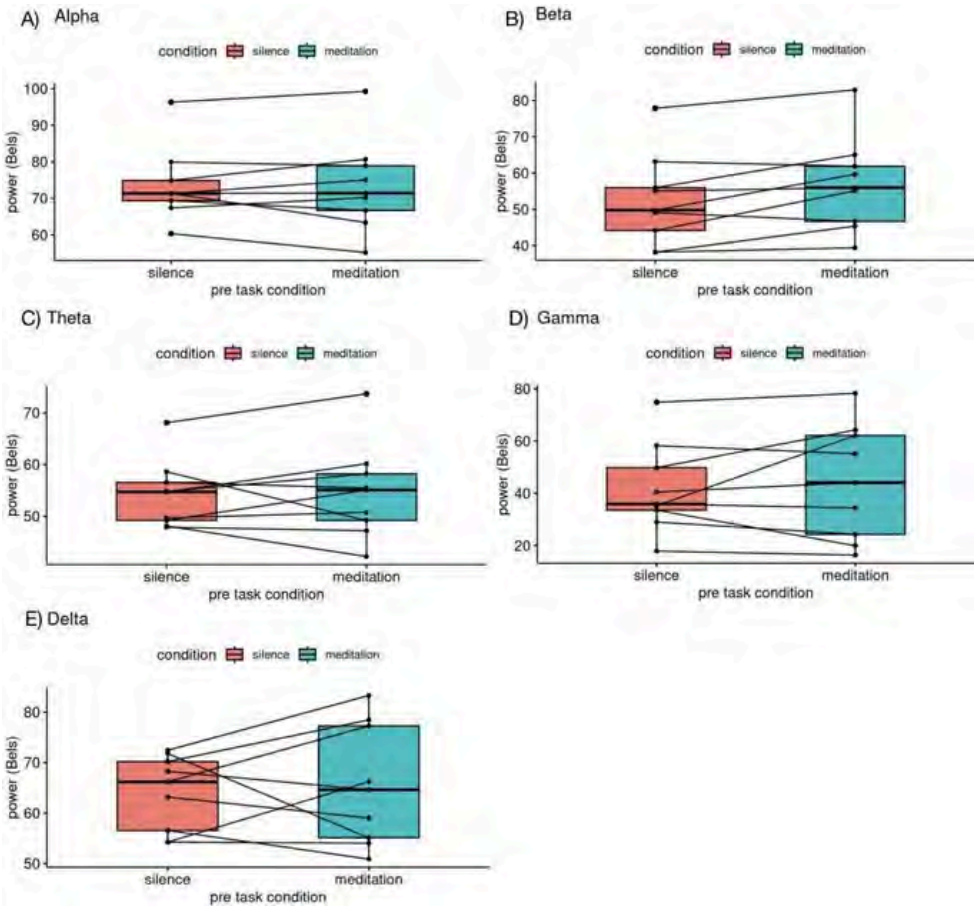


Figure 3. ***EEG power in Bels during the silence (control) and meditation conditions across frequency bands (n = 9). Panels show alpha (a), beta (b), theta (c), gamma (d), and delta (e) power, with individual participant values illustrating within-subject changes between conditions.***

Paired Wilcoxon signed-rank tests were conducted to compare EEG power between the meditation and silence conditions across frequency bands (theta, alpha, beta, gamma, and delta). There was no significant difference in theta power between conditions, (W = 20, p = 0.82 (two-tailed), n = 9). Similarly, alpha power did not differ significantly between meditation and silence, (W = 24, p = 0.91 (two-tailed), n = 9).

No significant differences were observed for beta power, (W = 38, p = 0.074 (two-tailed), n = 9), although this comparison showed a non-significant trend towards a condition-related difference. Gamma power also did not differ significantly between conditions, (W = 26, p = 0.73 (two-tailed), n = 9), nor did delta power, (W = 26, p = 0.73 (two-tailed), n = 9) (Figure 3).

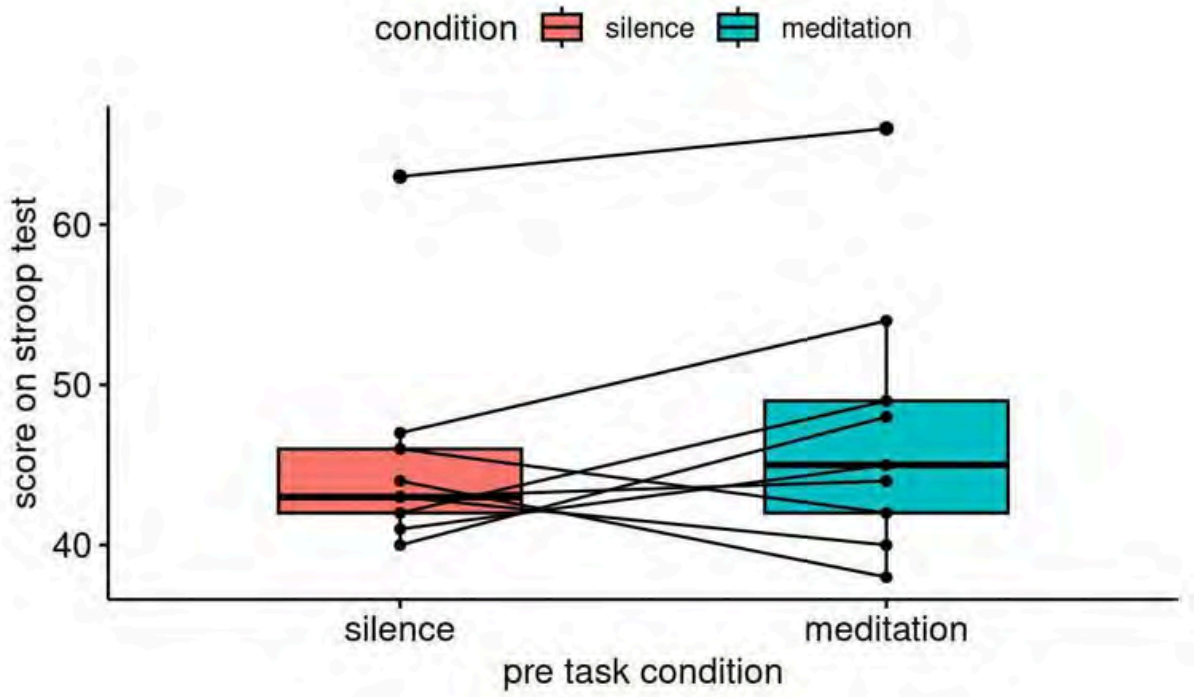


Figure 2. ***Boxplot showing change in the number of correct Stroop task responses between the silence (control) and meditation conditions (n = 9), with individual participant data illustrating within-subject differences.***

A Wilcoxon signed-rank test was conducted to examine whether task performance increased following meditation compared to silence. Although median scores were higher in the meditation condition, this difference did not reach statistical significance, W = 32, p = 0.1425 (one-tailed), n = 9. As illustrated in Figure 2, the trend was in the hypothesised direction, suggesting a potential improvement in performance following meditation that may warrant investigation in a larger sample.

Discussion

We found that the change in Stroop task performance between the silence and meditation conditions demonstrated a trend in the hypothesised direction, however, no statistically significant differences were observed. This may suggest that short-term meditation cannot fully reproduce the robust cognitive benefits observed in long-term practitioners (MacLean et al, 2010). Instead, the strongest benefits may arise only after structural reorganisation of neural networks associated with attentional control, which are thought to develop through sustained meditation practice (Lazar et al., 2005).

A recent study reported contrasting findings. Using a highly similar experimental design, the authors observed statistically significant improvements in Stroop test performance immediately following a single meditation session (Sleimen-Malkoun et al., 2023). These findings contrarily suggest that meditation can produce short-term functional effects on attentional control even in the absence of long-term training. The discrepancy between these results and those of the present study may therefore be attributable to methodological differences, most notably sample size. Although the current results showed a directional change consistent with the hypothesis, the effect may not have been large enough to reach statistical significance within the relatively small sample.

Alternatively, it is possible that participants in the present study had the capacity to experience a state of enhanced attentional ability but did not enter a sufficiently strong meditation-related cognitive state during the intervention. Supporting this interpretation, no significant differences were observed in EEG power across any frequency band between conditions, suggesting that the meditation may not have produced a measurable state-dependent neural change. This may reflect difficulties in participants fully engaging with the guided meditation instructions. As the instructions were delivered simultaneously with the meditation procedure, participants may not have fully understood or internalised them. A potential improvement would therefore be to provide participants with a brief explanation of the meditation technique prior to the guided session, allowing them to familiarise themselves with the instructions before the intervention begins.

It is important to distinguish between the lasting effects of a short-term induced state and the long-term trait effects of meditation. Numerous studies comparing experienced meditators with non-meditators report superior behavioural performance on attentional tasks in experienced practitioners when there is no immediate meditation session beforehand (Luo et al, 2024; Jo et al, 2016). This suggests that sustained meditation practice may be superior in its ability to produce more stable attentional benefits that persist beyond the immediate meditation state. In contrast, short-term meditation sessions may only produce transient functional changes that are most evident when the meditation occurs immediately prior to the cognitive task. While such short-term effects may be useful in contexts requiring immediate attentional focus, such as before

an examination, they may be less practical in environments such as the workplace where consistent cognitive performance is required over extended periods.

EEG analysis

Previous research has linked beta2 power to focused executive processing and the processing of bottom-up sensory information (Razumnikova, 2007). Interestingly, beta power in our present study did show a directional trend, with beta activity tending to increase following the meditation condition. Although this difference did not reach statistical significance, it is possible that a larger sample size or clearer instruction prior to the intervention may have produced a detectable effect.

In conclusion, the present study investigated whether a brief session of focused attention meditation could produce measurable cognitive and neural changes associated with attentional control. Although a directional trend in Stroop task performance was observed following the meditation condition, no statistically significant behavioural or EEG differences were detected. These findings suggest that short-term meditation may be insufficient to reliably induce measurable cognitive or neural changes in attentional control. However, as a potential limitation, the absence of significant EEG effects may also indicate that participants did not enter a sufficiently strong meditation-related cognitive state during the session.

Our results support a distinction between transient state effects and more stable trait effects of meditation. While brief meditation may produce short-term functional changes under certain conditions, the more robust attentional benefits reported in experienced meditators are likely to depend on sustained practice and longer-term neural adaptation. Future research would benefit from larger sample sizes, clearer pre-meditation instruction, and further investigation into psychological factors that may influence attentional performance, such as how long-term meditation practice shapes beliefs about one’s attentional abilities, which may in turn influence task performance.

References

Jo, H.-G. et al. (2016) ‘Meditation and attention: A controlled study on long-term meditators in behavioural performance and event-related potentials of attentional control’*, International Journal of Psychophysiology*, 99, pp. 33–39. DOI: 10.1016/j.ijpsycho.2015.11.016.

Lazar, S.W. et al. (2005) ‘Meditation experience is associated with increased cortical thickness’, *NeuroReport*, 16(17), pp. 1893–1897. DOI: 10.1097/01.wnr.0000186598.66243.19.

Luo, X. et al. (2024) ‘Differences in attentional function between experienced mindfulness meditators and non-meditators’, *Frontiers in Psychiatry*, 15. DOI: 10.3389/fpsyt.2024.1341294.

Lutz, A. et al. (2008) ‘Attention regulation and monitoring in meditation’, *Trends in Cognitive Sciences*, 12(4), pp. 163–169. DOI: 10.1016/j.tics.2008.01.005.

MacLean, K.A. et al. (2010) ‘Intensive meditation training improves perceptual discrimination and sustained attention’, *Psychological Science,* 21(6), pp. 829–839. DOI: 10.1177/0956797610371339.

R Core Team (2025). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: https://www.r-project.org/ (Accessed: 10 March 2026)

Razumnikova, O.M. (2007) ‘Creativity related cortex activity in the Remote Associates Task’, *Brain Research Bulletin*, 73(1–3), pp. 96–102. DOI: 10.1016/j.brainresbull.2007.02.008.

Sleimen-Malkoun, R., Devillers-Réolon, L. and Temprado, J.-J. (2023) ‘A single session of mindfulness meditation may acutely enhance cognitive performance regardless of meditation experience’, *PLOS ONE,* 18(3). DOI: 10.1371/journal.pone.0282188.

Stroop, J.R (1935). Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology*, [online] 18(6), pp.643-662. DOI: 10.1037/h0054651. Travis, F. and Shear, J. (2010) ‘Focused attention, open monitoring and automatic self-transcending: Categories to organize meditations from Vedic, Buddhist and

Chinese traditions', *Consciousness and Cognition*, 19(4), pp. 1110–1118. DOI:

10.1016/j.concog.2010.01.007.

van-Leeuwen, S., Singer, W. and Melloni, L. (2012) 'Meditation increases the depth of information processing and improves the allocation of attention in space', *Frontiers in Human Neuroscience*, 6. DOI: 10.3389/fnhum.2012.00133.

Corporate Wellness Magazine, *"The power of mindfulness: Incorporating meditation into the Workplace"*, Available at: https://www.corporatewellnessmagazine .com/article/the-power-of-mindfulness-incorporating-meditation-into-the-workplace (Accessed: 10 March 2026)

Hands-on Meditation,*"BoxBreathingRelaxationExercise|5MinutesBeginnerPace | Anxiety Reduction Pranayama Technique",* Available at: https://www.youtube.com /watch?v=oN8xV3Kb5-Q (Accessed: 10 March 2026)

Authors Bio:

My name is Lucy Haun Eade. I am a Neuroscience student at the University of Bristol, and I am incredibly interested in meditation and the effects it has on the brain. I am also particularly interested in how attending to Buddhist practice changes perceptions of what is real both permanently and during meditative states.

Serum Hydrogen Sulphide and Nitric Oxide Levels in Treated and Drug-Naive Patients with Schizophrenia: A Comparative Cross-Sectional Study

Authors: Atrim Das, Anamika Chakravarty

ORCID: 0009-0000-6441-3769

Programme: Molecular Neuroscience MSc

Affiliation: SchoolofPhysiology, Pharmacology and Neuroscience

Email: qf25081@bristol.ac.uk

Abstract

Schizophrenia is a chronic psychiatric disorder with a complex and unresolved neurobiological basis. Recent evidence implicates endogenous gasotransmitters, specifically hydrogen sulphide (H_2S) and nitric oxide (NO), in the disorder's pathophysiology, due to their modulatory effects on N-methyl-D-aspartate (NMDA) receptors and intracellular second-messenger systems. To investigate this relationship, a cross-sectional study was conducted at North Bengal Medical College comparing serum H_2S and NO concentrations in 100 patients diagnosed with schizophrenia. The cohort was evenly divided into antipsychotic-treated (n=50) and drug-naive (n=50) groups. Clinical symptom severity was quantified using the Positive and Negative Syndrome Scale (PANSS) and the Brief Psychiatric Rating Scale (BPRS). Biochemical analysis revealed significant deficits in both gasotransmitters among untreated patients. Mean serum H_2S levels were substantially lower in the drug-naive group (25.42 µmol/L) compared to the treated group (148.30 µmol/L; $p<0.001$). A parallel trend was observed for serum NO, with mean concentrations of 11.00 µmol/L in drug-naive patients versus 37.74 µmol/L in those receiving pharmacotherapy ($p<0.001$). Furthermore, Spearman correlation analysis demonstrated a strong inverse relationship between the concentrations of both gasotransmitters and psychiatric rating scores. These objective reductions in H2S and NO directly scale with increased symptom severity, suggesting their potential utility as accessible peripheral biomarkers for monitoring disease progression and therapeutic response in schizophrenia.

Keywords schizophrenia, hydrogen sulphide, nitric oxide, gasotransmitters, NMDA receptor, biomarker

Glossary of Abbreviations

BPRS: Brief Psychiatric Rating Scale

CBS: Cystathionine beta-synthase

cAMP: Cyclic adenosine monophosphate

cGMP: Cyclic guanosine monophosphate

CSE: Cystathionine gamma-lyase

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition

H2S: Hydrogen sulphide

LTP: Long-term potentiation

NBMCH: North Bengal Medical College and Hospital

NMDA: N-methyl-D-aspartate

NMDAR: NMDA receptor

NO: Nitric oxide

NOS: Nitric oxide synthase

nNOS: Neuronal nitric oxide synthase

PANSS: Positive and Negative Syndrome Scale

PKA: Protein kinase A

SD: Standard deviation

3-MST: 3-Mercaptopyruvate sulfurtransferase

1. Introduction

Schizophrenia is a profoundly disabling psychiatric condition affecting approximately 24 million people worldwide. This translates to roughly 1 in 300 individuals globally and 1 in 222 adults (World Health Organisation, 2022). Within India, the current prevalence is estimated at 0.5%, with a lifetime prevalence of 1.4% (Hegde et al., 2023). Despite these relatively low prevalence rates, the disorder ranks among the top 25 global causes of disability. The clinical profile encompasses positive symptoms such as hallucinations and delusions, negative symptoms including affective flattening and avolition, and pervasive cognitive deficits. Because no reliable biological marker currently exists for routine clinical use, schizophrenia remains entirely a clinical diagnosis. The underlying neurobiology involves significant disruptions across dopaminergic, serotonergic, and glutamatergic pathways. Researchers have increasingly focused on the hypofunction of the N-methyl-D-aspartate receptor (NMDAR), a glutamate receptor subtype essential for synaptic plasticity, learning, and memory. Pharmacological models strongly support this mechanism. NMDAR antagonists like ketamine and phencyclidine successfully induce both the positive and negative symptoms of schizophrenia in healthy subjects (Bressan and Pilowsky, 2003). Emerging evidence highlights two endogenous gaseous signalling molecules, hydrogen sulphide (H_2S) and nitric oxide (NO), as critical modulators of NMDAR function. Originally viewed strictly as an environmental toxin, H_2S is now recognised as a physiological gasotransmitter synthesised in the brain primarily by cystathionine beta-synthase (CBS). It amplifies NMDAR-mediated responses through the cAMP/PKA signalling cascade and regulates intracellular calcium dynamics via multiple second messenger pathways (Abe and Kimura, 1996; Kimura, 2000). NO is generated from L-arginine by nitric oxide synthase (NOS) after NMDAR activation. It functions as a retrograde messenger, influencing neurotransmitter release and synaptic plasticity through the cGMP pathway (Knowles and Moncada, 1994). [Fig.1]. Clinical investigations have documented reduced plasma H_2S in patients with schizophrenia compared to healthy controls, noting that lower concentrations correlate with more severe cognitive impairment (Xiong et al., 2018). Similar findings exist for NO, with depleted metabolites observed in drug-naive patients and subsequent normalisation following antipsychotic therapy (Lee and Kim, 2008; Nakano et al., 2010). There is a distinct lack of research directly

comparing the systemic levels of both gasotransmitters between pharmacologically treated and drug-naive cohorts, particularly within Indian demographic groups. The present study is designed to address this specific clinical gap.

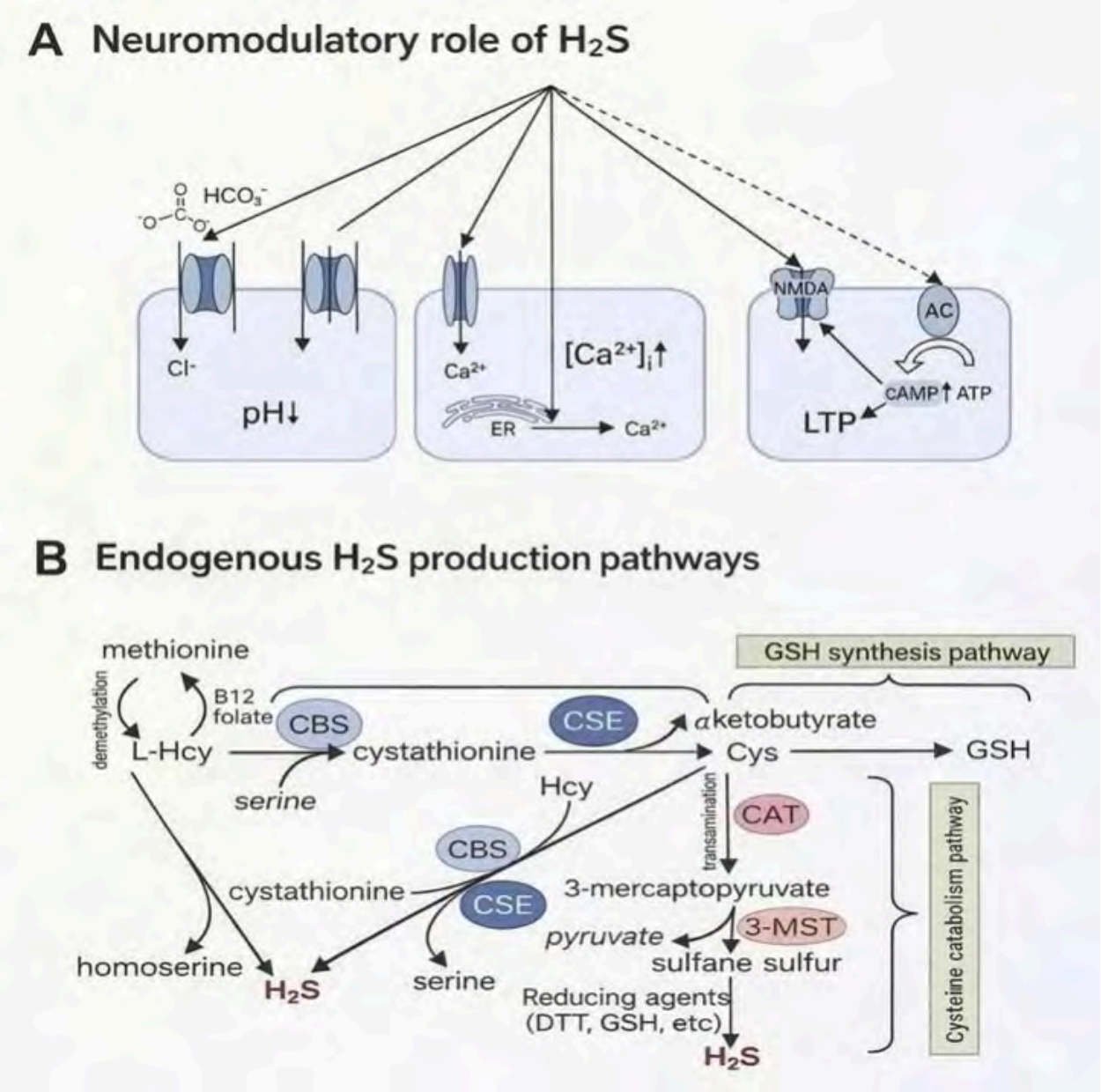


Figure 1. *Biosynthesis and neuromodulatory pathways of hydrogen sulphide (H_2S): (A) H2S modulates neural activity through three distinct second messenger systems. It enhances NMDAR phosphorylation via the cAMP/PKA pathway, drives hippocampal long-term potentiation (LTP) through the Ca2+/NMDAR pathway, and regulates calcium balance and neuroinflammation via the PLC/PKC pathway. (B) H_2S is generated from L-cysteine and homocysteine via CBS in the brain, CSE in peripheral tissues, and the CAT/3-MST pathway. The gas is subsequently oxidised, utilised for protein sulfhydration, or stored as polysulfides.*

In the context of schizophrenia, impaired trans-sulfuration decreases H_2S production. This deficiency directly contributes to NMDAR hypofunction and corresponding cognitive deficits.(CBS = cystathionine beta-synthase; CSE = cystathionine gamma-lyase; 3-MST = 3-mercaptopyruvate sulfurtransferase; LTP = long-term potentiation).

2. Methods

2.1 Study Design and Setting

This research is a cross-sectional analytical study at the Department of Psychiatry within North Bengal Medical College and Hospital in Darjeeling, India. The hospital's Institutional Ethics Committee formally approved the protocol, and the study was registered with the Central Drugs Standard Control Organisation. All participants or their designated legal guardians provided written informed consent before any clinical evaluation.

2.2 Participants

The study consecutively enrolled 100 patients from the outpatient and inpatient psychiatric departments. The cohort was divided evenly into 50 patients actively receiving antipsychotic treatment and 50 drug-naive patients. The drug-naive classification was strictly defined to include individuals who had either never taken antipsychotic medication or had completely discontinued their pharmacotherapy for a minimum of six consecutive weeks. Two senior psychiatry faculty members independently confirmed each diagnosis using standard DSM-5 criteria. To control for confounding variables, participants were required to be between 18 and 60 years of age. Any subject presenting with comorbid medical or psychiatric conditions, active substance use disorders, or a refusal to provide consent was excluded. Both groups were matched for age and gender to maintain statistical integrity.

2.3 Clinical Assessment

Clinicians quantified symptom severity using two validated psychiatric rating instruments. The Positive and Negative Syndrome Scale (PANSS) was used to evaluate 30 distinct clinical items spanning positive symptoms, negative symptoms, and general psychopathology (Kay et al., 1987). Each item was graded on a 7-point scale, generating a total severity score ranging from 30 to 210, where higher numerical values indicate more severe psychiatric illness. Additionally, the medical team utilised the Brief Psychiatric Rating Scale (BPRS) to assess 18 specific symptom constructs, including hostility, suspiciousness, and hallucinatory behaviour (Overall and Gorham, 1962). Clinicians ranked these symptoms from 1, indicating

the complete absence of a symptom, to 7, representing extreme severity. The BPRS provided a focused, complementary metric for establishing the patients' clinical baselines and evaluating their response to therapeutic interventions.

2.4 Biochemical Measurements

Quantification of serum hydrogen sulphide (H_2S) was performed using a spectrophotometric methylene blue method. To isolate the sulphide species, zinc acetate was added to the serum samples to induce precipitation. After re-dissolving this precipitate, N, N-dimethyl-p-phenylenediamine and ferric chloride were introduced to generate methylene blue, which was then measured at a wavelength of 670 nm using a UV-Vis spectrophotometer. The final H_2S concentrations were recorded in µmol/L. Because serum nitric oxide (NO) is highly unstable, it was measured indirectly through its stable metabolites, nitrate and nitrite, using the Griess reaction. The protocol required reducing nitrate to nitrite utilising a cadmium catalyst. Under acidic conditions, the nitrite reacted with sulfanilamide to form a diazonium ion, which subsequently coupled with N-(1-naphthyl)ethylenediamine. This chemical interaction produced a pink azo dye that was quantified colourimetrically at 540 nm using the same UV-Vis spectrophotometer. Personnel in the biochemistry laboratory conducted all assays while remaining completely blinded to the patients' clinical data.

2.5 Statistical Analysis

All statistical analyses were executed using IBM SPSS version 21. The Kolmogorov-Smirnov test was applied to evaluate data normality. Given that the serum concentrations of H2S and NO did not follow a normal distribution, the non-parametric Mann-Whitney U test compared the rank-ordered values between the treated and drug-naive groups. For the clinical severity metrics (BPRS and PANSS), independent-samples t-tests were appropriate for making group comparisons. To assess the relationship between gasotransmitter concentrations and psychiatric symptom severity, Spearman's rank correlation coefficient was calculated. The threshold for statistical significance was established at $p < 0.05$ for all two-tailed tests.

3. Results

3.1 Sample Characteristics

The final study cohort included 100 participants evenly divided into two testing groups. The treated group consisted of 23 males and 27 females with a mean age of 33.58 years (SD = 10.22). The drug-naive group included 22 males and 28 females

with a mean age of 33.94 years (SD = 10.36). Statistical testing confirmed no significant baseline differences between the cohorts regarding age (p = 0.870), gender (p = 0.84), religion, family structure, marital status, educational attainment, or socioeconomic background. Demographically, 70% of the participants resided in rural areas, 71% belonged to lower socioeconomic strata (Class V), and 46% had completed primary-level education. The duration of the illness was significantly longer in the treated cohort (mean 40.46 months, SD = 18.76) compared to the drug-naive cohort (mean 28.30 months, SD = 20.44; p = 0.0023). This variance aligns with the clinical reality that treated patients generally possess a longer psychiatric history before stabilisation.

3.2 Symptom Severity

Clinical evaluations revealed a significantly heavier symptom burden among the drug-naive patients across both diagnostic scales. The mean BPRS score reached 70.14 (SD = 12.60) in the untreated group, compared to 38.30 (SD = 14.82) in the treated group (p < 0.001). Similarly, the mean PANSS total score was 126.06 (SD = 21.08) for drug-naive patients versus 70.34 (SD = 25.89) for those receiving pharmacotherapy (p < 0.001).

3.3 Serum H_2S and NO Levels

Biochemical analysis uncovered a profound deficit in gasotransmitter concentrations among the untreated individuals. Serum H_2S levels were nearly six times lower [Figure 2] in drug-naive patients (mean 25.42 µmol/L, SD = 30.33) than in treated patients (mean 148.30 µmol/L, SD = 71.09; p < 0.001). Serum NO followed an identical pattern, with drug-naive patients presenting a mean concentration of 11.00 µmol/L (SD = 9.86) against the treated group's 37.74 µmol/L (SD = 22.34; p < 0.001). Beyond the distinct group comparisons, the baseline data demonstrated a significant inverse association between H_2S levels and PANSS scores. Patients presenting with the lowest circulating H_2S levels consistently exhibited the most severe psychiatric symptomatology.

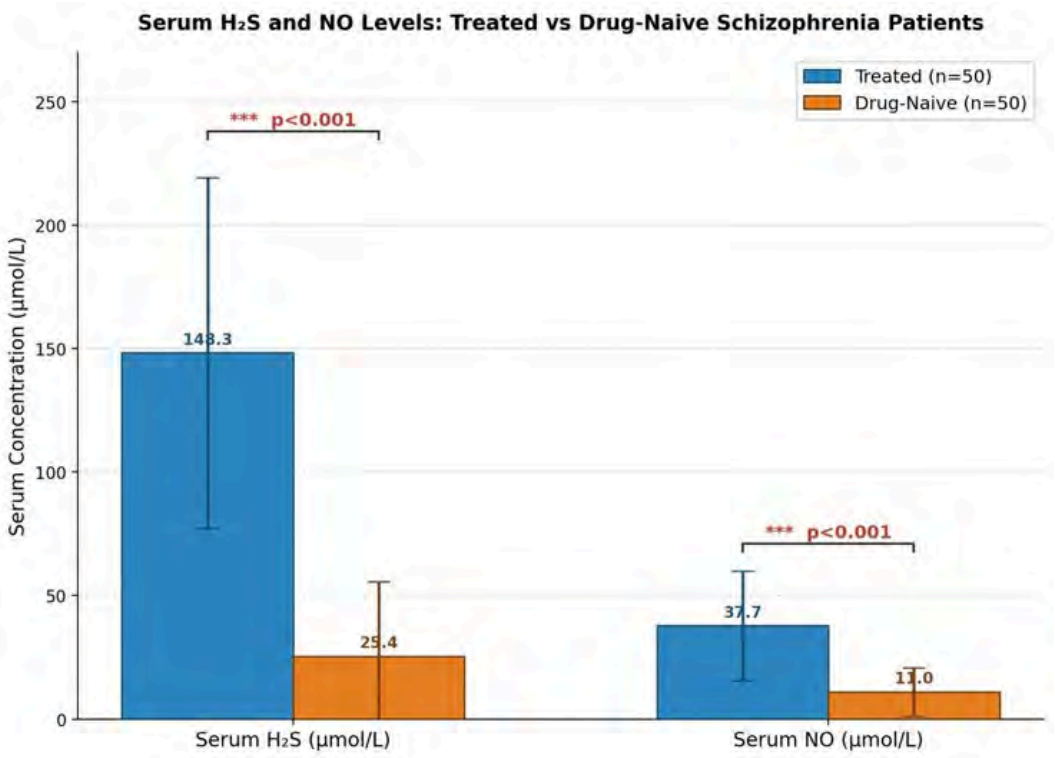


Figure 2. *Serum H_2S and NO levels (mean + SD) in treated (n=50) and drug-naive (n=50) schizophrenia patients. Both gasotransmitters were significantly lower in drug-naive patients (Mann-Whitney U test; *** p<0.001 for both comparisons). Error bars represent one standard deviation.*

3.4 Correlations with Symptom Severity

To determine how these biochemical deficits translate to clinical presentation, Spearman correlation analysis was applied across the entire patient cohort [Figure 3]. The data revealed a robust inverse relationship between circulating gasotransmitter concentrations and psychiatric symptom severity. Specifically, H_2S levels demonstrated a strong negative correlation with both the BPRS (rho = -0.614, $p < 0.001$) and the total PANSS score (rho = -0.602, $p < 0.001$). Nitric oxide mirrored this pattern, exhibiting similarly strong negative correlations with the BPRS (rho = -0.653, $p < 0.001$) and the PANSS (rho = -0.634, $p < 0.001$). Clinically, this inverse trajectory indicates that as systemic levels of H_2S and NO decline, patients experience progressively higher severity scores and more pronounced psychopathology. All identified correlations reached high statistical significance ($p < 0.001$, two-tailed).

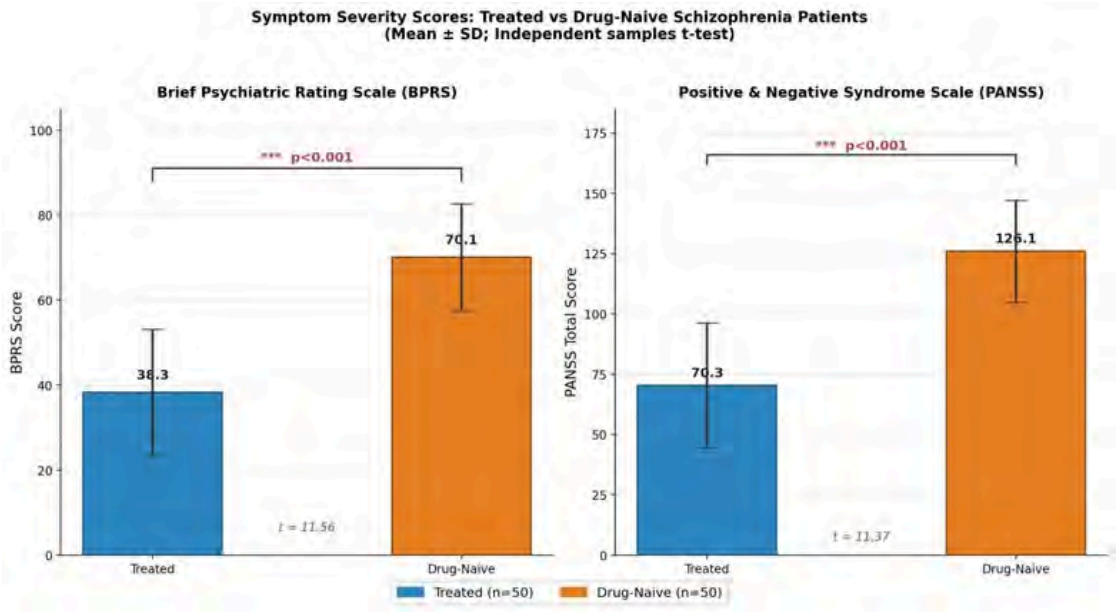


Figure 3. *Representative scatterplot visualisation of the inverse associations between serum H_2S and NO levels and symptom severity scores across all 100 participants. Lower gasotransmitter levels were associated with higher BPRS and PANSS scores, with all reported Spearman correlations significant at $p < 0.001$.*

4. Discussion

The results of this study provide compelling comparative evidence that systemic deficienciesinhydrogen sulphide (HS_2)and nitric oxide (NO)are fundamentally tied to the neurobiology of untreated schizophrenia. By demonstrating that both gasotransmitters are substantially reduced in drug-naive patients and partially restored in those receiving antipsychotic therapy, the data strongly support the hypothesis that gasotransmitter dysregulation is a core component of the disorder's pathophysiology. The pronounced inverse correlation between these molecular levels and clinical symptom severity further underscores their biological and diagnostic significance.

The magnitude of the H_2S deficit is particularly striking, revealing a nearly sixfold reduction inthe drug-naive cohort. Whilepreviousresearch byXiong et al. (2018) identified lower plasma H_2S in patients compared to healthy individuals, the current study strengthens this foundation by confirming the deficit through a direct head-to-head comparison of treated versus untreated clinical populations. This approach effectively isolates the impact of the illness and its pharmacological management. Biologically, this severe reduction aligns with established impairments in the brain's trans-sulfuration pathway involving the CBS and CSE enzymes (Berry et al., 2020).

Because H_2S is required to drive cAMP/PKA-mediated NMDAR phosphorylation and facilitate hippocampal long-term potentiation, its depletion provides a clear, mechanistic explanation for the severe cognitive and psychotic symptoms observed in untreated patients (Abe and Kimura, 1996; Kimura, 2000).

The nitric oxide findings are equally impactful, validating and significantly expanding upon earlier literature. Prior studies have shown depleted NO metabolites in unmedicated patients and corresponding increases following targeted treatments like risperidone (Lee and Kim, 2008; Nakano et al., 2010). By evaluating a larger patient cohort and applying comprehensive dual assessments with both the BPRS and PANSS, this study robustly confirms that antipsychotic intervention effectively restores systemic NO levels. This restorative effect likely stems from the modulation of the NMDA-NO-cGMP signalling axis. As antipsychotics suppress dopaminergic hyperactivity in the mesolimbic pathway, they indirectly normalise glutamatergic transmission and downstream nitric oxide synthase activity, a mechanism supported by extensive meta-analyses in the field (Maia-de-Oliveira et al., 2012; Oliveira et al., 2008).

The clinical relevance of these biochemical shifts is evident in the strong negative correlations identified between both gasotransmitters and psychiatric severity. With correlation coefficients consistently demonstrating robust inverse relationships (rho < −0.60), the data confirm that serum H_2S and NO concentrations reliably track the overall psychopathological burden. The most severely ill patients consistently presented with the lowest gasotransmitter reserves across two independent, validated rating scales. As illustrated [Figure 3], the individual data points for all 100 participants reveal a clear dose-dependent biological gradient: as gasotransmitter concentrations decline, BPRS and PANSS scores rise systematically, confirming that this inverse relationship is not driven by outliers but reflects a robust, continuous association across the full spectrum of symptom severity. The correlation with nitric oxide was marginally stronger than with H_2S, which likely reflects NO's immediate function as a second messenger following NMDAR activation and its dense integrative role across both dopaminergic and serotonergic networks (Lorrain and Hull, 1993). While the findings present a strong case for the clinical utility of these biomarkers, several methodological considerations must be acknowledged to contextualise the results appropriately.

First, the composition of the drug-naive cohort introduces a degree of heterogeneity. The group included both individuals who had never received antipsychotic medication and those who had undergone a minimum six-week pharmacological washout. Although a six-week washout is a widely accepted clinical standard, it

cannot guarantee complete pharmacodynamic clearance, particularly for depot formulations or in patients with altered drug metabolism. This heterogeneity may have attenuated the true magnitude of gasotransmitter depletion in the untreated state, meaning the observed differences between groups could, if anything, be an underestimate of the biological effect in genuinely treatment-naive patients.

Second, the cross-sectional design, while appropriate for establishing associations, cannot resolve the direction of causality. It remains unclear whether reduced H_2S and NO concentrations are a primary neurobiological feature that predisposes individuals to more severe psychopathology, or whether they are a downstream consequence of the disease process itself. Establishing causality would require prospective longitudinal studies that track gasotransmitter levels from illness onset through treatment initiation, ideally with serial sampling at defined pharmacological milestones.

Third, several potential confounding variables were not systematically controlled. Dietary factors, particularly sulphur-containing amino acid intake from protein-rich foods, directly influence endogenous H_2S biosynthesis via the trans-sulfuration pathway. Similarly, tobacco smoking is a well-documented modulator of both NO and H_2S metabolism, as nicotine upregulates neuronal nitric oxide synthase activity, and smokers exhibit altered serum NO profiles independent of psychiatric status. Given that the study population was predominantly rural and of lower socioeconomic status, dietary patterns and smoking prevalence may have varied systematically between groups, introducing unmeasured confounding. Future studies should record and stratify by smoking status and dietary protein intake to isolate the psychiatric signal from these physiological variables.

A further interpretive consideration concerns the compartmental validity of peripheral measurements. This study quantified gasotransmitter concentrations in serum, which reflects systemic rather than central nervous system levels. H_2S and NO are highly reactive, short-lived molecules whose concentrations in the brain and cerebrospinal fluid may not be faithfully mirrored by peripheral blood. The blood-brain barrier selectively regulates the passage of gaseous signalling molecules and enzymatic activity in peripheral tissues, including erythrocytes, vascular endothelium, and hepatocytes, which can independently modulate serum concentrations in ways that are decoupled from central neurochemistry. Consequently, while the serum deficits observed here are biologically plausible and consistent with the NMDAR hypofunction hypothesis, they should be interpreted as a peripheral correlate of central dysregulation rather than a direct measure of brain gasotransmitter activity. Future studies incorporating cerebrospinal fluid sampling or neuroimaging-based

proxies of NMDAR function would provide stronger mechanistic evidence for the central relevance of these findings. The broader challenge of validating peripheral biomarkers for psychiatric disorders warrants explicit consideration. Unlike biomarkers for many somatic conditions, psychiatric biomarkers must contend with the absence of a definitive neuropathological gold standard against which peripheral measurements can be calibrated. Schizophrenia is a clinically and biologically heterogeneous syndrome, and gasotransmitter dysregulation represents one of several converging pathophysiological mechanisms rather than a singular, disease-defining abnormality. The sensitivity and specificity of serum H_2S and NO as diagnostic markers have not been formally evaluated in this study, and the absence of a healthy control group means that the observed values cannot be benchmarked against normative ranges. Furthermore, the overlap in gasotransmitter concentrations between groups, as evidenced by the standard deviations reported, suggests that individual-level diagnostic utility would be limited without further refinement of reference intervals and cut-off thresholds. These caveats do not diminish the scientific value of the findings, but they underscore that translation from a research biomarker to a clinically actionable test requires prospective validation in independent cohorts with healthy controls and standardised pre-analytical protocols. Moving forward, longitudinal investigations incorporating healthy baseline controls, standardised treatment protocols, cerebrospinal fluid sampling where feasible, and systematic recording of lifestyle confounders will build directly upon this foundation, further refining our understanding of how these critical signalling molecules dictate the trajectory of schizophrenia.

5. Conclusion

This study demonstrates that serum H_2S and NO are markedly reduced in drug-naive schizophrenia patients compared with antipsychotic-treated patients, and that both gasotransmitters correlate strongly and negatively with symptom severity. Serum H_2S and NO are minimally invasive, inexpensive, and measurable in resource-limited settings, making them promising candidates as biochemical markers of disease severity and treatment response. Replication in larger, longitudinal studies with healthy controls and standardised treatment protocols is needed before clinical translation.

References

Abe, K., & Kimura, H. (1996). The possible role of hydrogen sulfide as an endogenous neuromodulator. *Journal of Neuroscience*, 16(3), 1066–1071. https://doi.org/10.1523/JNEUROSCI.16-03-01066.1996

Berry, T., Abohamza, E., & Moustafa, A. A. (2020). Treatment-resistant schizophrenia: Focus on the transsulfuration pathway. *Reviews in the Neurosciences*, 31(2), 219–232. https://doi.org/10.1515/revneuro-2019-0057

Bressan, R. A., & Pilowsky, L. S. (2003). Glutamatergic hypothesis of schizophrenia. *Revista Brasileira de Psiquiatria*, 25(3), 177–183. https://doi.org/10.1590/S1516-44462003000300011

Hegde, P. R., Nirisha, L. P., Basavarajappa, C., et al. (2023). Schizophrenia spectrum disorders in India: A population-based study. *Indian Journal of Psychiatry*, 65, 1223–1229. https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_50_23

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13(2), 261–276. https://doi.org/10.1093/schbul/13.2.261

Kimura, H. (2000). Hydrogen sulfide induces cyclic AMP and modulates the NMDA receptor. *Biochemical and Biophysical Research Communications*, 267(1), 129–133. https://doi.org/10.1006/bbrc.1999.1916

Knowles, R. G., & Moncada, S. (1994). Nitric oxide synthases in mammals. *Biochemical Journal*, 298, 249–258. https://doi.org/10.1042/bj2980249

Lee,B.H.,&Kim,Y.K.(2008).Reducedplasmanitricoxidemetabolitesbeforeand after antipsychotic treatment in patients with schizophrenia compared to controls. *Schizophrenia Research*, 104, 36–43. https://doi.org/10.1016/j.schres.2008.05.013

Lorrain, D. S., & Hull, E. M. (1993). Nitric oxide increases dopamine and serotonin release in the medial preoptic area. *Neuroreport*, 5, 87–89. https://doi.org/10.1097/00001756-199310000-00023

Maia-de-Oliveira, J. P., Trzesniak, C., Oliveira, I. R., et al. (2012). Nitric oxide plasma/serum levels in patients with schizophrenia: A systematic review and meta-analysis. *Revista Brasileira de Psiquiatria*, 34, 149–162. https://doi.org/10.1016/j.rbp.2012.07.002

Nakano, Y., Yoshimura, R., Nakano, H., et al. (2010). Association between plasma nitric oxide metabolites levels and negative symptoms of schizophrenia: A pilot study. *Human Psychopharmacology*, 20, 139–144. https://doi.org/10.1002/hup.677

Oliveira, J. P., Zuardi, A. W., & Hallak, J. E. (2008). Role of nitric oxide in patients with schizophrenia: A systematic review of the literature. *Current Psychiatry Reviews*, 4, 219–227. https://doi.org/10.2174/157340008786847920

Overall, J. E., & Gorham, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports*, 10, 799–812. https://doi.org/10.2466/pr0.1962.10.3.799

World Health Organization.(2022). *Schizophrenia*.https://www.who.int/news-room/fact-sheets/detail/schizophrenia

Xiong, J. W., Wei, B., Li, Y. K., et al. (2018). Decreased plasma levels of gasotransmitter hydrogen sulfide in patients with schizophrenia: Correlation with psychopathology and cognition. *Psychopharmacology*, 235, 2267–2274. https://doi.org/10.1007/s00213-018-4930-3

Authors Bio:

My name is Atrim Das, and I am an MSc Molecular Neuroscience student at the University of Bristol, with a background in Biotechnology. My academic interests lie at the intersection of neuroscience, biotechnology, artificial intelligence, and translational medicine. Through my research experiences, I have explored areas ranging from psychiatric biomarkers and clinical biochemistry to gene therapy, molecular cloning, and AI-driven healthcare tools. I am particularly interested in how molecular and computational approaches can help us understand brain disorders and develop more accessible, precise, and patient-centred solutions. I hope to contribute to research that bridges laboratory science, clinical insight, and real-world impact.

Serum Hydrogen Sulphide as a Biochemical Correlate of Major Depressive Disorder: A Cross-Sectional Case-Control Study from a Tertiary Care Centre in Sub-Himalayan India

Authors: Atrim Das

ORCID: 0009-0000-6441-3769

Programme: Molecular Neuroscience MSc

Affiliation: SchoolofPhysiology,Pharmacology and Neuroscience

Email: qf25081@bristol.ac.uk

Abstract

Schizophrenia is associated with profound cognitive impairment, and disrupted neuroplasticity is increasingly implicated in its pathophysiology. Brain-derived neurotrophic factor (BDNF) is a key mediator of synaptic plasticity and is consistently reduced in schizophrenia. This prospective longitudinal study measured serum BDNF and neurocognitive performance in 40 first-episode, drug-naive patients at baseline (T0) and after eight weeks of atypical antipsychotic treatment (T8) at North Bengal Medical College and Hospital, India. At baseline, mean serum BDNF was markedly low (49.95 ± 19.92 pg/ml) and increased substantially to 211.03 ± 32.54 pg/ml at T8 ($p < 0.001$). Significant improvements were observed across all cognitive domains, including memory, executive function, processing speed, and verbal fluency. Baseline BDNF correlated strongly with cognitive performance across all measures. Post-treatment, these associations largely persisted for memory, processing speed, and verbal fluency, but the correlation between BDNF and the Mini-Mental State Examination was no longer significant, most likely reflecting range restriction once acute psychosis partially resolved. These findings suggest that serum BDNF may serve as a clinically accessible correlate of neuroplastic recovery in early schizophrenia, though controlled studies with healthy comparators and standardised treatment protocols are needed to establish causality.

Keywords schizophrenia, BDNF, neurocognition, atypical antipsychotics, neuroplasticity

Glossary of Abbreviations

BDNF: Brain-Derived Neurotrophic Factor
CNS: Central Nervous System
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
ELISA: Enzyme-Linked Immunosorbent Assay
FEDN: First-Episode Drug-Naive
LTP: Long-Term Potentiation
MMSE: Mini-Mental State Examination
NBMCH: North Bengal Medical College and Hospital
PANSS: Positive and Negative Syndrome Scale
PGIMS: PGI Memory Scale
SCWT: Stroop Colour Word Test
SD: Standard Deviation
TMT: Trail Making Test
TrkB: Tropomyosin Receptor Kinase B
VFT: Verbal Fluency Test

1. Introduction

Schizophrenia is a profoundly disabling psychiatric condition affecting approximately 24 million people worldwide, or roughly 1 in 300 individuals (World Health Organisation, 2022). While not the most prevalent illness, it ranks among the top 25 causes of global disability due to the substantial burden it places on patients, families, and healthcare systems. The disorder typically emerges before the age of 25 and follows a chronic, relapsing course defined by three main symptom categories: positive symptoms such as hallucinations, negative symptoms including emotional blunting, and cognitive deficits encompassing impaired memory and slowed processing speed (Bortolato et al., 2015). These cognitive impairments

cause the greatest functional damage over a lifetime. Antipsychotic medications often manage hallucinations and delusions effectively, yet cognitive difficulties typically persist across both acute and remission phases and remain the strongest predictor of long-term functional outcome (Green et al., 2004). The biological basis of this cognitive decline remains incompletely understood. Converging evidence implicates disrupted brain plasticity and altered neurodevelopment, leading many researchers to conceptualise schizophrenia as a neuroplasticity disorder (Favalli et al., 2012). Central to this framework is brain-derived neurotrophic factor (BDNF), the most abundant neurotrophin in the adult central nervous system. BDNF is concentrated in the hippocampus and prefrontal cortex, which are the primary regions subserving learning, memory, and executive function (Miranda et al., 2019). BDNF signals through its high-affinity receptor, Tropomyosin Receptor Kinase B (TrkB), to promote neuronal survival, dendritic arborisation, and synaptic strengthening required for long-term potentiation and memory consolidation (Klien et al., 1990; Lee and Silva, 2009). Because serum BDNF correlates with cerebrospinal fluid concentrations, a standard blood draw provides a practical proxy for central neurotrophin activity. Extensive research confirms that individuals with schizophrenia have lower serum BDNF than healthy controls. Green et al. (2011) showed that peripheral BDNF is reduced across both medicated and unmedicated schizophrenia samples, supporting BDNF as a robust illness-associated marker. However, the meta-analysis also highlighted substantial heterogeneity across studies, including differences in sample type, medication exposure, illness stage, and assay method, leaving the relationship between BDNF change and cognitive recovery unresolved. Atypical antipsychotics appear to partially restore BDNF levels, coinciding with symptomatic improvement (Zhang et al., 2018), though the precise relationship between circulating BDNF, antipsychotic treatment, and specific cognitive domains remains unclear. This ambiguity is particularly pronounced in first-episode, drug- naive (FEDN) populations. Studying individuals before any antipsychotic exposure removes the confounding effects of chronic illness and long-term medication, providing the clearest window into the relationship between BDNF and cognition (Man et al., 2018). Recent studies have begun to examine BDNF-cognition relationships in first-episode or drug-naive schizophrenia. Wu et al. (2020) reported altered BDNF-related neuroendocrine profiles in early schizophrenia, while Rahul

(2017) examined the effect of atypical antipsychotic treatment on serum BDNF and cognitive function in first-episode schizophrenia. However, evidence remains limited in Indian clinical populations, particularly for longitudinal studies combining serum BDNF with multiple domain-specific cognitive measures before and after antipsychotic exposure. The present study addresses this gap by tracking serum BDNF and comprehensive neurocognitive performance in FEDN patients before treatment and after eight weeks of atypical antipsychotic therapy. Three hypotheses were tested: (1) untreated patients would show significantly lower BDNF than at post-treatment follow-up; (2) cognitive performance would improve measurably after eight weeks; and (3) rising BDNF levels would correlate with better cognitive test scores. Conducted in North Bengal, this research provides regionally specific data to evaluate BDNF as a potential clinical correlate of cognitive recovery in early schizophrenia.

2. Methods

2.1 StudyDesignandSetting

A within-subjectslongitudinaldesignwasemployed,witheachparticipantservingas their own baselinecontrol.ThestudywasconductedattheDepartmentof Psychiatry, North Bengal Medical College and Hospital (NBMCH), Darjeeling, West Bengal, India.EthicalapprovalwasgrantedbytheInstitutionalEthicsCommitteeof NBMCH, andthestudywasregisteredwiththeCentralDrugsStandardControl Organisation.Written informed consentwasobtained fromall participantsortheir legal guardianspriortoenrolment.

2.2 Participants

Participantswererecruited fromthe hospital'spsychiatryoutpatientand inpatient departments.Inclusion criteria were agebetween 18and60years,first-time diagnosisofschizophreniaaccording toDSM-5 criteria,confirmed drug-naive status (no priorantipsychoticexposure),andatleastprimaryschooleducation.Exclusion criteria includedcomorbidpsychiatricconditions,significantmedicalillness,and substance use disorders.Atargetsample of53 participantswascalculated based on power analysisusingBDNFvariancedatafromPalominoetal.(2006),assuming 80% powerandα=0.05.Of47individualsinitiallyscreened,40completedboth assessmenttime-points,seven were losttofollow-up(fourwithdrew consent,three were unavailable at T8), and an additional shortage of ELISA kits precluded further recruitment.Thefinal cohortof40participantsfallsbelow theapriori target,which reduces statisticalpowerandshouldbeconsideredwheninterpretingeffectsizes and the generalisabilityoffindings.

2.3 ProcedureandNeurocognitiveAssessment

All participants were assessed at two time-pointsofT0(baseline,beforeany medication) and T8 (after eight weeks of atypicalantipsychotictreatment). Medication selection was made independently bythetreatingclinician,notthe

research team, and flexible dosing was used in accordance with standard clinical practice. Symptom severity was quantified using the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987), a 30-item instrument scored on a 7-point scale. Trained clinical psychologists administered a standardised battery of neuropsychological tests at both time-points of the Mini-Mental State Examination (MMSE; Folstein et al., 1975), the PGI Memory Scale (PGIMS; Pershad and Wig, 1976), the Stroop Colour Word Test (SCWT; Stroop, 1935), the Verbal Fluency Test (VFT; Benton and Hamsher, 1989), and the Trail Making Test (TMT; Reitan and Wolfson, 1993). The same psychologist administered both assessments to minimise inter-rater variability; however, practice effects on repeated cognitive testing cannot be excluded and represent a potential confound.

2.4 Serum BDNF Measurement

Fivemillilitresofvenousbloodwere collectedatT0andT8,processed immediately, and stored at −70°C until batchanalysis. Serum BDNF concentration was quantified using a Human BDNF ELISA Kit(detection range 3.125–2000 pg/ml). The laboratory analyst was blinded to participants' clinical diagnoses and treatment status to prevent measurement bias.

2.5 Statistical Analysis

Data were analysed using IBM SPSS version 22. Paired-samples t-tests compared T0 and T8 values for BDNF, PANSS scores, and all cognitive measures. Pearson correlation coefficients examined associations between serum BDNF and cognitive test scores at each time point. Chi-square tests assessed categorical associations between BDNF ranges and cognitive performance categories. Statistical significance was set at $p < 0.05$ (two-tailed). Multiple comparisons were not formally corrected, and the findings should therefore be interpreted with appropriate caution, particularly for the correlation analyses.

Measure	Domain assessed	Outcome interpretation
PANSS	Positive, negative, and general psychopathology symptoms	Higher score indicates greater symptom severity
MMSE	Global cognitive status, including orientation, attention, and language	Higher score indicates better global cognition
PGIMS TDS	Memory dysfunction across verbal and non-verbal tasks	Higher score indicates greater memory impairment
SCWT Word Card	Attention and processing speed	Longer completion time indicates poorer performance
SCWT Interference Card	Executive control, response inhibition, and cognitive flexibility	Longer completion time indicates poorer performance
VFT Semantic	Category-based verbal retrieval	Higher word count per minute indicates better semantic fluency
VFT Phonemic	Letter-based verbal retrieval	Higher word count per minute indicates better phonemic fluency
TMT-A	Psychomotor speed, visual scanning, and sequencing	Longer completion time indicates poorer performance
TMT-B	Set-shifting, divided attention, and executive function	Longer completion time indicates poorer performance

Table 1. Clinical and neurocognitive outcome measures used in the study

3. Results

3.1 Sample Characteristics

The final cohort comprised 40participants(21women,19men;meanage 35.78 years). The majority residedinrural areas(67.5%),spokeBengaliastheir primary language (72.5%), and wereclassifiedinlowersocioeconomicbrackets(70%) per the Modified B.G. Prasad Classification(2023).Mostparticipantsweremarried (80%) and had completed primary or secondaryschooling.Overhalf(52.5%) had experienced symptoms for one year orlessatenrolment.Themeanageof 35.78 years is higher than the typical onset agereportedintheliterature(<25years). This may reflect delayed help-seeking, diagnosticdelay,orstigma-relatedbarriers to care in this regional population, and may limitcross-studycomparisons.

3.2 Symptom Severity and BDNF Levels

Eight weeks of atypical antipsychotic treatmentwasassociatedwithasubstantial reduction in symptom severity. Mean PANSStotalscoredecreasedfrom113.78 (± 14.76 SD) at T0 to 54.93 (± 10.45 SD) atT8($p<0.001$).SerumBDNFincreased from 49.95 (± 19.92 SD) pg/ml at T0 to 211.03(±32.54SD)pg/mlatT8($p < 0.001$), representing an approximately 4.2-fold increase.Bothchangesareillustrated [Fig 1]. At baseline, participants with lower BDNFconcentrationstendedtohave higher PANSS scores ($\chi^2 = 20.53$, $p = 0.008$), consistentwithpriorreports(Singh et al., 2020).

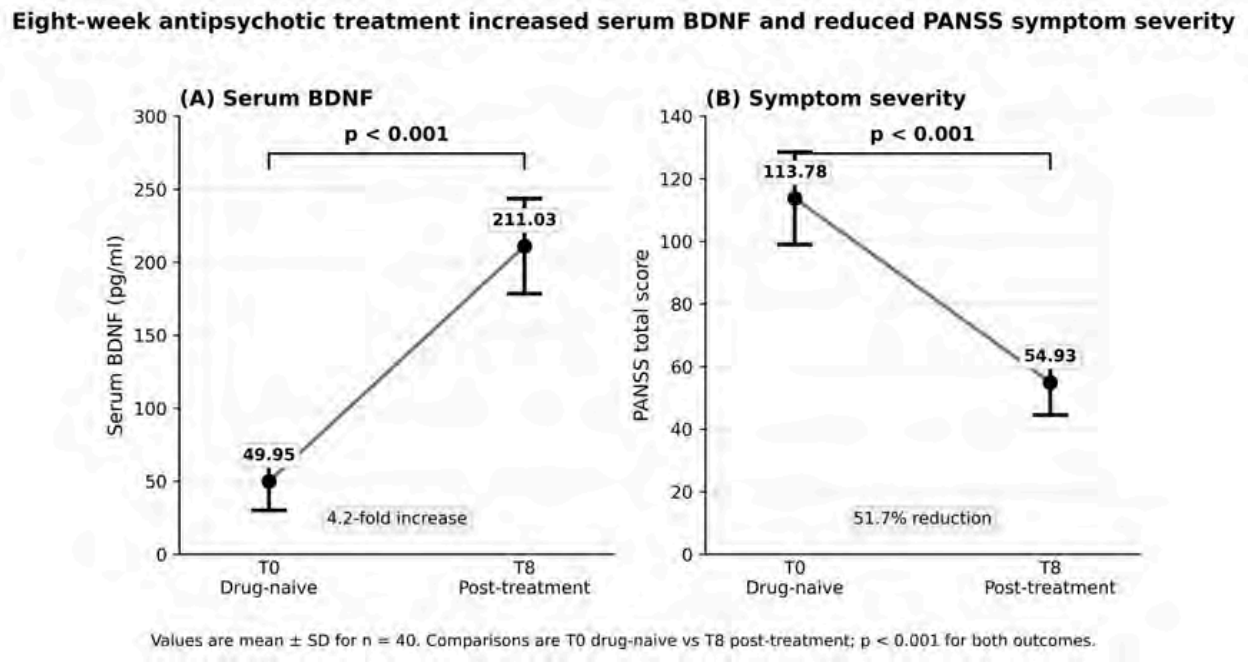


Figure 1. Paired plots of (A) serum BDNF (pg/ml) and (B) PANSS total score at T0 (drug-naive) and T8 (post-treatment) in 40 patients with first-episode schizophrenia. Group means are shown as filled circles with SD bars overlaid in black. Both changes were statistically significant (p < 0.001). PANSS = Positive and Negative Syndrome Scale.

3.3 Neurocognitive Outcomes

Significant improvements wereobserved across all cognitive domains between T0 and T8. Mean MMSE score increased from 14.40 (± 2.68 SD) to 17.95 (± 2.93 SD) (p < 0.001). PGI Memory Scale Total Dysfunction Score decreased from 15.83 (± 2.41 SD) to 8.48 (± 1.79 SD) (p < 0.001). SCWT Word Card completion time decreased from 105.33 (± 10.58 SD) to 74.35 (± 9.95 SD) seconds (p < 0.001), and SCWT Interference completion time decreased from 196.90 (± 18.70 SD) to 135.13 (± 14.71 SD) seconds (p < 0.001). Semantic VFT score increased from 11.55 (± 2.83 SD) to 15.43 (± 3.13 SD) words per minute (p < 0.001), and phonemic VFT score increased from 4.18 (± 1.29 SD) to 5.33 (± 1.47 SD) (p < 0.001). TMT-A completion time decreased from 106.9 to 76.4 seconds, and TMT-B from 217.3 to 146.8

seconds (both p < 0.001). The magnitude of improvement across all eight measures is summarised [Fig 2].

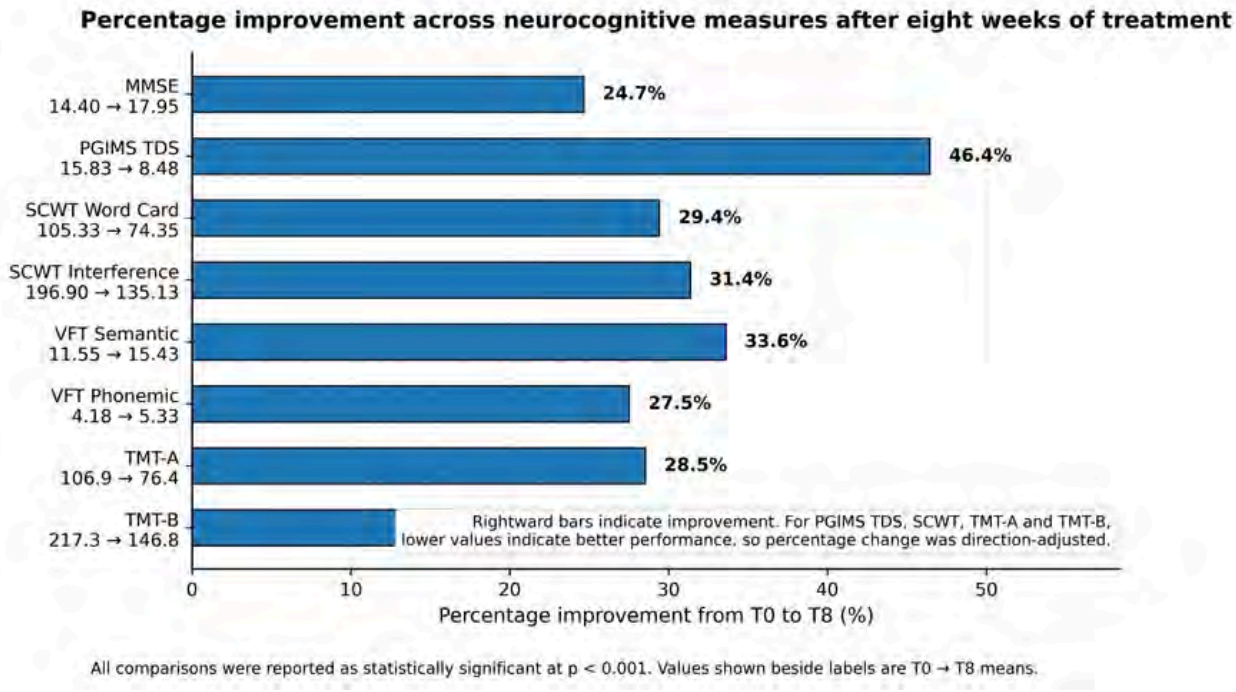


Figure 2. *Direction-adjusted percentage change is shown for eight cognitive outcomes after eight weeks of atypical antipsychotic treatment. For PGIMS TDS, SCWT Word Card, SCWT Interference, TMT-A and TMT-B, lower scores indicate better performance, so values were inverted to ensure all rightward bars represent improvement. All comparisons were significant at p < 0.001.*

3.4 Correlations Between BDNF and Neurocognition

At T0, serum BDNF correlated significantly with performance across all cognitive measures, as summarised [Fig 3]. Higher BDNF was associated with better general cognitive screening (MMSE: r = 0.81, p < 0.001), greater verbal fluency (VFT semantic: r = 0.79, p < 0.001; phonemic: r = 0.47, p = 0.002), lower memory dysfunction (PGIMS TDS: r = −0.66, p < 0.001), and faster processing speed (SCWT Word Card: r = −0.56, p < 0.001; Interference: r = −0.48, p = 0.002). At T8, significant correlations persisted for memory (PGIMS TDS: r = −0.53, p < 0.001), processing speed (SCWT Word Card: r = −0.49, p = 0.007; Interference: r = −0.47, p = 0.003), and verbal fluency (semantic: r = 0.47, p = 0.002; phonemic: r = 0.32, p = 0.046). The correlation between BDNF and MMSE was no longer statistically significant at T8 (r = 0.25, p = 0.115).

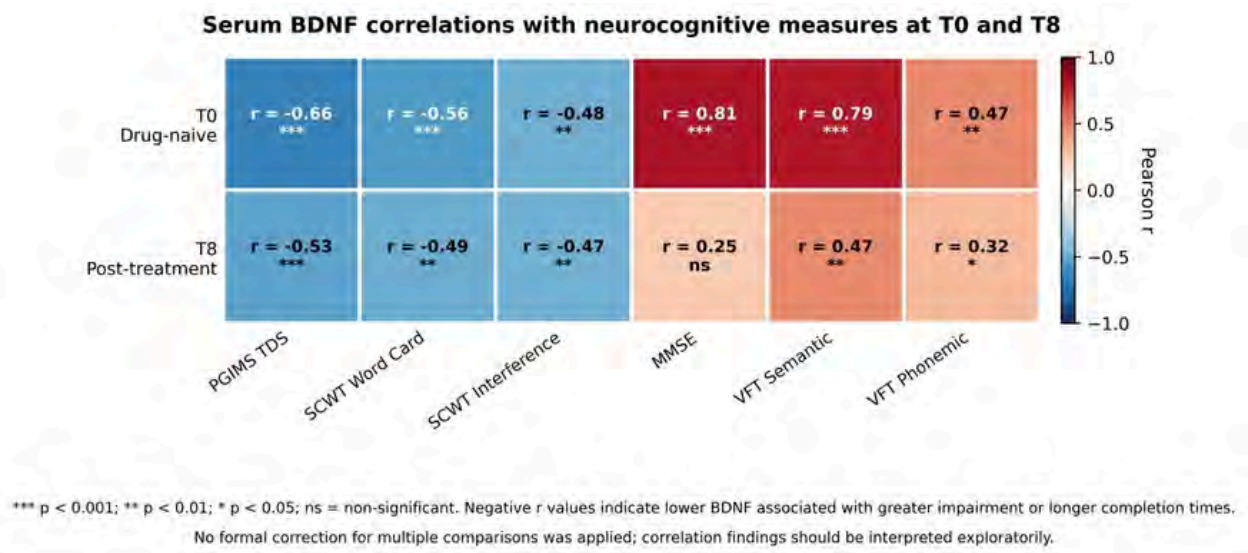


Figure 3. *Pearson correlation coefficients between serum BDNF and six neurocognitive measures at T0 (drug-naive) and T8 (post-treatment), displayed as a heatmap. Colour intensity reflects correlation magnitude (blue = negative; red = positive); cells marked 'ns' indicate p ≥ 0.05. The loss of significance for BDNF–MMSE at T8 (r = 0.25, ns) contrasts with the persistence of significant correlations for memory and executive function measures. No formal correction for multiple comparisons was applied, and correlation findings should be interpreted exploratorily. PGIMS TDS = PGI Memory Scale Total Dysfunction Score; SCWT = Stroop Colour Word Test; VFT = Verbal Fluency Test.*

4. Discussion

This studyprovides evidence that eight weeks of atypical antipsychotic treatment in first-episodeschizophrenia is associated with substantial increases in serum BDNF and concurrentimprovements across multiple cognitive domains. These findings are consistentwiththe hypothesis that antipsychotic-mediated upregulation of BDNF signallingmaycontribute to neuroplastic recovery. However, the within-subjects design andabsence of a healthy control group preclude causal inference. The markedlylowbaseline BDNF (49.95 ± 19.92 pg/ml) is consistent with prior reports in FEDN populations (Tanili and Faktor, 2018; Singh et al., 2020), and the post-treatmentincrease aligns with findings from Zhang et al. (2018) and Gonzalez-Pinto et al. (2010),who similarly observed BDNF rises following atypical antipsychotic initiation.Theapproximately 4.2-fold increase is notable. However, because serum BDNF valuesvary substantially across assay platforms and sample-processing protocols,theabsence of a matched healthy control group prevents firm conclusions about whetherpost-treatment levels approached normative ranges. Future studies incorporatingahealthy control group assessed with the same ELISA kit and processingprotocol would be necessary to address this question. The baseline

cognitive profile consisting of marked MMSE impairment (14.40/30), elevated PGIMS dysfunction score (15.83), prolonged SCWT Interference time (196.90 seconds), and reduced verbal fluency, maps onto the established neuropsychological signature of schizophrenia (Bilder et al., 2000; Zanelli et al., 2010) and replicates regional findings from Bhattacharyya et al. (2013) and Javed et al. (2018). The strong baseline correlation between BDNF and MMSE (r = 0.81) is among the highest reported in this context and is consistent with the known role of BDNF–TrkB signalling in hippocampal and prefrontal synaptic plasticity (Lewis et al., 2005; Miranda et al., 2019). The meta-analytic evidence linking reduced BDNF to neurocognitive deficits in schizophrenia (Ahmed et al., 2015) also supports the plausibility of this association, though the cross-sectional nature of the baseline correlation does not establish a causal direction. A clinically informative pattern emerged in the post-treatment correlation data. The significant BDNF–MMSE association observed at baseline was absent at T8 (r = 0.25, p = 0.115). Importantly, this should not be interpreted as a ceiling effect, since the mean post-treatment MMSE score of 17.95/30 remains well below the normal range (typically ≥ 24), indicating that substantial cognitive impairment persisted at T8. A more plausible explanation is range restriction. Once acute psychosis had partially resolved, MMSE scores clustered within a narrower band, reducing the variance available for correlation. This interpretation is consistent with the known insensitivity of the MMSE to subtle cognitive variation in partially recovered patients (Folstein et al., 1975). In contrast, correlations between BDNF and more demanding assessments like memory, processing speed, and verbal fluency persisted at T8, suggesting these domains continue to track ongoing neuroplastic processes beyond acute symptom resolution. This distinction has practical implications for biomarker research, proving BDNF may be most informative when paired with sensitive, domain-specific cognitive instruments rather than broad screening tools. Several methodological limitations warrant explicit acknowledgement. First, the absence of a healthy control group means that observed BDNF changes cannot be benchmarked against normative trajectories. Second, medication was not standardised. Flexible dosing and the use of multiple antipsychotic agents introduce pharmacological heterogeneity that may independently influence BDNF and cognition. Stratification by antipsychotic type and dose in future studies would help disentangle these effects. Third, the final sample of 40 participants falls below the a priori power target of 53, reducing the study's ability to detect smaller effects and increasing the risk of inflated effect size estimates. Fourth, repeated cognitive testing at T8 may have introduced practice effects, potentially inflating apparent cognitive gains independently of treatment. Fifth, treatment compliance was not formally monitored, and unmeasured non-compliance could have attenuated or obscured true treatment effects. Finally, the relatively older mean age of this cohort (35.78 years) compared with the typical onset age of schizophrenia (< 25 years) may reflect delayed help-seeking or diagnostic delay in this regional population, and may limit generalisability to younger FEDN samples. Despite these limitations, the findings contribute to a growing evidence base supporting serum BDNF as a clinically accessible correlate of cognitive recovery in

early schizophrenia. Because BDNF measurement requires only a standard blood draw, it is feasible in resource-limited settings. Future investigations should incorporate healthy control groups, standardised medication protocols with documented compliance, longer follow-up periods, and neuroimaging to map circulating BDNF to structural and functional brain changes.

5. Conclusion

This prospective longitudinal study demonstrates that serum BDNF is markedly reduced in drug-naive first-episode schizophrenia and increases substantially following eight weeks of atypical antipsychotic treatment, in parallel with significant improvements across memory, executive function, processing speed, and verbal fluency. The persistence of BDNF–cognition correlations for domain-specific measures after treatment, alongside the loss of the BDNF–MMSE association, most likely attributable to range restriction rather than a ceiling effect, highlights the importance of selecting sensitive cognitive instruments for biomarker research. These findings support the potential utility of serum BDNF as a clinically accessible correlate of neuroplastic recovery in early schizophrenia, while underscoring the need for controlled studies with standardised protocols to establish the direction and magnitude of this relationship.

References

Ahmed,A.O.,Mantini,A.M., Fridberg, D.J., et al. (2015) 'Brain-derived neurotrophic factor (BDNF)andneurocognitive deficits in people with schizophrenia: a meta-analysis',PsychiatryResearch, 226(1), pp. 1–13.

Benton,A.andHamsher,K. (1989) Multilingual Aphasia Examination. New York: Oxford UniversityPress.

Bhattacharyya,K.,Guha,P., Ghosal, M., et al. (2013) 'Association of thought disorderwithcognitivedysfunctions in schizophrenia: a cross-sectional study in a tertiarycentreinKolkata,India', German Journal of Psychiatry, 16(4), pp. 137–142.

Bilder,R.M.,Goldman,R.S., Robinson, D., et al. (2000) 'Neuropsychology of first-episodeschizophrenia:initial characterization and clinical correlates', American JournalofPsychiatry,157(4), pp. 549–559.

Bortolato,B.,Miskowiak,K.W., Kohler, C.A., et al. (2015) 'Cognitive dysfunction in bipolardisorderandschizophrenia: a systematic review of meta-analyses', NeuropsychiatricDiseaseand Treatment, 11, pp. 3111–3125.

Favalli,G.,Li,J.,Belmonte-De-Abreu, P., et al. (2012) 'The role of BDNF in the pathophysiologyandtreatment of schizophrenia', Journal of Psychiatric Research, 46(1),pp.1–11.

Folstein, M.F., Folstein, S.E. and McHugh, P.R. (1975) 'Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician', Journal of Psychiatric Research, 12(3), pp. 189–198.

Gonzalez-Pinto, A., Mosquera, F., Palomino, A., et al. (2010) 'Increase in brain-derived neurotrophic factor in first episode psychotic patients after treatment with atypical antipsychotics', International Clinical Psychopharmacology, 25(4), pp. 241–245.

Goren, J.L. (2016) 'Brain-derived neurotrophic factor and schizophrenia', Mental Health Clinician, 6(6), pp. 285–288.

Green, M.F., Kern, R.S. and Heaton, R.K. (2004) 'Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS', Schizophrenia Research, 72(1), pp. 41–51.

Green, M.J., Matheson, S.L., Shepherd, A., et al. (2011) 'Brain-derived neurotrophic factor levels in schizophrenia: a systematic review with meta-analysis', Molecular Psychiatry, 16(9), pp. 960–972.

Javed, S., Azmi, S.A. and Akhouri, D. (2018) 'Memory impairment in patients of schizophrenia', Age, 20(29), pp. 30–39.

Kay, S.R., Fiszbein, A. and Opler, L.A. (1987) 'The Positive and Negative Syndrome Scale (PANSS) for schizophrenia', Schizophrenia Bulletin, 13(2), pp. 261–276.

Klien, R., Conway, D., Parada, L.F., et al. (1990) 'The trkB tyrosine kinase gene codes for a second neurogenic receptor that lacks catalytic domain', Cell, 61(4), pp. 647–656.

Lee, Y.S. and Silva, A.J. (2009) 'The molecular and cellular biology of enhanced cognition', Nature Reviews Neuroscience, 10(2), pp. 126–140.

Lewis, D.A., Hashimoto, T. and Volk, D.W. (2005) 'Cortical inhibitory neurons and schizophrenia', Nature Reviews Neuroscience, 6(4), pp. 312–324.

Man, L., Lv, X., Du, X.D., et al. (2018) 'Cognitive impairments and low BDNF serum levels in first-episode drug-naive patients with schizophrenia', Psychiatry Research, 263, pp. 1–6.

Miranda, M., Morici, J.F., Zanoni, M.B., et al. (2019) 'Brain-derived neurotrophic factor: a key molecule for memory in the healthy and the pathological brain', Frontiers in Cellular Neuroscience, 13, p. 363.

Palomino, A., Vallejo-Ilarramendi, A., Gonzalez-Pinto, A., et al. (2006) 'Decreased levels of plasma BDNF in first-episode schizophrenia and bipolar disorder patients', Schizophrenia Research, 86(1–3), pp. 321–332.

Pershad, D. and Wig, N.N. (1976) 'A battery of simple tests of memory for use in India', Neurology India, 24(2), pp. 86–93.

Reitan, R.M. and Wolfson, D. (1993) The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation. 2nd edn. Tucson: Neuropsychology Press.

Singh, J., Verma, R., Raghav, R., et al. (2020) 'Brain-derived neurotrophic factor (BDNF) levels in first-episode schizophrenia and healthy controls: a comparative study', Asian Journal of Psychiatry, 54, p. 102370.

Stroop, J.R. (1935) 'Studies of interference in serial verbal reactions', Journal of Experimental Psychology, 18(6), pp. 643–662.

Tanili, A.T. and Faktor, H.S. (2018) 'Investigation of serum brain-derived neurotrophic factor (BDNF) levels in patients diagnosed with schizophrenia without antipsychotictreatmenthistory',Journal ofNeurobehavioral Sciences,5(3),pp.165–

171.

Vinogradov, S., Fisher, M., Holland, C., et al. (2009) 'Is serum brain-derived neurotrophic factor a biomarker for cognitive enhancement in schizophrenia?', Biological Psychiatry, 66(6), pp. 549–553.

World Health Organization (2022) Schizophrenia [Online]. Available at: https://www.who.int/news-room/fact-sheets/detail/schizophrenia (Accessed: June 2024).

Zanelli, J., Reichenberg, A., Morgan, K., et al. (2010) 'Specific and generalized neuropsychological deficits: a comparison of patients with various first-episode psychosis presentations', American Journal of Psychiatry, 167(1), pp. 78–85. Zhang, Y., Fang, X., Fan, W., et al. (2018) 'Brain-derived neurotrophic factor as a biomarker for cognitive recovery in acute schizophrenia: 12-week results from a prospective longitudinal study', Psychopharmacology, 235(4), pp. 1191–1198.

Authors Bio:

My name is Atrim Das, and I am an MSc Molecular Neuroscience student at the University of Bristol, with a background in Biotechnology. My academic interests lie at the intersection of neuroscience, biotechnology, artificial intelligence, and translational medicine. Through my research experiences, I have explored areas ranging from psychiatric biomarkers and clinical biochemistry to gene therapy, molecular cloning, and AI-driven healthcare tools. I am particularly interested in how molecular and computational approaches can help us understand brain disorders and develop more accessible, precise, and patient-centred solutions. I hope to contribute to research that bridges laboratory science, clinical insight, and real-world impact.

Editor Details

Co-Editor-in-Chief
- Emma Segal
- *iu22655@alumni.bristol.ac.uk*

Co-Editor-in-Chief
- Megan Stannett
- *in22570@alumni.bristol.ac.uk*

Deputy-Editor-in-Chief
Arts, Law and Social Sciences
- Danlu Chen
- *danlu.chen@bristol.ac.uk*

Deputy-Editor-in-Chief
Health and Life Sciences
- William Wheatley
- *will.wheatley@bristol.ac.uk*

Deputy-Editor-in-Chief
Science and Engineering
- Sanjana Sundararaman
- *Lr25497@bristol.ac.uk*

Copy Editor
- Ozioma Onyeama
- *pj23401@bristol.ac.uk*

Subject Editors
- Adrian-Mihai Costea
- *adrian.costea@bristol.ac.uk*

- Akanksha Pai
- *dk22299@bristol.ac.uk*

- Alexandra Norris
- *alex.norris.2023@bristol.ac.uk*

- Allegra Gervasi
- *kj25701@bristol.ac.uk*

- Anna Shahbazyan
- *anna.shahbazyan.2025@bristol.ac.uk*

- Emily Birt
- *emily.birt@bristol.ac.uk*

- Hou Hin Ip (Izaac Ip)
- *le23036@bristol.ac.uk*

- Heston Brocklehurst
- *kv22866@bristol.ac.uk*

- Hussain Almaainah
- *zu23636@bristol.ac.uk*

- Senali Dananja Galappaththi
- *ll25775@bristol.ac.uk*

- Shakirah Rashid
- *pz23066@bristol.ac.uk*

- Solomon Nderitu
- *mc25672@bristol.ac.uk*

Peer Reviewer Details

Arts, Law and Social Sciences

- Alannah Mylechreest
- *gq22090@bristol.ac.uk*

- Alexander Matthews
- *wx25775@bristol.ac.uk*

- Ava Kate Brierley
- *fk23233@bristol.ac.uk*

- Emily MacLean
- *bs24881@bristol.ac.uk*

- Grace Smith
- *ek24431@bristol.ac.uk*

- Isla Mercer
- *nh23518@bristol.ac.uk*

- Jed Dixon
- *ky24535@bristol.ac.uk*

- Kaixuan Chen
- *in23517@bristol.ac.uk*

- Laura Agard
- *laura.agard@bristol.ac.uk*

- Leander Bartsch
- *iu25335@bristol.ac.uk*

- Megan Gear
- *hu24737@bristol.ac.uk*

- Monica Cheng
- *monica.cheng@bristol.ac.uk*

- Sarah Morcom
- *bc22311@bristol.ac.uk*

- Sagrika Mehta
- *al25422@bristol.ac.uk*

Health and Life Sciences

- Anna Holliday
- *en23518@bristol.ac.uk*

- Jayden Wordley
- *jayden.wordley@bristol.ac.uk*

- Keisha Tannesha Santoso
- *jx22984@bristol.ac.uk*

- Manar Amasha
- *manar.amasha@bristol.ac.uk*

- Millie Chilcott
- *hs22262@bristol.ac.uk*

- Quincy Mulraine
- *bb22328@bristol.ac.uk*

- Samantha Siomko
- *jf23062@bristol.ac.uk*

- Tamie Karunaratne
- *Dj23802@bristol.ac.uk*

Science and Engineering

- Adrian Gallo
- *bd25228@bristol.ac.uk*

- Celine Hwang
- *gv24688@bristol.ac.uk*

- Lorenzo Bravo
- *jj22921@bristol.ac.uk*

- Saranya Srinivasan
- *bu25681@bristol.ac.uk*

- Timothy Holland
- *tim.holland@bristol.ac.uk*