South African Journal of Science

volume 119
number 3/4

New perspectives from South Africa on the study of faces

First fossil scarab beetle from Africa

The Global Syndemic – undernutrition, obesity and climate change collide

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 119
Number 3/4
March/April 2023

South African Journal of Science

eISSN: 1996-7489

Leader

Book Reviews

Invited Commentaries

Commentary

Published by
the Academy of Science of South Africa (www.assaf.org.za) with financial assistance from the Department of Science & Innovation.

Design and layout
Elzahn Swarts
E: swarts.elzahn@gmail.com

Correspondence and enquiries
sajs@assaf.org.za

Submissions
Submissions should be made at www.sajs.co.za

Review Article

Research Articles

Face Science

Guest Leader

Research Articles

Cover caption
What's in a face? Our faces are central to our identity and to how we perceive our environment, communicate, and express ourselves. A special section on Face Science in this issue presents cutting-edge research and new perspectives on the study of faces within South Africa.

The year that was 2022: Looking back and looking ahead

Last year in 2022 we introduced the 'year that was' editorial in which we share our reflections on the previous year and on how we did in reaching our goals for the year. As an open-access multidisciplinary journal, those goals already encompass a wide range: publish original research that is relevant to Africa, multidisciplinary, and suitable for non-specialist readers; promote the visibility of published research; and encourage academic debate and discussions. But as a journal that will be 120 years old next year, we have continually added to that list to adapt to changing times. Diversity, equity, inclusion and accessibility are a global focus in scholarly publishing today, and it is important that we examine our journal's accessibility beyond subscriptions and article-processing charges (as a diamond open-access journal we cannot do better) to make our content more accessible for visually impaired readers and readers for whom English is not their first language, and to make publishing more accessible for authors who do not have the training or the confidence to write their research and submit it, thereby improving inclusion, equity and diversity in our context. We touch on the activities we undertook in 2022 to aim to achieve these goals. Some of the reflections will be familiar to our readers; in order to reflect fully on the year, we revisit some themes we have discussed before. We acknowledge that our footprint is small, our stride is short and there is a long road yet to travel.

Publishing peer-reviewed research relevant to Africa

We have a very high desk rejection rate (79% in 2022), with many of our rejections being on the basis of submissions being out of the scope of the journal. We are a journal from and about Africa, and we are a multidisciplinary journal, so even excellent submissions that are of interest only to a very narrow subgroup of readers will not be considered for our journal and will be better placed elsewhere. Because of the criterion of African relevance, almost half (48%) of all desk rejections are submissions from authors outside of Africa, with about 23% and 29% being from South Africa and the rest of Africa, respectively.

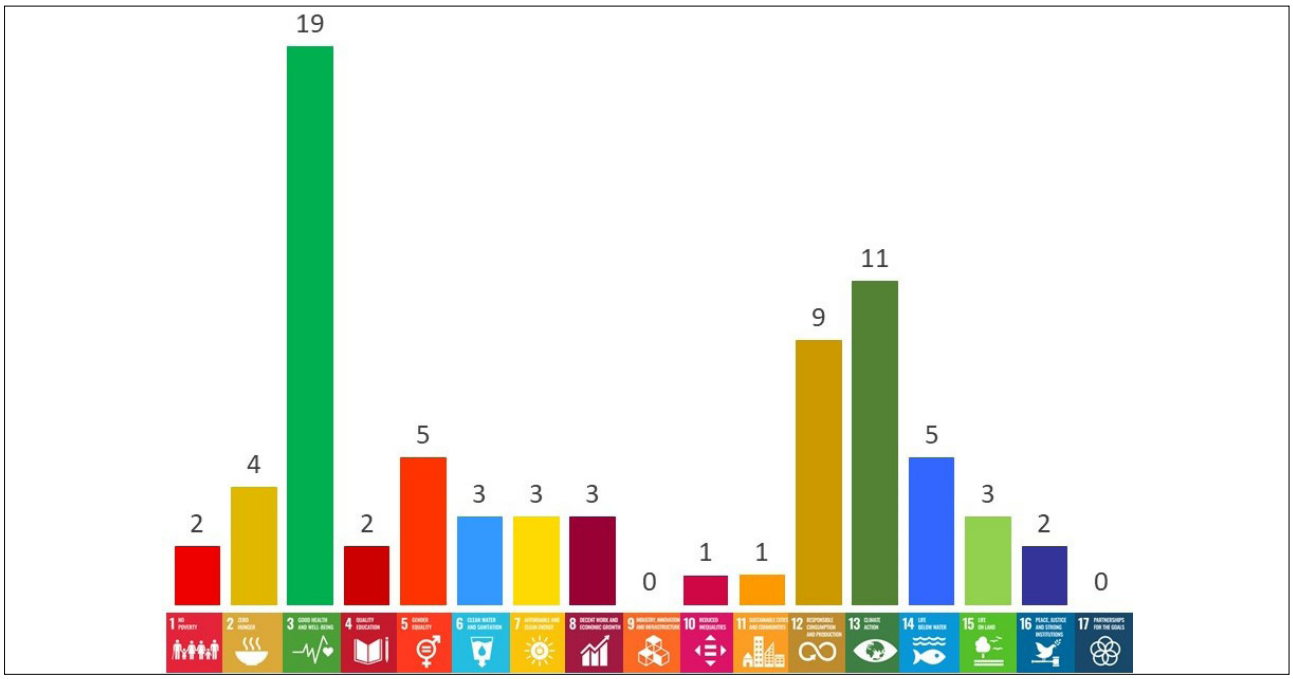
A total of 490 original research and review articles were submitted to the journal in 2022 – 35% of these submissions were from South Africa, and 26% were from elsewhere on the continent (a small increase on last year's 24%). We published 82 peer-reviewed articles across the eight issues published in 2022. About 73% of the published authors were from South Africa, with 15% from the rest of Africa – an increase on last year's 10%. Last year, reflecting on 2021 submissions, we indicated that we would like greater participation from African countries outside of South Africa. The small increases in the number of submissions and publications from authors from African countries outside of South Africa are encouraging, but there is scope for further improvement.

Our average turnaround time from submission to final decision (which excludes revision rounds) was 142 days in 2022. The struggle to find reviewers continued in 2022, with an average number of reviewers approached per submission in 2022 of seven – compared with an average of six invitations per submission in 2021. In total, 840 reviewers were approached in 2022, of whom 243 completed a peer review and 12 completed two or more peer reviews. In 2022, the average time to accept a review invitation was 14 days – equivalent to the time taken to complete a review after acceptance. One of our biggest challenges in improving turnaround times is the high number of reviewers invited who do not respond, either to accept or to decline a review invitation. We are again grateful to every reviewer who completed a review, as well as to those who, if unable to review, responded and recommended others. Without our peer reviewers and associate editors, we would not have peer-reviewed content to publish, and we are appreciative of their time and contributions (a list of those who reviewed for us in 2022 can be found here).

Publishing multidisciplinary research

The articles published in 2022 fall within 22 research categories, as classified by Dimensions, and contributed to 15 of the 17 Sustainable Development Goals (Figure 1).

The eight issues published in 2022 – two more than our usual six issues per year – included three special issues. The first, entitled 'How do you do social distancing in a shack: COVID-19 in the South African context', demonstrated our commitment as a journal to bringing together multidisciplinary perspectives to address complex issues. Social and health scientists, epidemiologists and ethicists, amongst



Source: Digital Science Dimensions

Figure 1: Number of articles published in 2022 that contribute to each of the 17 Sustainable Development Goals.

others, all made contributions with very tight deadlines. We are very grateful to our guest editorial team, Jonathan Jansen and Shabir Madhi, for demonstrating what results can accrue when editors from very different fields work together. This underlying theme of interdisciplinarity was also demonstrated in the second special issue, entitled 'Waste as a resource: South African perspectives on circularity'. This issue was done in partnership with the Community of Practice (CoP) 'Waste to Value: Transitioning South Africa towards a Waste-to-Resource Circular Economy', and it was a privilege to work with the guest editorial team from a range of disciplines on a topic of central importance to our country and our planet – another topic which can best be addressed only by multidisciplinarity and communities of practice such as theirs.

The third special issue of the year was something of a departure for our journal. This issue emanated from the contributions of the HSRC Radical Reason 'Conversations with Global Thinkers' and we were exceptionally fortunate to work with Rachel Adams and Crain Soudien on this issue, which broke boundaries for us, not only in terms of methods and approaches used from social science, philosophical and decolonial perspectives, but also in terms of the structure of contributions, some of which were in the form of discussions and debates. We are committed as a journal to providing a platform for a range of voices across many disciplines because it is precisely at points of unfamiliarity and contestation that new ideas and approaches to solutions for complex problems, including questions of climate change and environmental degradation, commonly begin to grow. It is the nature of academia that there will be strong differences of opinion on a range of issues, and we welcome and support open academic debate in our journal.

Making published research visible

We have a modest, but growing, social media presence, which is focused on promoting published content and increasing its visibility and reach across a wider readership. We encourage readers to follow us on Twitter, Facebook and LinkedIn and to engage with us through these platforms.

We continued to issue media releases to local mainstream media ahead of the publication of each issue. Media reports on articles published in the journal can be viewed on our website here. In 2022, there were 125 media mentions of published articles, with a global online reach of 6.5 million.

Enhancing accessibility and inclusivity

As a journal with an African focus, we have a vested interest in developing and supporting scientific writing in Africa. In 2022 we again provided online workshops on scientific writing and peer reviewing, and we have an ongoing monthly online forum to support new scholars in writing and peer review. These activities are free and open to all, with participation from across the globe.

This said, we are well aware of the limitations of what we do and of our reach. We regularly receive manuscripts where we gain the impression that otherwise competent authors have not, through their training and research experience, developed the level of methodological and writing skills which would be optimal for scientific publication in a global knowledge economy. Our small contribution of workshop and other support is not enough to address this issue, and we aim to have more engagements with diverse stakeholders on this issue. As always, if there are readers of the journal who wish to engage with us, please do contact us. It is a responsibility of the entire research community in South Africa and on our continent to nurture growth and excellence, and we all need to work together.

In this regard, we have received thus far only positive feedback on our Inclusive Language Policy promulgated in 2022. We developed this policy with an eye on accessibility and increasing participation in the journal and to promote inclusion through writing, and we welcome feedback, both positive and critical.

Looking ahead

As a journal in a changing context, we have to grow and to change. There are challenges and opportunities. We believe it is too early, for example, to assess the likely impact of ChatGPT and similar technologies on our work. While we explore the possibility of publishing peer review reports in the future, we remain cognisant that double-anonymous peer review is important in mitigating potential bias and enabling inclusivity. For example, a recent article[1] in *Nature Ecology & Evolution* provides evidence for review outcome gaps based on author demographics. An issue raised in this article is that of the demographics of peer reviewers, something that is not always thought about.

Our journal will be 120 years old in 2024, and we are planning to use this milestone to reflect on the past and think about the future. We have so much that is excellent on which we are able to build, and we are also aware that, as the world changes in all sorts of ways, we have a responsibility to change where appropriate. Diversity, access, rigour, and debate are all key to good science and we are dependent on our readers and contributors to keep challenging us and helping us to improve. Mistakes and missteps are part of this process; the important thing for us is to be a contributing part of a community which focuses on learning and knowledge for the good of society.

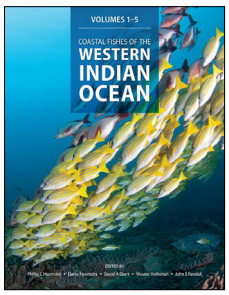
Reference

1. Smith OM, Davis KL, Pizza RB, Waterman R, Dobson KC, Foster B, et al. Peer review perpetuates barriers for historically excluded groups. Nat Ecol Evol. 2023. https://doi.org/10.1038/s41559-023-01999-w

HOW TO CITE:
Swartz L, Fick L, Grobler N, Wagener H. The year that was 2022: Looking back and looking ahead. S Afr J Sci. 2023;119(3/4), Art. #15803. https://doi.org/10.17159/sajs.2023/15803

Who's who of fishes of the Western Indian Ocean

BOOK TITLE:
Coastal fishes of the Western Indian Ocean



EDITORS:
Phillip C. Heemstra, Elaine Heemstra, David A. Ebert, Wouter Holleman, John E. Randall

ISBN:
9781998950409 (hardcover set, limited print run, 2698 pp)
9781990951282 (PDF volume 1, 643 pp)
9781990951299 (PDF volume 2, 620 pp)
9781990951305 (PDF volume 3, 473 pp)
9781990951312 (PDF volume 4, 471 pp)
9781990951329 (PDF volume 5, 491 pp)

PUBLISHER:
South African Institute for Aquatic Biodiversity, Makhanda, South Africa

PUBLISHED:
2022

REVIEWER:
Sean T. Fennessy[1]

AFFILIATION:
[1]Oceanographic Research Institute, Durban, South Africa

EMAIL:
seanf@ori.org.za

HOW TO CITE:
Fennessy ST. Who's who of fishes of the Western Indian Ocean. S Afr J Sci. 2023;119(3/4), Art. #15676. https://doi.org/10.17159/sajs.2023/15676

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 March 2023

The highly anticipated *Coastal Fishes of the Western Indian Ocean* was finally published towards the end of 2022. Anyone working in the fields of marine fish identification, fisheries and fish biology, in the Western Indian Ocean (WIO) region and beyond, can put away their well-thumbed copies of *Smith's Sea Fishes*, and reach for their laptops (or phones) instead because *Coastal Fishes of the Western Indian Ocean* is digital – and freely available (https://www.saiab.ac.za/coastal-fishes-of-the-western-indian-ocean.htm). A limited number of printed copies of the 5-volume box sets will be made available depending on demand; I reviewed the digital version.

I had always assumed that JLB Smith's *The Sea Fishes of Southern Africa* (Central News Agency, 1949) was the first consolidated effort at an African east coast fish guide, but, according to the *Origins* chapter of the WIO book, that honour goes to *Fishes of Zanzibar* (Van Voorst, 1866). Smith did, though, conceive of a WIO fishes book in the early 1950s, shortly after the publication of his pioneering 1949 *Sea Fishes*. Building on this, Margaret Smith and Phil Heemstra of the (then) JLB Smith Institute of Ichthyology in Grahamstown produced the mammoth *Smiths' Sea Fishes* (Macmillan South Africa), subsequent to the 1986 edition of which informal discussions were held at the institute which raised the daunting feasibility of producing a WIO fishes guide. This momentum was fortunately sustained by the institute's directors, with manuscripts incrementally being sent to Heemstra, and by 2014 most had been submitted. The bulk of the book consists of taxonomic identification keys from around 100 contributors, all experts on their respective taxa. In recognition, the first volume includes their biographical sketches, which is a nice touch – several of these luminaries have regrettably passed away.

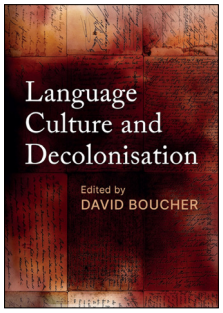
Their contributions had to be dealt with by the editorial team, all research associates of the publisher, the South African Institute of Aquatic Biodiversity (SAIAB) – a national facility of the National Research Foundation. This was a colossal task, the extent of which was not always appreciated by unsympathetic and impatient users-to be, who kept demanding "When will the WIO book be ready??". I count myself among their ranks, though I had the benefit of receiving some pre-publication keys when I was grappling with tricky specimens.

The book is a guide identifying the 3200+ species of coastal fishes known from the WIO, which has 15–20% of the world's known fishes. "Coastal" means those that occur in water less than ~200 m in depth – so those that inhabit inshore waters (including estuaries, beaches and rocky shores), extending onto the continental shelf and the upper part of the continental slope. Also included are some oceanic and deep-water species which move into shallower waters at times. The WIO is considered as the east and southern coasts of Africa (to Cape Point), Madagascar and the tropical islands (including Comoros, Seychelles, Maldives, Lakshadweep, Chagos), across to the southern tip of India and to as far south as St Paul Island (40° S).

Following the 'Origins' chapter, broad context and background to the WIO fishes is provided in subsequent chapters in Volume 1; these chapters are mines of information on diverse pertinent topics by several experts, including oceanography of the WIO, evolution of fishes, fish biology and genetics, origins and geology of WIO reefs, WIO fisheries, and a fascinating account of the early ichthyologists who first started collecting fishes in the region. A list of acronyms and a massive bibliography feature here too. Also here are instructions on how to use the book – right from the basics in case you do not know which family your fish specimen belongs to (if you do know, there are alphabetical indices of all family names (scientific and common) with their corresponding volume number to take you to the correct volume). If your specimen is cartilaginous (hagfishes, sharks, rays, chimaeras), the key to Families of Chondrichthyans takes up the remainder of Volume 1. If it is a bony fish, the key to their Orders is in Volume 2, commencing with the more primitive orders (eels, herrings, flyingfishes, seahorses, scorpionfishes, etc). Keys to the 'typical' fishes (Order Perciformes) are in Volumes 3 to 5, culminating in the more advanced families such as flatfishes, triggerfishes, filefishes, pufferfishes, etc. Preceding the systematic accounts, there are highly readable descriptions of the origins, evolution, diversity and anatomy of the cartilaginous and bony fishes. There are photos and/or illustrations for virtually every species in the book; there are 129 colour plates, and, with an estimated average of 10 species per plate, these alone produce well over 1000 colour images of species – let alone the innumerable black-and-white illustrations supplementing the systematic accounts. Some of these have been reproduced from earlier identification guides, but many are *de novo.* There is an overall glossary and numerous other subsidiary glossaries, which I found useful when interpreting terminology in the keys. Continuing the tradition of *Smith's Sea Fishes*, relevant literature is provided for each species in the book, and, for most, additional wider information on life history, interest to fisheries, etc. is also provided.

Since the book was launched in September 2022, I have had occasion to use several of the keys in the book, with success; the sheer size of the work means I cannot claim that there are no taxonomic errors, and trust the editorial and proofreading teams. There are, and will always be, specimens whose identity cannot be resolved in taxonomic works such as these. Resolution of, and ongoing disagreements between, taxonomists, and continual discoveries of new species all mean that the identification of species can be a moving target. But, being electronic and online, the keys can be updated frequently, and planning is already underway to expedite this. Some readers will regret that there is not a one-volume compilation to carry around with them, but the scope and scale of the fishes of the WIO rendered this impossible; besides, you can have it on your smartphone. A colleague of mine has even concatenated the volumes into one PDF file to make it more easily searchable. This epic publication sustains the enduring legacy of SAIAB and JLB in Makhanda – the editors and contributors can be very proud. And no, this review was not done by ChatGPT.

BOOK TITLE:
Language, culture and decolonisation



EDITOR:
David Boucher

ISBN:
9780796926128 (paperback, 384 pp)

PUBLISHER:
HSRC Press, Cape Town,
South Africa; ZAR327

PUBLISHED:
2022

REVIEWER:
Kathy Luckett[1,2]

AFFILIATIONS:
[1]Centre for Higher Education Research, Teaching and Learning, Rhodes University, Makhanda, South Africa
[2]Centre for Higher Education Development, University of Cape Town, Cape Town, South Africa

EMAIL:
kathy.luckett@uct.ac.za

HOW TO CITE:
Luckett K. Reflections from South Africa on *Language, Culture and Decolonisation*. S Afr J Sci. 2023;119(3/4), Art. #15640. https://doi.org/10.17159/sajs.2023/15640

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 March 2023

Reflections from South Africa on *Language, Culture and Decolonisation*

The chapters in this book originated at a series of conferences hosted by the Johannesburg Institute for Advanced Study, Department of Politics and International Relations, University of Johannesburg. As the title suggests, the focus of the book is on relations between language, culture and decolonisation, in particular on strategies for countering the injustices, erasures, and manipulations of the languages and cultures of the colonised. Given the book's origin, South Africa and South African authors feature prominently, with the majority of chapters focusing on the role of language in relation to the decolonising project in South Africa and/or Africa and the African diaspora.

In the Introduction, the editor sets out the salience of a book that investigates language oppression for taking forward the contemporary project of intellectual decolonisation in a context where political decolonisation has failed to realise justice for colonised peoples. Boucher claims that the book 'rise(s) to the challenge of exploring strategies for taking possession of language(s), and breaking the silence that...the seeing power of race...has imposed on the invisible and inaudible' (p. 9). He offers three (overlapping and not necessarily exclusive) types of colony to explain why intellectuals in the colonies have historically theorised and practised different forms of resistance and associated approaches to language and culture. The three categories are white settler colonies, colonies administered by indigenous elites, and colonies where the 'natives' were considered so primitive that they first had to be 'civilised' in preparation for self-rule. In a context where we are currently inundated with generalised, abstract calls to action for decolonisation, Boucher's recourse to historical context is welcome: "We must not, however, fall into the trap of believing that a 'general theory' of coloniality and decoloniality will explain each instantiation of the relation between colonizer and colonized. The lived experience of each liberation movement, and the centrality of language to its character and identity, is always contingent and circumstantial" (p. 9). Thus the book aims to introduce the reader to the complexity of the language question and its constitutive histories in the post/neo-colonies where people move in and out of colonial languages, indigenous languages and creoles, for different purposes and with varying degrees of proficiency. One of the key issues discussed in the book is the debate around the role of indigenous languages and creoles in relation to the formation of national consciousness and reclamation of African sovereignty.

In Chapter 1 (p. 19–34), Boucher usefully outlines three approaches taken by colonised intellectual and political leaders in response to language oppression:

1. The appropriation of the colonial language for revolutionary purposes – a pragmatic approach often found in large multi-lingual countries such as India (and by the ANC in South Africa), or in diasporic contexts, such as Negritude's subversion of French. This approach tends to view language as a universal human capacity used to express a given reality.
2. The assertion and revival of the indigenous languages – as constitutive of colonised identities, dignity and self-consciousness and necessary for authentic cultural expression (the imposition of colonial languages is seen as culturally and psychically destructive). This approach is based on linguistic relativism (articulated in the Sapir-Whorf hypothesis) where languages are understood to have different (even incommensurable) conceptual and sematic structures.
3. The positive (re)clamation of pidgins and creoles as new languages and expressions of hybrid cultures infused with new identities, values and creativity – usually in contexts with histories of slave-slave-settler interaction and transculturation. This approach is based on a view of language as fluid, dynamic and relational socio-cultural practice as opposed to innate capacity or discrete cultural system.

I use this framework to reflect on two of the chapters that highlight the contradictions and challenges of the language question as the authors grapple with taking up approach 2 and 3, respectively. Chapter 3 (p. 47–64) by South African philosopher John Lamola is a good example of a critical discussion around the second approach to the language question in postcolonial contexts – an approach that adheres to a strong theory of decolonisation.[1] Lamola explores the claim that authentic African philosophy can only be articulated by African people in their indigenous languages and that the idea of African philosophy articulated in a colonial language is not only a philosophical absurdity, but also has 'grave psycho-social ramifications' – for example, double consciousness or self-division (Du Bois, Biko). On the other hand, Lamola also wishes to avoid trapping African philosophy in a myriad of 'tribalised social ontologies' that Hountondji (1983) has long since rejected as 'ethnophilosophies'. Lamola discusses the works of established African philosophers such as Senghor, Mbiti, Wiredu, Eze, Oruka and Tangwa to argue that by aligning themselves with positions close to language approach 1, they represent the African elite who advocate Western modernisation for Africa, thus reproducing neo-colonialism. Lamola claims that 'as a philosophy emanating from the historical milieu of a traumatised people, (African) philosophy still has to make peace with the searing challenge of the psycho-political effects of language in all its varied implications and complexities as raised by Fanon and Ngugi'; concluding that 'professional African philosophy remains an intractable defective enterprise'.

By contrast, Chapter 4 by Brian Sibanda contests Lamola's position from a reading of decolonialisation that best aligns with language approach 3. Sibanda is critical of Afro-radicals who seek to abolish colonial languages in order to build new nativist hierarchies and 'return to some impossible pristine place' (p. 68). Instead he argues that the decolonial project is not a philosophy of revenge but one that seeks to liberate all people, both coloniser and colonised, by disestablishing hierarchies, decentring colonial languages and promoting linguistic multiplicity

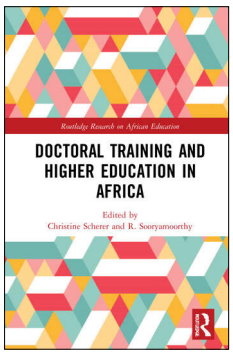
within a new inclusive pluriversality in which indigenous languages are accorded equal status – thus making transcultural conversations possible. Using the seemingly contradictory examples of Achebe (approach 1) and Ngugi (approach 2), Sibanda argues that both colonial and indigenous languages have a role in a collective decolonial project that is built on recognising difference and multiple local particularities in the quest to rehumanise the world.

My own brief reflections on these two chapters are firstly that Lamola (writing in academic philosophical discourse in English) falls into the trap of homogenising Africa and Africans and so fails to adequately address the problems posed by the multilingual nature of Africa as well as the desire of many Africans to appropriate modernity in their own styles and languages. On the other hand, while Sibanda's approach to the language question seems more realistic, he fails to adequately take unequal power relations into account, especially the global hegemony of the English language and thus the need for political power and will to challenge its dominance and protect and nurture indigenous languages.

Reference

1. Jansen JD, Walters CA. The decolonization of knowledge: Radical ideas and the shaping of institutions in South Africa and beyond. Cambridge: Cambridge University Press; 2022. https://doi.org/10.1017/9781009082723

BOOK TITLE:
Doctoral training and higher education in Africa



EDITORS:
Christine Scherer and Radhamany Sooryamoorthy

ISBN:
9781032025667
(hardback, GBP130.00, 224 pp)
9781003183952
(ebook, GBP29.24, 224 pp)

PUBLISHER:
Routledge, Oxfordshire, UK

PUBLISHED:
2022

REVIEWER:
Emmanuel M. Mgqwashu[1]

AFFILIATION:
[1]Director: Centre for Higher Education Professional Development (CHEPD), Faculty of Education, North-West University, Potchefstroom, South Africa

EMAIL:
Emmanuel.Mgqwashu@nwu.ac.za

HOW TO CITE:
Mgqwashu E. To 'train' or to 'educate' for doctoral work? – that is the question: A review of *Doctoral Training and Higher Education in Africa*. S Afr J Sci. 2023;119(3/4), Art. #15740. https://doi.org/10.17159/sajs.2023/15740

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 March 2023

To 'train' or to 'educate' for doctoral work? – that is the question: A review of *Doctoral Training and Higher Education in Africa*

Doctoral Training and Higher Education in Africa is both timely and essential for a transforming and decolonising higher education. A combination of an ageing professoriate and a growing number of younger, newer and less experienced academics entering academia in Africa[1] make this book even more critical. Its deliberateness in offering varying insights on doctoral training from across Africa enables the reader access to unique discourses and practices that guide doctoral support in ways that other similar publications on the subject have not.

The timeliness of the book also lies in its contribution to the role higher education needs to play in economic development. Africa is a context in which, more than before, higher education can be said to be increasingly becoming one of the hopes for ensuring economic development, innovation and the necessary socio-political advancement for society at large.[2] Doctoral work is certainly critical in the context of such imperatives, for it is through it that an informed knowledge base equipped to tackle local and global challenges could be found.

The book is also critical as a political project. As some will rightly argue, disguised in neo-liberal discourses such as globalisation and/or internationalisation, former colonisers seem to have managed to control, and in some sense shape, higher education on the African continent.[3] Calls for Africans to take back knowledge domains, redefine and locate them within the local have been deafening.[4] In some respects, this book represents a response to such calls. Focusing on what it calls doctoral training, the book engages with the subject matter in ways that fit distinctly African needs and contexts, thereby revitalising African higher education.

One of the aspects that seems to have received cursory attention in the book, despite the attention it has attracted in the recent past, is the need to continue to deconstruct what it means to generate knowledge in a transforming and decolonialising higher education context.[5] More specifically, given the subject of the book, it would have benefitted a reader to be exposed to an extended discussion on the possible implications for doctoral supervision of a candidate whose work needs to be locally responsive, while globally relevant. There is urgency for this as we are all grappling with ensuring that doctoral research begins from a decolonial lens, so that the knowledge generated leads to the democratisation of scholarship and learning.[6] This, I argue, is a niche that has not received the full attention it urgently requires in most publications. Given the transformation and decolonisation imperatives because of the massification of higher education, as well as concerns for epistemological access and recognition and incorporation of the knowledges of the South[7], focused attention on this aspect would have added another layer of richness.

Doctoral candidates are arguably a cohort with potential to build a scholarly trajectory as future experts in higher education decolonial research practices – a concern raised by many scholars, among whom is Chilisa[3(p.1–2)]:

> *Northern research methodologies exclude from knowledge production the formerly colonised, historically marginalized and oppressed groups, which today are most often represented …* [by] *broad categories of non-Western, third world, developing, underdeveloped, First Nations, indigenous peoples, third world women, African American women, and so on.*

What the book seems to be emphasising as key to success in Africa's doctoral journeys is the international collaboration with scholars from the North and/or Westernised contexts, and thus underplays the urgency to focus on introducing future scholars in Africa to decolonising research techniques such as indigenous research methodologies, with the intention to reduce the influence and current flow of research from domination by the North, as well as what Radcliffe[8(p.330)] refers to as the influence of the 'white and neoliberal universities' on the 'prevailing political economy of knowledge production'. What seems to be emphasised in the book, in other words, is the need for 'training' future African scholars through (among other things) sending them overseas to universities that have had a long history of 'success' in doctoral education, thereby ignoring the extent to which such contexts have been for centuries dominated by what Radcliffe[8] sees as 'white and neoliberal' worldviews. While the book rightly sees room for such 'training', I argue that there is an even more urgent need for 'educating' first.

A distinction between 'training' and 'educating', particularly in the context of doctoral work within the context of a transforming and decolonising higher education sector, could have received some extended attention in the book. A tendency to 'train' before we 'educate' is one of the underlying reasons rightly identified in the book as contributing to the doctoral throughput conundrums in most universities in Africa. By implication, the book distinguishes wisely and carefully between the two, with *training* as having to do with the act of imparting a special skill or behaviour commonly offered to individuals to perform at operational level, and *educating* as having to do with a process of systematic learning that develops a sense of judgement and reasoning. While the former involves learning something with a goal of performing a specific skill or behaviour, the latter is a systematic process of learning something with a goal of acquiring knowledge necessary to make informed, critically reflective and reflexive decisions. As eloquently argued, and effectively illustrated in the book, exposing doctoral candidates to disciplinary conceptual, theoretical and knowledge foundations, as well as relevant research traditions and paradigms prior to embarking on a doctoral research journey, continues to evidence the value of *educating* for doctoral research, before *training*.

To conclude, despite the aspects I consider to have needed some attention in the book that are either underrepresented or not covered at all, I am persuaded that, by not shying away from such issues as access to education, proactive recruitment, funding, practitioner expertise, enrolment and drop-out, this book is a great resource for higher education administrators and policymakers, as well as researchers and academics.

References

1. Hlengwa A. How are institutions developing the next generation of university teachers? Crit Stud Teach Learn. 2019;7(1):1–18. https://doi.org/10.14426/cristal.v7i1.170
2. McKenna S. Crossing conceptual thresholds in doctoral communities. Innovations in Education and International, 2017;54(5):458–466. https://doi.org/10.1080/14703297.2016.1155471
3. Chilisa B. Situating knowledge systems. In: Indigenous research methodologies. London: Sage; 2012. p. 1–42.
4. Makgoba MW, Seepe S. Knowledge and identity: An African vision of higher education transformation. In: Seepe S, editor. Towards an African identity of higher education Pretoria: Vista University and Skotaville; 2004. p. 13–58.
5. Leibowitz B, Mgqwashu EM, Kasanda C, Lefoka P, Lunga V, Shalyefu RK. Decolonising research: The use of drawings to facilitate place-based biographic research in southern Africa. J Decolonising Discipl. 2019;1(1):27–46. http://hdl.handle.net/20.500.12137/94
6. Kovach M. Conversation method in indigenous research. First Peoples Child Fam Rev. 2010;5(1):40–48. https://doi.org/10.7202/1069060ar
7. Mbembe AJ. Decolonizing the university: New directions. Arts Humanit High Educ. 2016;15(1):29–45. https://doi.org/10.1177/1474022215618513
8. Radcliffe SA. Decolonising geographical knowledges. Trans Inst Br Geogr. 2017;42(3):329–333. https://doi.org/10.1111/tran.12195

AUTHORS:
Michael J. Wingfield[1]
Brenda D. Wingfield[1]

AFFILIATION:
[1]Forestry and Agricultural Biotechnology Institute (FABI), Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Michael Wingfield

EMAIL:
mike.wingfield@fabi.up.ac.za

HOW TO CITE:
Wingfield MJ, Wingfield BD. Musings on mentorship. S Afr J Sci. 2023;119(3/4), Art. #15483. https://doi.org/10.17159/sajs.2023/15483

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
mentor, mentee, role model, career development, protégé

PUBLISHED:
29 March 2023

Musings on mentorship

Significance:

Mentorship has always been a key component of postgraduate student education, and it is becoming increasingly important in the academic environment. Strong mentorship is widely recognised as a key aspect of professional development and career success. Our view is that mentorship is a lifelong process extending beyond just teaching, and encompasses a wide range of aspects of the relationship between someone who is more experienced and someone who is less experienced. We see this as a two-way process, where both mentee and mentor benefit. It is ideally a personal relationship in which mentors are well respected in their fields, earning the respect of their mentees, who benefit from the ideas, insights and skills of their mentors. We argue that the lives and careers of both students and academics would benefit from universities establishing formal, non-bureaucratic structures to promote positive mentorship.

The notion of mentorship is becoming increasingly common in the academic environment. Its growing importance is reflected in the fact that some institutions include it in Key Performance Areas of employment. Mentorship is an interesting and complex concept because it encompasses far more than merely the idea of 'teaching others'. The word has been part of the English language since the late 1600s and it references Greek mythology. In Homer's *Odyssey,* when Odysseus set off to fight in the Trojan War, he left his son, Telemachus, in the care of his elderly friend named Mentor. Mentor was extremely wise, and his role was to act as a personal guide and advisor to the young man, providing inspiration, good counsel and good example.[1]

Today, mentorship is broadly used in the work environment and refers to many different aspects of the relationship between a less experienced and a more experienced person. Our views expressed here are, of course, based on our own academic experiences, having spent our careers in university situations where we have had the privilege of guiding mainly postgraduate students in the broad fields of microbiology and genetics.

Academics, most often towards the end of their most active years, are frequently asked to comment on what has contributed to any success they may have achieved. In our case, we have often stated publicly that whatever our accomplishments, mentorship has been central. But without being merely a shallow platitude, what exactly do we mean by mentorship? Over many years, and over many glasses of wine, we have debated this at some length. At the core, we believe mentorship is a process, not a moment in time. And it is a lifelong process. We have gained a great deal from our own mentors, and we hope our mentees have gained from us. Within a lifetime, one can have many mentors who inspire and advise. They all assist in how we overcome challenges and develop our own interests and careers.

What is mentorship? How does it differ from teaching? Most references to mentorship come from the formal work environment where senior employees are tasked with mentoring more junior staff members. This can raise the question as to where the difference lies between training, provision of advice, and mentorship. Are senior staff members in organisations or advisors to students really mentors, rather than colleagues, friends, and teachers? Our feeling is that true mentorship is a relationship that cuts more deeply than this. Moreover, whether formal or informal, it is a two-way process with learning and friendship benefitting both mentee and mentor.

Because mentorship involves two people, the relationship is intensely personal. The mentor needs to be someone whom the mentee admires, and who is well respected in their field and able to discuss ideas, skills, and insights. The mentor is also a champion of the mentee, a confidante on the academic journey. This relationship is a safe space for the mentee, and the mentor a safe companion with whom to share disappointments as well as highlights and successes. The process between the two is about developing a career by way of shared experiences. Mentors are an understanding support, but not necessarily an uncritical one – it is a collaboration around mutually beneficial empowerment and openness. Mentorship should be about building confidence, transferring skills, providing a positive role model, and understanding.

We began to think intensively about mentorship when, in 2014, we were awarded one of the first six South African government identified and supported DST (now the Department of Science and Innovation; DSI) Centres of Excellence (CoEs) managed by the National Research Foundation. The CoEs have been extremely successful and have considerably influenced South African science. Given their broad and diverse subject areas, they have incorporated mentorship in all activities. In the case of our CoE, Tree Health (now Plant) Biotechnology[2], there was sufficient funding to establish a mentorship experiment at another level. This formal mentorship programme involved identifying second- and third-year undergraduate students and pairing them with mentors from our postgraduate programme. The students who entered the programme were given some funding towards their studies and the relationships that developed between the two individuals were evaluated, by both mentors and mentees. In cases where students (mentees) and mentors were shown to have developed positive and meaningful relationships (importantly not master/slave or teacher/student), they were given a second year of support. The programme was very successful, and many past undergraduate students regard this programme as the key to their later success in postgraduate studies and early careers. Given our successful experiment, we believe that formally structured mentorship programmes have a large impact on postgraduate research programmes. The process works in at least two ways: postgraduate students experience what being a mentor means, while mentees benefit from exposure to a research environment. This enables mentees to make better choices about their future study options.

Our view, shaped by our own experience, is that mentorship should entail setting positive examples with kindness, generosity, and understanding. What is important, particularly from our experience as mentors, is that we try not

to require our graduate students to do more than we are prepared to do ourselves. In this respect, we are passionate about our research, committing long hours to this endeavour. From a mentorship perspective we tend to expect those who seek our mentorship to do the same.

Occasionally there is reference to mentors as 'tormentors' – people who are firm and demanding. Nothing is further from a successful mentorship. While we understand well that effective mentorship needs some clear structure and that there must be high levels of mutual respect and commitment, we see mentorship as a relationship. A relationship closer to a friendship bound by a clear desire by two people to learn from each other, but typically with one having more experience than the other.

We understand that our philosophy regarding postgraduate student mentorship does not conform to that of some others. In addition to 'tormenters', another approach goes by the name of 'mummy mentorship', perhaps tongue-in-cheek. We have observed mentor colleagues treating their mentees like children, even referring to them as 'the kids'. Our experience has been that postgraduate students, who are, after all, adults, typically do not respond well to being patronised. When students fail, as they will do from time to time, it is a mentor who needs to provide encouragement and support to them to continue the journey and show how improvements can be made. Guidance and encouragement are not the same as 'parenthood'.

In the past, we have spoken of 'mentorship chains' in the academic environment. This is mentorship that begins from the early years of postgraduate education, viz. a master's degree. Here, PhD students were mentors, and they themselves were mentored by postdoctoral fellows and so on up the ranks of academic life. Retirement does not end mentorship – as we ourselves have discovered. The academic and research environment is complex and achieving success can be very demanding. We often shudder at seeing early-career colleagues taking on inordinately large numbers of postgraduate students. Clearly this detrimentally affects the academics' own early careers, but it can also have negative impacts on the students too. Merely having a PhD is not necessarily the experience required to advise a PhD student. Both the qualification and experience are needed for a successful outcome. This, and other ideas on mentorship, are discussed in Malcolm Gladwell's interesting book.[3,4]

Postgraduate student mentorship is based strongly on inter-personal relationships. There is consequently no ideal model for everyone – there is no widely applicable 'blueprint'. Each student is an individual and needs to be treated as such. Personal strengths and weaknesses of graduate student mentees need to be identified and a good mentor 'fit' found. It is also worth remembering that PhD supervisors may not always be the most appropriate mentors for their students – they are in a teaching role as experts in the specific field. This topic needs further consideration and academic mentorship should be given more philosophical and practical attention.

We have raised various thoughts on graduate student mentorship that have emerged from our careers as academics. We are not experts in the field of mentorship and others may differ with our ideas. The fact that there are disagreements and differing approaches highlights the importance of the topic and the fact that more careful attention in academic environments is required. It seems that mentorship in universities is generally a topic left to develop (or not) informally and without guidance or debate. This is a situation that needs to change. We would advocate formal (but not bureaucratic) structures that promote an understanding of graduate student mentorship and how this might best be applied. This would surely improve the lives and experiences of academics, their students and overall research excellence.

Acknowledgements

The invitation to write this Commentary arose from the fact that we have both been awarded the Harry Frederik Oppenheimer Fellowship by the Oppenheimer Memorial Trust, and our having made mention to the importance of mentorship in the adjudication process. We are grateful to the many mentors who have contributed to our careers as scientists and to our past students from whom we have also gained experience in mentorship. Our friend (and writing mentor) Prof. Jane Carruthers kindly agreed to 'massage' our first draft for which we are deeply grateful. In many respects, the final product is as much hers as ours.

Competing interests

We have no competing interests to declare.

References

1. Roberts A, Chernopiskaya A. A historical account to consider the origins and associations of the term mentor. Hist Educ Soc Bull. 1999;64:81–90.
2. Steenkamp ET, Wingfield MJ. Global forest research, science education and community service positively impacted by a unique Centre of Excellence in Tree Health Biotechnology. South For. 2013;75(2):71–80. https://doi.org/10.2989/20702620.2013.800757
3. Gladwell M. Outliers. New York: Little, Brown and Company; 2008. p. VII–IX.
4. Wingfield B. How long does it take to get a PhD? S Afr J Sci. 2010;106(11/12), Art. #469. https://doi.org/10.4102/sajs.v106i11/12.469

AUTHORS:
Penelope J. Mograbi[1,2]
Emma Archer[3]
Christo Fabricius[4]
Rachel Wynberg[5]
John Donaldson[6]

AFFILIATIONS:
[1]Centre for African Ecology, School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom
[3]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
[4]CARMa-Afrika, Sustainability Research Unit, Nelson Mandela University, George, South Africa
[5]Department of Environmental and Geographical Science, University of Cape Town, Cape Town, South Africa
[6]South African National Biodiversity Institute, Cape Town, South Africa

CORRESPONDENCE TO:
Penelope Mograbi

EMAIL:
Penny.Mograbi@gmail.com

HOW TO CITE:
Mograbi PJ, Archer E, Fabricius C, Wynberg R, Donaldson J. The sustainable use of wild species benefits biodiversity and human well-being in South Africa. S Afr J Sci. 2023;119(3/4), Art. #15739. https://doi.org/10.17159/sajs.2023/15739

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
natural resources, IPBES, biodiversity economy, livelihoods, overexploitation

FUNDING:
Global Challenges Research Fund; South African National Research Foundation (grant no. 119789 and 84429); South African Department of Science and Innovation (grant no. 84429)

PUBLISHED:
29 March 2023

The sustainable use of wild species benefits biodiversity and human well-being in South Africa

Significance:

A recent report from the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) assessed how the sustainable use of wild species benefits people and nature, and which policies work best to prevent unsustainable exploitation. In the context of an accelerating and alarming biodiversity crisis, the assessment findings have important implications for South Africa, a megadiverse country with a population that relies extensively on the use of wild species for food, energy, medicine, and income, amongst many other purposes. This Commentary reflects on implications of the IPBES assessment for South Africa, drawing on insights from local contributing authors.

Introduction

The use of wild species is widespread and occurs across almost all aquatic and terrestrial ecosystems, from subsistence to global economies, and is embedded in local and global systems for food, medicine, hygiene, energy and many other uses.[1] This is certainly true of South Africa – a megadiverse country with high endemism levels and a growing human population that continues to depend on wild species to meet basic needs.

Despite a perceived disjuncture between conservation and development, the sustainable use of biodiversity can contribute significantly to South Africa's National Development Plan 2030 by reducing poverty and inequality and supporting more inclusive rural and urban economies.

Findings from the *Sustainable Use of Wild Species* assessment report, produced by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), are thus highly relevant for South Africa, and offer important insights and implications for the conservation and the sustainable use of biodiversity in the country. The report represents the first global synthesis on the use of wild species for food, energy, materials, medicine, recreation, ceremony, inspiration, and a range of other vital contributions to human well-being. It builds on IPBES' Global Assessment findings showing that overexploitation is the largest threat to life in the oceans and the second largest threat, after habitat transformation, to life on land.[2] Globally, the report can potentially influence implementation of the Convention on Biological Diversity (CBD) and the Kunming-Montreal Global Biodiversity Framework, which includes the sustainable use of biodiversity as one of four goals, and the focus of several targets.[3,4] It is also relevant to achieving the country's Sustainable Development Goals (SDGs, Box 1). The report's findings are timely for South Africa, coinciding with the development of a draft White Paper on the conservation and sustainable use of biodiversity, the redrafting of national biodiversity laws, and the development of strategies to bolster the biodiversity-based economy – a sector projected to generate ZAR47 billion by 2030.[5] The leadership role played by South Africa in the assessment is noteworthy. The five South African authors of this Commentary were among the 85 experts nominated to conduct the assessment, with one of the authors (J.D.) appointed as co-chair, thus enabling South African perspectives to inform the global assessment and its policy recommendations.

The approach adopted for the assessment categorised five broad 'practices' in the use of aquatic animals, terrestrial animals, trees and other plants, fungi and algae. The practices were fishing, gathering, logging, terrestrial animal harvesting (including hunting), and non-extractive practices (e.g. nature-based tourism). Each practice was analysed by inter-disciplinary expert teams across specific 'uses' (e.g. food and feed, materials, energy) over the last 20 years to assess the status and trends of sustainable use (Figure 1), identify causes of change in abundance and distribution of utilised wild species, explore likely futures of sustainable use, and assess policies and tools to promote sustainable use. Multiple forms of knowledge were used to develop credible, legitimate, and inclusive evidence. Innovative approaches to draw on traditional knowledge brought together many different local and indigenous voices using a range of platforms to explore different ways of knowing and being, and to recognise the centrality of wild species to the identities, cultures and livelihoods of many Indigenous peoples and local communities.

Key messages from the report's Summary for Policymakers (https://zenodo.org/record/7411847) show that the sustainable use of wild species is critical for people and nature, with over 50 000 harvested wild species documented (and many more likely used) as central to the needs and identities of billions of people, including Indigenous peoples and local communities and an estimated 70% of the world's poor. Although some 34% of wild species were estimated to be used sustainably globally, widespread unsustainable use of terrestrial and marine species has raised extinction risks for many species. Overexploitation (including global trade in wild species), landscape and seascape changes, climate change, pollution and invasive alien species impact wild species abundance and distribution, and impact negatively on those who rely on wild species for their subsistence, income, and well-being.

Here we present and discuss key messages from the assessment and their implications for South Africa.

The sustainable use of wild species is critical for South African people and nature

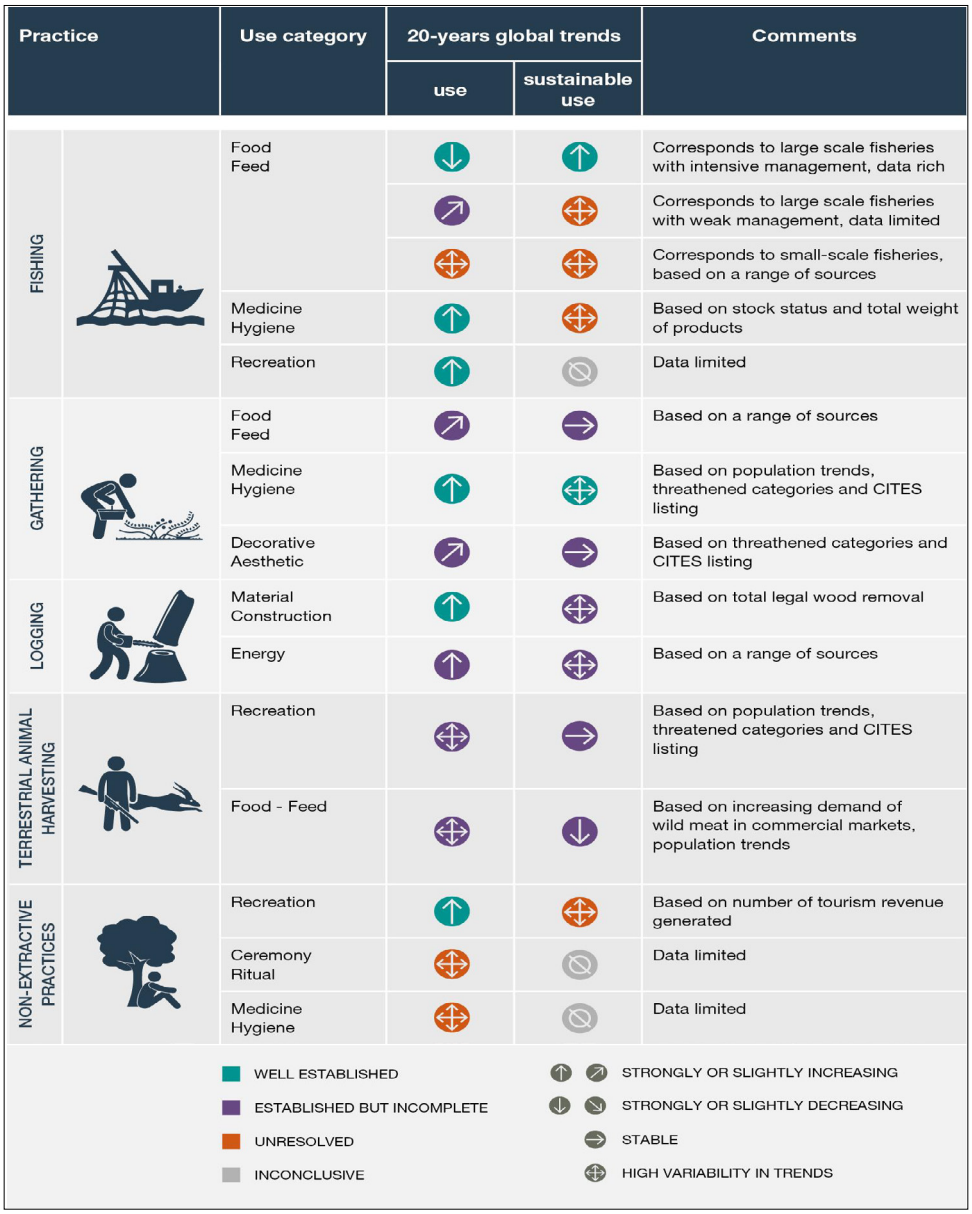
A central message of the report is that the sustainable use of wild species is critical for people and nature – a finding that resonates with the largely ubiquitous use of wild species in South Africa. For example, wild-caught commercial fisheries are worth approximately ZAR8 billion and employ some 27 000 people directly (up to 100 000 indirectly), and subsistence fishing is valued at about ZAR16 million and supports approximately 29 000 individual

subsistence fishers (2013 estimates).[6] Although the financial contribution of small-scale fisheries to GDP is low (<1%), the sector is critical for providing employment and food to poor coastal communities[6], as well as enabling well-being through connections to the ocean. More than 2000 plant species are used and traded for medicinal use, with an estimated value of about ZAR8 billion.[7] Biodiversity-related tourism contributed over ZAR30 billion in 2015 and some 418 000 biodiversity-related jobs in South Africa, with each job in conservation supporting a further five that depend on biodiversity use.[7] The tourism sector's growth can potentially contribute ZAR14 billion to GDP and double the number of jobs by 2030.[8]

Of relevance to South Africa is the high reliance of vulnerable people and communities on wild species. Wild species not only serve as a fall-back option for rural households during times of economic stress, but also add to people's livelihood security, especially for rural dwellers[9], and, if well managed, can generate significant revenue and employment[7].

Discussions frequently focus on the extractive use of wild species (i.e. plant harvesting, hunting, fishing), but wild species are also used extensively for non-extractive purposes. South Africa's wildlife-watching tourism industry is one example, but other non-tangible benefits provide an important part of our cultural heritage. Wild species are an intrinsic part of our national identity (e.g. proteas as the national emblem) and time in nature provides physical and mental well-being and, for some, connections with ancestral spirits.[10]

The report describes the global trends and regional variations in wild species use, based on systematic reviews of over 1600 sources (Figure 1). At a global level, use of wild species is mostly increasing, but its sustainability is highly variable and often unknown (Figure 1). In certain instances, wild species may be more intensively used in South Africa than elsewhere. For example, South Africa's iconic landscapes and megafauna support significant hunting, wildlife tourism and photography sectors. In



Source: IPBES[1] (CC BY 4.0)

Figure 1: Global trends in use and sustainable use of wild species from 2000 to the present. The figure shows only the top two to three most documented use categories for each practice based on systematic literature reviews using >1600 sources. Trends in use refer to an assessment of the overall state of use for wild species in relation to the specified practice, i.e. has overall use increased strongly, increased, stayed the same, decreased or decreased strongly. The multi-directional arrow depicts highly variable trends across areas or sectors for a given category of practice-use. The colours of the arrows refer to the confidence levels associated with those trends. Trends in sustainable use specifically refer to whether the intensity and form of use have been deemed sustainable over the 20-year period. The comments column contains brief reference to how the trend was determined.

Box 1: How does the sustainable use of wild species contribute to achieving the United Nations' Sustainable Development Goals?

The sustainable use of wild species directly contributes, or has the potential to contribute, to *all* of the United Nations' Sustainable Development Goals (SDGs). Conversely, unsustainable use can be detrimental to achieving global targets. For example, trade in wild foods can compete with local subsistence use and can jeopardise local food security (SDG2). Unsanitary handling and consumption of wild animals can cause illness and increase the risk of zoonotic disease (SDG3). However, sustainable use of wild species can reduce poverty by providing food (SDG1, SDG2, SDG3), medicine (SDG1, SDG3) and income (SGD1, SDG8, SDG9, SDG10). Many uses of wild species have been shown to benefit women most (SDG5, SDG10). A strong body of evidence also reveals the critical role played by wild species for people in vulnerable situations (SDG10). Income derived from trade in wild species can provide cash to support children's education and wildlife watching may provide valuable educational experiences (SDG4). Including indigenous and local knowledge into formal education systems could support local biodiversity stewardship and sustainable use of wild species (SDG4, SDG14, SDG15). Sustainable logging practices can protect water quality and reduce soil erosion (SDG8), with wood biomass forming an important energy source (SDG7). The dependence on natural resources is not limited to rural regions, with increasing documentation of wild food harvesting in urban environments (SDG11). Promoting sustainable use of wild species through cross-sectoral and multilateral cooperation (SDG17) also supports responsible consumption and production (SDG12), reduces illegal trade of wild species and associated international criminal networks (SDG16), has synergies with many climate action activities (SDG13), and supports biodiversity conservation (SDG14, SDG15).

Source: Adapted from Table 1.3 in IPBES[1]

rural areas, access to electricity is both erratic and expensive, leading to increased use of wild species for fuelwood, while many of the country's population still use traditional plant- and animal-based medicines. Many medicinal plants are harvested unsustainably, with 184 species in decline and 56 listed as threatened as a result of overexploitation[7], highlighting how unsustainable use of wild species not only threatens biodiversity, but affects livelihoods and erodes cultural practices and identity.

National-level data for wild species use are under-studied and, where available, are often not helpful for decision-making. The National Biodiversity Assessment (2018) reported almost all taxonomic groups face increased extinction risk.[7] But the true scale of the crisis remains underreported: 90% of the more than 770 recorded harvested marine taxa in South African waters have not had stock status assessments done, and the 10% of species that have been assessed, show that more than one third of stocks are overexploited or collapsed.[7] A lack of appropriate, repeatable, and comparable indicators that reflect the linked social-ecological nature of sustainable use is a finding echoed in the report, with a specific call for more information on non-extractive and social uses.

South Africa's wild species are under threat, mirroring the global biodiversity crisis reported by IPBES.[2] The major threat to terrestrial and freshwater systems is the degradation and transformation of land, driven by agriculture and aquaculture, and urban, industrial and mining development, whilst fishing is the largest pressure on marine systems.[7] Multiple threats to wild species may interact in complex ways. For example, the illegal succulent trade entices the involvement of local people who have limited socio-economic opportunities, leading to the over-harvesting of endemic species, such as those of the *Conophytum* genus, to supply a lucrative global horticultural market fuelled by social media platforms (e.g. #planttiktok). A perfect storm of overexploitation, a drying climate, overgrazing and mining are resulting in an unprecedented species decline in the world's most biodiverse desert ecosystem.[11] Illegal trade in abalone, rhino horn and a range of other wild species is also rife, exacerbated by the lack of enforcement, corruption, and the involvement of organised crime.[12]

Pathways and levers to promote sustainable use in South Africa

The final key messages outline potential pathways and levers to achieve the sustainable use of wild species in a dynamic future. Here 'levers' refer to ways of realising change when applied to 'leverage points' – areas where policy interventions could create vital change. The report grouped leverage points into seven policy categories, covered briefly below.

1. Policy options that are inclusive and participatory strengthen the sustainable use of wild species

The sustainable use of wild species is strengthened when decision-making processes are transparent and inclusive. South Africa has progressive policies and laws to enable public participation. Civil society groups have successfully advocated for citizens' participation in decision-making, lobbying for changes in legislation against, for example, unsustainable hunting practices, captive breeding and hunting of big cats, local communities' fishing rights, and settling of land claims in protected areas. However, spaces for policy influence have increasingly closed[13], hindered by a lack of awareness of citizens' rights, the reluctance of government officials to promote functional (as opposed to 'on-paper') participation, a state that has veered away from alliances with civil society, and severe inequality which hinders the rural poor from participating on equal terms with wealthy urban citizens. This is a crucial leverage point for both biodiversity and human well-being; it is well recognised that conservation strategies that meaningfully involve local communities lead to positive conservation and socio-economic outcomes.[14]

2. Policy options that recognise and support plural knowledge systems enhance the sustainable use of wild species

Policy and decision-making processes that bring scientists and traditional knowledge holders together to co-learn from diverse forms of knowledge can help to promote the sustainable use of wild species. Scientific knowledge of sustainable use, especially ecological aspects, is relatively well developed in South Africa, thanks in part to specialised academic institutions with a history of biodiversity research. Many South Africans, especially those living close to nature in rural areas, have a rich body of knowledge about wild resources and a history of using them in their cultures, traditions, and livelihoods[10], but such knowledge has yet to be properly recognised and integrated into decision-making. Reasons for this include the fragmentation of academic disciplines and government departments, particularly in the areas of agriculture, fisheries and biodiversity, and an underappreciation for the value of local knowledge. It is also essential to have inter- and trans-disciplinary teams that include those from the social sciences and humanities in addition to biophysical scientists, to best understand and address knowledge on the integrated nature of wild species and ecosystems use with humans and societies. This is a need that is often identified but rarely implemented meaningfully.

3. Policy instruments and tools need to ensure fair and equitable distributions of costs and benefits from the sustainable use of wild species

Policies underpinned by social equity, including the fair and equitable distribution of costs and benefits, are more likely to advance the goals of conservation and development.[1,14] South Africa has progressive biodiversity laws which aim to safeguard indigenous knowledge holders and local communities, ensuring that they share in the benefits of biodiversity use. These include the *National Environmental Management Biodiversity Act* and the recently promulgated *Indigenous Knowledge Systems Act*. These laws aim to address historical injustices of so-called biopiracy – the misappropriation of genetic resources and traditional

knowledge without consent or compensation, often tied to patenting and the development of commercial products. Dozens of benefit-sharing agreements have been brokered in South Africa to comply with these laws and achieve restorative justice, linked to globally traded species such as *Pelargonium sidoides* (to treat bronchitis), *Aspalathus linearis* (rooibos tea), *Sceletium tortuosum* (to treat anxiety and depression), *Aloe ferox* (wide medicinal and cosmetic uses), and many others. Although the agreements have succeeded in distributing finances and, in part, recognising traditional knowledge holders, evidence suggests they have also been highly fraught and contested, often with ambivalent outcomes.[15] Moreover, harvesters continue to receive a low price for wild resources, such as medicinal plants[16], linked largely to inequities in global and national trade chains. Such cases point towards the complexities of realising social justice in practice, even with supportive policies in place.

Rural South Africans who live with wild resources are often the most vulnerable to the costs (e.g. crop damage from wild animals) and seldom receive a fair share of the benefits (e.g. national parks' tourism proceeds[17]). Limited access to land and resources, a pervasive theme in South Africa's history of dispossession, also restricts benefit sharing of both tangible (e.g. natural resources) and intangible benefits (e.g. cultural identity and place-based attachment[18]). Policies that enable local communities to benefit from the use of wild resources at sustainable levels that match the costs of damages, alongside government and NGO support, contracts and off-take agreements, help to address the problem.

4. Context-specific policies are needed to ensure the sustainable use of wild species

One-size-fits-all policies that do not consider regional variations in wild species abundance, social and governance contexts, and land and resource ownership, are likely to be resisted by resource users who have experience of conserving and sustainably using wild resources. More functional involvement of local resource users in developing regulations, coupled with more flexible policies, could be an important lever towards more sustainable wild resource use and management. Recognition of customary law and practices forms an important part of such measures.

5. Monitoring wild species and practices is crucial to prevent species decline

South Africa, like the rest of the world, has inadequate indicators and monitoring tools for sustainable use, undermining effective decision-making and positive outcomes. Although fisheries and logging sectors have well-developed sustainability indicators, those for terrestrial harvesting and non-extractive practices are lacking. Where indicators do exist, as for national fisheries stock assessments, these are not computed for all exploited species. Across all practices, emphasis needs to be placed on developing relevant and integrated social-ecological metrics, rather than simply focusing on harvestable resources. National policy should include multiple forms of knowledge to achieve comprehensive, appropriate indicators, utilising the national wealth of local knowledge.

Monitoring is one of the central principles of good governance of common pool resources. It not only enables stakeholders to track change and adaptively manage use levels and methods, but also holds people accountable. Participatory monitoring tools are available and these, coupled with local and traditional knowledge and involvement of knowledge holders, could be a valuable lever to avoid unsustainable wild resource use practices.[19] Monitoring can be further strengthened through use of appropriate technologies (e.g. smartphone apps), community engagement and citizen science, an area where South Africa has good experience and expertise.

6. Policy instruments that are aligned and maintain coherence and consistency will be more effective

South Africa, as a signatory to global multilateral agreements (e.g. CBD and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)), is, in principle, strongly aligned with international policy agreements relating to the sustainable use of wild species. The country also has a strong body of customary law and practices that support sustainable use. However, clashing policies at the national and sub-national level (e.g. between provinces, between sectors such as agriculture, mining, and environment, or between statutory and customary governance systems) confuse resource users and are an obstacle to sustainable use. Illicit trade in wild resources can also flourish due to a lack of consensus among neighbouring countries around laws and penalties or a lack of consistent policies and laws across the supply chain.[12] There is a need to assess misalignment and, where necessary, adapt policies, and their implementation, to be more compatible across regions and sectors.

7. Robust institutions are essential to the future sustainable use of wild species

Rules and codes of conduct that support collaborative and decentralised engagement and hold users and decision-makers accountable are at the centre of sustainable resource use, while weak institutions may struggle to exercise control. This is particularly true for customary institutions managing common pool resources in cases where commercial pressure has intensified beyond the sustainable off-takes associated with traditional measures, or where customary governance has broken down. Mopane worms, for example, are in high demand in urban areas, leading to increased harvesting pressures and, often, unsustainable use that is poorly regulated in communally managed areas.[20] Analysis of likely futures for the sustainable use of wild species suggests there will be increased pressures associated with climate change, disruptive technological advancement and increasing consumption. Institutions will need to respond to these, and other changes, through constant negotiation and adaptive approaches.

Conclusion

The *Sustainable Use of Wild Species* assessment helps to guide a South African approach for sustainably using wild resources and underscores the importance of preventing biodiversity loss to maintain the benefits provided to people and nature while contributing towards achieving the country's SDGs. In South Africa, there is significant potential in finding solutions that marry development, biodiversity, and livelihoods in the use of wild species and habitats. Yet significant challenges remain, centred indisputably on policy implementation rather than the policies themselves.

A review of the conservation and use of biodiversity in South Africa, undertaken more than 20 years ago[21], concluded that major constraints precluding more effective management included a lack of capacity, inadequate skills and expertise, insufficient budgets, legal fragmentation and weak political commitment. It is disheartening to note that all of these continue to be key hurdles in securing a sustainable biodiversity economy. Moreover, while inclusive decision-making is well entrenched in the Constitution and in sectoral laws, there are major differences in the level of participation by different interest groups, and in the extent of engagement by different government departments. Local communities and traditional knowledge holders seldom have the means and finances to participate on equal terms, and while increasing efforts aim to forefront different ways of knowing, research continues to be dominated by paradigms set in the Global North.

The IPBES *Sustainable Use of Wild Species* assessment provides the impetus for placing biodiversity at the centre of sustainable development. Taking on board its recommendations, securing political commitment, allocating sufficient budgets towards its implementation, and developing the skills and capacity required, provides the opportunity for South Africa to showcase both its biodiversity heritage and its ability to implement viable solutions for people and nature.

Acknowledgements

We acknowledge the contributions of the global team that made the IPBES *Sustainable Use of Wild Species* assessment report possible: the other co-chairs (Marla R. Emery and Jean-Marc Fromentin), coordinating lead authors and chapter experts, the technical support unit (Agnès Hallosserie, Marie-Claire Danner and Daniel Kieling), and the internal and external reviewers. P.J.M. thanks the IPBES Global

Fellowship Programme for training and support. We acknowledge funding from: the Global Challenges Research Fund to the University of Edinburgh (P.J.M.); the South African National Research Foundation (NRF) through an NRF-Rated Researchers grant (no. 119789 to E.A.); and the NRF and South African Department of Science and Innovation through the South African Research Chairs Initiative (grant no. 84429 to R.W.). We also thank the two anonymous readers for their constructive feedback which has improved this Commentary.

Competing interests

We have no competing interests to declare. All authors were among the 85 experts who conducted the assessment and J.D. was co-chair.

References

1. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). Thematic assessment of the sustainable use of wild species of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn: IPBES; 2022. https://doi.org/10.5281/zenodo.6448567
2. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn: IPBES; 2019. https://doi.org/10.5281/zenodo.3831673
3. Convention on Biological Diversity (CBD). Kunming-Montreal Global Biodiversity Framework. Montreal: CBD; 2022 [cited 2023 Jan 31]. Available from: https://www.cbd.int/doc/c/e6d3/cd1d/daf663719a03902a9b116c34/cop-15-l-25-en.pdf
4. Obura D, Agrawal A, DeClerck F, Donaldson J, Dziba L, Emery MR, et al. Prioritizing sustainable use in the Kunming-Montreal Global Biodiversity Framework. PLoS Sustain Transform. 2023;2(1), e0000041. https://doi.org/10.1371/journal.pstr.0000041
5. Government of South Africa. Address by President Cyril Ramaphosa at the launch of the Biodiversity Economy Operation Phakisa, Kalahari Waterfront, Thohoyandou, Limpopo, 25 August 2018 [webpage on the Internet]. c2018 [cited 2023 Jan 30]. Available from: https://www.thepresidency.gov.za/speeches/address-president-cyril-ramaphosa-launch-biodiversity-economy-operation-phakisa%2C-kalahari
6. Brick K, Hasson R. Valuing the socio-economic contribution of fisheries and other marine uses in South Africa [document on the Internet]. c2016 [cited 2023 Mar 02]. Available from: https://cer.org.za/wp-content/uploads/2016/08/Socio-economic-Report_Web.pdf
7. Skowno AL, Poole CJ, Raimondo DC, Sink KJ, Van Deventer H, Van Niekerk L, et al. National Biodiversity Assessment 2018: The status of South Africa's ecosystems and biodiversity: Synthesis report. Pretoria: South African National Biodiversity Institute; 2019 [cited 2023 Jan 30]. Available from: http://opus.sanbi.org/jspui/handle/20.500.12143/6362
8. African Leadership University School of Wildlife Conservation (ALUSWC). The state of the wildlife economy in South Africa. Country case study for the State of the Wildlife Economy in Africa report. Kigali: African Leadership University School of Wildlife Conservation; 2020 [cited 2023 Jan 30]. Available from: http://www.sahunters.co.za/images/attach/conservation/sow_report/State%20of%20the%20Wildlife%20Economy%20in%20Africa%20-%20Case%20study%20-%20South%20Africa%20(16.11.2020).pdf
9. Shackleton C, Shackleton S. The importance of non-timber forest products in rural livelihood security and as safety nets: A review of evidence from South Africa. S Afr J Sci. 2004;100(11):658–664. https://hdl.handle.net/10520/EJC96169
10. Cocks ML, Dold T, Vetter S. 'God is my forest' – Xhosa cultural values provide untapped opportunities for conservation. S Afr J Sci. 2012;108(5/6), Art. #880. https://doi.org/10.4102/sajs.v108i5/6.880
11. South African National Biodiversity Institute (SANBI). SA's succulent plants experience unprecedented rates of decline [webpage on the Internet]. c2022 [cited 2023 Jan 03]. Available from: https://www.sanbi.org/media/sas-succulent-plants-experience-unprecedented-rates-of-decline
12. South African Anti-Money Laundering Integrated Task Force (SAMLIT). Financial flows associated with illegal wildlife trade in South Africa. SAMLIT; 2021 [cited 2023 Jan 31]. Available from: https://www.fic.gov.za/Documents/SAMLIT_IWT%20Report_November2021.pdf
13. Wynberg R, Fig D. Realising environmental rights: Civic action, leverage, and litigation. In: Langford M, Cousins B, Dugard J, Madlingozi T, editors. Socio-economic rights in South Africa: Symbols or substance? New York: Cambridge University Press; 2014.
14. Oldekop JA, Holmes G, Harris WE, Evans KL. A global assessment of the social and conservation outcomes of protected areas: Social and conservation impacts of protected areas. Conserv Biol. 2016;30(1):133–141. https://doi.org/10.1111/cobi.12568
15. Wynberg R. Biopiracy: Crying wolf or a lever for equity and conservation? Res Policy. 2023;52(2), Art. #104674. https://doi.org/10.1016/j.respol.2022.104674
16. Nwafor I, Nwafor C, Manduna I. Constraints to cultivation of medicinal plants by smallholder farmers in South Africa. Horticulturae. 2021;7(12), Art. #531. https://doi.org/10.3390/horticulturae7120531
17. Anthony B, Scott P, Antypas A. Sitting on the fence? Policies and practices in managing human-wildlife conflict in Limpopo Province, South Africa. Conserv Soc. 2010;8(3):225–240. https://doi.org/10.4103/0972-4923.73812
18. Cundill G, Bezerra JC, De Vos A, Ntingana N. Beyond benefit sharing: Place attachment and the importance of access to protected areas for surrounding communities. Ecosyst Serv. 2017;28:140–148. https://doi.org/10.1016/j.ecoser.2017.03.011
19. Cundill G, Fabricius C. Monitoring in adaptive co-management: Toward a learning based approach. J Environ Manage. 2009;90(11):3205–3211. https://doi.org/10.1016/j.jenvman.2009.05.012
20. Sekonya JG, McClure NJ, Wynberg RP. New pressures, old foodways: Governance and access to edible mopane caterpillars, *Imbrasia (=Gonimbrasia) belina*, in the context of commercialization and environmental change in South Africa. Int J Commons. 2020;14(1):139–153. https://doi.org/10.5334/ijc.978
21. Wynberg R. A decade of biodiversity conservation and use in South Africa: Tracking progress from the Rio Earth Summit to the Johannesburg World Summit on Sustainable Development. S Afr J Sci. 2002;98(5/6):233–243. https://hdl.handle.net/10520/EJC97486

AUTHORS:
Agnes Erzse[1]
Adam Balusik[1]
Petronell Kruger[1]
Evelyn Thsehla[1]
Boyd Swinburn[2]
Karen J. Hofman[1]

AFFILIATIONS:
[1]SAMRC/Wits Centre for Health Economics and Decision Science – PRICELESS SA, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Professor of Population Nutrition and Global Health, School of Population Health, University of Auckland, Auckland, New Zealand

CORRESPONDENCE TO:
Agnes Erzse

EMAIL:
agnes.erzse@wits.ac.za

HOW TO CITE:
Erzse A, Balusik A, Kruger P, Thsehla E, Swinburn B, Hofman KJ. Commentary on South Africa's syndemic of undernutrition, obesity, and climate change. S Afr J Sci. 2023;119(3/4), Art. #14776. https://doi.org/10.17159/sajs.2023/14776

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
syndemic, food environment, climate change, nutrition

FUNDING:
SAMRC/Wits Centre for Health Economics and Decision Science – PRICELESS (23108)

PUBLISHED:
29 March 2023

Commentary on South Africa's syndemic of undernutrition, obesity, and climate change

Significance:

The 2019 Lancet Commission on Obesity describes the clustering of three global problems – undernutrition, obesity, and climate change – as the Global Syndemic. The syndemic holds major health and economic consequences for South African individuals, their families and society as a whole. In this Commentary, we discuss how the syndemic presents itself in South Africa, how it arises in the context of the broader food system, and what can be done about it.

What is the problem?

Malnutrition in all forms continues to be the most preventable cause of ill health globally. The *2018 Global Nutrition Report* shows that 88% of countries (124 out of 141) had a double burden of malnutrition (i.e. the co-occurrence of two of the three main forms of malnutrition: overnutrition, undernutrition, micronutrient deficiency) and 29% (41 countries) faced a triple burden of malnutrition.[1]

This compounding burden is concentrated in low- and middle-income countries (LMICs)[2] that are experiencing rapid urbanisation and a shift towards consumption of ultra-processed food and beverage products[3]. This nutrition transition coincides with an increase in motorised transportation and contributes to the growing prevalence of obesity[4], as well as increases in greenhouse gas (GHG) emissions fuelling climate change[5]. While LMICs produce the fewest GHG emissions[6], they are disproportionally affected by the negative impact of climate change on food production and the productivity of croplands that further constrains their already low levels of nutrition and food security and consequent high rates of malnutrition[7].

In 2019, The Lancet Commission on Obesity described the clustering of three global problems – undernutrition, obesity, and climate change – as the Global Syndemic.[8] In contrast to previously siloed perceptions of undernutrition, obesity, and climate change, the Global Syndemic model calls for a systemic understanding of these problems that co-exist in time and place, actively interact with each other, and have common underlying societal drivers. Using a systems approach has been recognised as helpful when looking at addressing the intersection of these challenges. The food system is a main factor which underpins the interaction of undernutrition, obesity, and climate change, and alone contributes to one third of anthropogenic GHG emissions.[9] It gathers all the elements and activities that relate to land use for agriculture, production, processing, distribution, and waste management around food (summarised as the food supply chain, food environment, and consumer behaviour).[10] While food systems are the backbone of human health, the currently 'broken' global food system is costing us not only a healthy environment, but also healthy and productive years of our lives.

The syndemic holds major health and economic consequences for individuals, their families and society. In South Africa, the combined cost of undernutrition (ZAR62 billion per year)[11] and obesity (ZAR33 million per year)[12] is ZAR62 330 million per annum while climate change modelling shows that the effect of rising temperatures on labour availability and productivity will cost up to 20% of per capita GDP[13].

The 2020 Economic and Recovery Plan for South Africa provides a summary of the multiple socio-economic challenges that the country faces. These include gender inequality, a high unemployment rate, widespread poverty, declining economic growth and declining investment.[14] These challenges are further compounded by a debt burden that sits around 80% of the country's GDP.[14] Resources are clearly limited, hence, addressing the intertwined issues presented by the syndemic requires common strategies.

This Commentary focuses on how the Global Syndemic manifests in South Africa, and on one of its major common drivers, the food system.

Manifestation of the syndemic

Food and nutrition insecurity are serious challenges and the natural environment in the context of climate change is a significant contributor.[15] South Africa is experiencing comparatively more severe impacts than average in terms of temperature and rainfall variability.[16] There has also been increasing drought, flooding, and changes in the timing of the rainy season.[17] Agriculture contributes around 2.5% to South Africa's GDP[18], but it utilises around 6% of the total labour force – a significant labour market in a country with high rates of unemployment[19].

Despite well-developed agricultural, food and nutrition policies, in 2019 around 10 million South Africans (17.3%) were affected by food insecurity. This figure worsened to around 1 in 5 (23.6%) due to COVID-19 and its impact on the food system.[20] Food and nutrition insecurity manifests in the high levels of triple burden of malnutrition. With 68% prevalence of overweight and obesity in adult women, 31% in adult men, and 13% in children[21], South Africa has the highest obesity prevalence in sub-Saharan Africa[22]. Obesity and overweight occur simultaneously with high rates of stunting among children under 5 (27%)[21], and even within the same households. At least one obese adult was found in 45% of households with stunted children.[23] Regarding micronutrients, 40% of children under 5 have zinc deficiency and 44% have vitamin A deficiency. Iron deficiency affects 61% of children and 31% of adult women.[24]

Climate change related crop failure and loss of livestock is anticipated to lead to food shortages, and a consequent increase in food prices, which will perpetuate high rates of the triple burden of malnutrition.[25]

Why is this happening?

To understand the key drivers of the syndemic and their interaction, we conceptualised a model (Figure 1) using Downs et al.'s[26] food environment typology. Figure 1 summarises the interaction between the elements in the food environment typology and climate change, and provides an opportunity to identify entry points for action. In this section we unpack Figure 1 by discussing how the issue of poor nutrition is driven by diet, how diet is influenced by the food environment, how the food environment subsequently forms part of and is influenced by the food system, and how the food system and climate change have a bi-directional impact on one another.

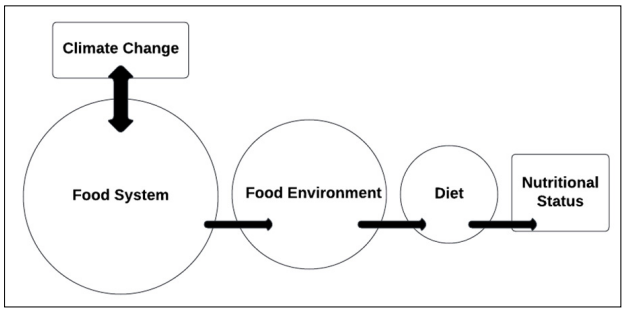


Figure 1: Conceptual model of the food environment typology and its elements driving the syndemic.

Nutritional status is determined by diet

South African diets are characterised by high fat and sugar intakes, with 40% of children in Grades 8–11 regularly consuming food high in fat (cakes, biscuits) and 50% consuming sugary beverages.[23] One quarter of South Africans' dietary intake consists of sugar, alcohol or fat.[27] The mean per capita sugar and salt intake exceeds the amount recommended by the World Health Organization, while the reverse is true for fruit and vegetable intake.[28] Furthermore, in 2017, 36% of households were categorised as having a low dietary diversity score with 60% in rural areas and 47% in urban slums.[29]

Diet is determined by food environment

How people make decisions about which foods to acquire, prepare and consume is determined by the food environment[26], which at its core is the physical, socio-cultural, economic, and political context in which individuals engage with the food system[30]. In other words, what people eat is dependent on what food is affordable, accessible, available, and desirable. These elements of the food environment are discussed in the sections below.

Price

One key driver of dietary patterns is the price of healthy foods compared to highly processed foods.[31] South Africans have been facing a worsening food environment and compromised purchasing power for healthier diets, which cost almost 70% more than less healthy alternatives.[32] In turn, a lower than recommended consumption of fruits and vegetables is partly explained by their high cost, which has risen more rapidly than that of other foods.[33] The Pietermaritzburg Economic Justice and Dignity Group's Household Affordability Index study (2021) showed how the cost of a basic nutritional food basket for a lower-income family increased by around 7% from September 2020 to August 2021.[33] This period coincides with a three-year drought in South Africa as well as the COVID-19 pandemic – potential reasons for disruptions to food production and consequent price increase.

Availability

Unhealthy dietary patterns of high intake of fat, sugar and salt are also influenced by the proximity to and the number or fast-food outlets. A 2016 study found that in Gauteng, South Africa's most densely populated province, fast-food outlets (n=1559) vastly outnumbered their healthier counterparts, formal grocery stores (n=709).[34] Furthermore, the distribution of food availability followed a social gradient, where grocery stores were predominantly available in higher socio-economic areas, while fast-food outlets were concentrated in areas with lower- to middle-income and predominantly black South African communities. A similar trend was reported for Cape Town.[35]

Manufactured demand

Food marketing has been shown to be a major influence on food attitudes, choices and literacy without cognitive effort or awareness.[36] Food marketing also socialises consumers to form emotional connections with food or develop consumption behaviour outside of traditional meal times.[37] Evidence suggests that industry practices (including food placement on shelves)[38,39] and in-store location marketing[39], subconsciously skews consumer choices towards less nutritious food options[40]. This phenomenon has been well documented and shown to form lifelong consumer behaviours, particularly among children who are the most vulnerable to persuasive messaging.[41] Unhealthy food marketing in South Africa has also been observed at alarming rates in settings where children gather, such as schools.[42] This has been observed even where the food and beverage industry actors have made public pledges to desist from such practices.[42,43] Child directed marketing of unhealthy foods and beverages, in violation of South African law regulating fair marketing practices in relation to infant and young child food products, is another example of concerning marketing practices.[44,45] Beyond children, marketing of unhealthy food and beverages have also been shown to target vulnerable socio-economic classes. For example, a study reporting on marketing practices found that food producers target advertisements of starchy food to poorer black South Africans.[46]

The food environment is determined by the food system

The choices that consumers make about food in their environment is dependent on the food systems that create these environments (Figure 1). Globally, and in South Africa, the food system (which is made up of the food environment, food supply chain and consumer behaviour) has been designed for two main purposes: to feed (but not necessarily nourish) people who can afford it, and to provide profits for those involved in food provision.[47] This makes our food system fundamentally an economic model, with inadequate focus on strengthening the other two key dimensions – social and environmental – of sustainable development. This is despite South Africa's commitment to the United Nation's 2030 Agenda for Sustainable Development and the associated Sustainable Development Goals (SDGs) including goals to end all forms of hunger and malnutrition (SDG 2), create responsible frameworks for the consumption and production of food (SDG 12), and address climate change (SDG 13).[48] While SDG 2 and SDG 13 address issues relating to the syndemic, SDG 12 specifically mentions the issues of wastage in the food system and encourages transitions towards more sustainable food systems.[49]

Food system is determined by and contributes to climate change

The bi-directional relationship between the food system and climate change in Figure 1 relates to two key elements in the food system, the food supply chain and underlying consumer behaviour. A helpful start in explaining how consumer behaviour influences the climate is Bennett's Law in agricultural economics, which states that "as people become wealthier, they switch from simple starchy plant-dominated diets to a more varied food input that includes a range of vegetables, fruit, dairy products, and especially meat"[50]. The consumer-driven diversification of food supply, a trend well documented in developing countries[51], requires more resources to produce, with negative consequences on the environment and climate. These consequences include the loss of natural ecosystems due to increased demand for land conversion for agricultural production[52], and twice as much carbon emissions from meat production than that of vegetables[53,54]. While consumer behaviour has a clear impact on the environment, the reverse is also true. Climate change influences consumer behaviour as it decreases the potential kilocalorie production and hence the quantity of food available and ultimately consumer choice.[55]

Further, climate change-induced decrease in crop yields across the globe influences food supply chains, forcing nations to retain food production and production capabilities for regional purposes.[56] While South Africa is a net exporter of agricultural products, it is dependent on the imports of inputs (such as fertiliser and plant-protection chemicals) required to produce this surplus that is exported.[57] As such, the indirect impact of climate change on the international food supply impacts South Africa.

Food systems are affected by changes in agricultural practices, forced migration due to climate change and destruction of food infrastructure due to climate-related hazards. While the increasing numbers of floods and droughts in South Africa already pose a severe economic threat to the agriculture sector and its ability to provide food[58], the Intergovernmental Panel on Climate Change (IPCC) estimates that agricultural productivity will further decline from 21% to 9% in sub-Saharan African by 2080 due to climate change[7]. Unless the negative impacts of climate change on food production are anticipated and mitigated, climate change will only serve to worsen food security for South Africans, further increasing under- and overnutrition.[25]

While production in the food supply chain is impacted by climate change, there are also elements in the supply chain which impact the environment. For example, South Africa generates around 10.2 million tonnes of food waste each year, which has both environmental effects due to resulting GHG emissions during decomposition and also increases food insecurity through wastage being factored into food prices.[59]

What can be done?

The 2018 Global Nutrition Report Executive Summary emphasises five critical steps to tackle this syndemic.[60] These steps below provide a good starting point to address the syndemic in South Africa and transform the food system so that it promotes environmental sustainability, human health, social equality, and economic prosperity.

Step 1: Breaking down of silos and *developing comprehensive programmes*. Efforts should focus on double or triple duty actions that simultaneously address the common drivers of two or more issues of the syndemic. In doing so, researchers, policymakers, and donors all need to strive to identify an evidence base of systemic drivers and actions. Engagement with all stakeholders is critical – including those affected by the challenges (people living with obesity, in an obesogenic food environment), those who intentionally or unintentionally create the unhealthy systems, and those who are trying to change these systems.

Step 2: Prioritising and investing in the data needed and capacity to use it. By better understanding and investing in geospatial data, the impact of climate change on the food system and its links to malnutrition and obesity can be better assessed.

Step 3: Scaling up and diversifying financing for nutrition. Fiscal policies, such as the sugar-sweetened beverage (SSB) tax[61], are great examples of triple duty actions. These can both incentivise consumers to make better nutritional choices and generate revenue that can be used to finance the sustainable transformation of the food system. Evidence from New Zealand shows how a junk-food and SSB tax was estimated to not only reduce GHG emissions but also provide potential savings for the health system.[62] Besides broader-scale nutrition financing initiatives like the one in New Zealand, solutions also exist on a smaller scale. For example, the provision of financial incentives and support schemes for small-scale farmers in South Africa could facilitate job creation, increase financial and food security, and could help reduce the effects of international food supply shortages by diversifying the food environment available to South Africans. Furthermore, small-scale farming systems are often more environmentally sustainable as farmers have more of a vested interest in the long-term productivity of the land, and local production helps reduce the climate emissions generated from the food import-export industry.[63]

Step 4: Focusing on healthy diets to drive better nutrition. Government measures play a critical role in reducing the consumption of ultra-processed foods, which not only fuels obesity and nutrition-related non-communicable diseases but also contributes to stunting and micronutrient deficiencies by displacing more nutritious whole foods[8], and reducing climate change and biodiversity damage linked to its globalised supply chains[64]. Evidence-based and tested policy recommendations such as marketing bans of unhealthy food and beverage products to children, and easy to decipher front-of-pack warning labels are needed.[8]

Step 5: Improving the targets and commitments that are driving actors. Strong governance of actors in the food system is essential to prevent any further damage to health systems and the food environment. Governments should introduce mandatory evidence-based restrictions rather than allowing industries to self-regulate. There is no evidence that voluntary actions by the food and beverage industry safeguard public health.[65] The introduction of the SSB tax (known as the Health Promotion Levy) in South Africa in 2016 is an example of how the government can drive actors and incentivise nutrition – in this case, the tax decreased the average number of SSBs being consumed per individual.[66] While this is a step in the right direction, increasing the tax to the recommended 20% (as opposed to the current approximately 10%) will be necessary to magnify its effects. Moreover, given that the agriculture sector is a key determinant in both climate change and nutrition, additional work needs to be done by the government to translate their commitments to the IPCC into measurable targets that the agricultural system can commit to.

Why is action not happening?

Despite continuous endorsement by international organisations, there has been patchy progress in implementing evidence-based policies in South Africa. This has been explained by what the Lancet Commission on Obesity calls policy inertia[8] – the combined effects of strong industry opposition to policies that attempt to regulate or modify commercial actives, and inadequate political will and government reluctance to take up the battle with industry and enact regulatory and fiscal policies. This is particularly relevant in South Africa, where trade liberalisation and the prioritisation of economic growth promotes a favourable political environment for industry actors.[67] This talks to *Step 5* in the proposed strategies above – stronger efforts are required to address the current incentive system for players in the food system.

One example of the latest opposition concerns the South African sugar industry that has attempted to influence government policies through its political practices. The industry has continuously distorted the scientific evidence linking SSBs to obesity[68], promoted ineffective voluntary actions[42], and weakened and delayed evidence-based policies including the SSB tax and front-of-pack nutritional labelling[68,69].

Despite civil society organisations and public opinion polls that suggest support for fiscal policies, such as the SSB tax, these have not translated into adequate public demand for enhanced policy action. There continues to be a lack of fiscal and regulatory policies, including taxes on unhealthy foods, and strengthening the existing tax on sugary drinks which has not been increased in rate and scope since its implementation in 2018.

Conclusion

This Commentary only touches the surface of the numerous issues surrounding the syndemic of undernutrition, obesity and climate change in South Africa. However, with increasing malnutrition and worsening climate, costing billions of rands annually, this is an opportune time to review the drivers of these major challenges and search for comprehensive and efficient approaches in tackling complexities of the syndemic. The causes of these issues are not singular in nature as they arise from several issues including pricing, marketing, and the availability of nutritious food. There is a wide range of double and triple duty policy options to simultaneously tackle the syndemic; these include the breaking down of silos of action, improving the collection and utilisation of data, scaling up nutrition financing, focusing on healthy diets in the systems, and improving the governance structure for actors in the food system. Actions will demand a more coherent policy action and breaking down of current incentive structures between industry actors and their governance.

Acknowledgements

We thank the team from the SAMRC/Wits Centre for Health Economics and Decision Science – PRICELESS SA for their research contribution and support. A.E., A.B., P.K., E.T., and K.J.H. are supported by the SAMRC/Wits Centre for Health Economics and Decision Science, PRICELESS, University of Witwatersrand School of Public Health, Faculty of Health Sciences, Johannesburg South Africa (23108) (https://pricelesssa.ac.za/).

Competing interests

We have no competing interests to declare.

References

1. 2018 Global nutrition report. Bristol: Development Initiatives; 2018. Available from: https://globalnutritionreport.org/reports/global-nutrition-report-2018/
2. Dobbs R, Sawers C, Thompson F, Manyika J, Woetzel J, Child P, et al. Overcoming obesity: An initial economic analysis [document on the Internet]. c2014 [cited 2022 Sep 01]. Available from: https://www.mckinsey.com/~/media/mckinsey/business%20functions/economic%20studies%20temp/our%20insights/how%20the%20world%20could%20better%20fight%20obesity/mgi_overcoming_obesity_full_report.ashx
3. Popkin BM, Adair LS, Ng SW. The global nutrition transition: The pandemic of obesity in developing countries. Nutr Rev. 2012;70(1):3–21. https://doi.org/10.1111/j.1753-4887.2011.00456.x
4. Hoffman D. Obesity in developing countries: Causes and implications. Food Nutr Agric. 2001;28:35–44.
5. Seferidi P, Scrinis G, Huybrechts I, Woods J, Vineis P, Millett C. The neglected environmental impacts of ultra-processed foods. Lancet Planet Health. 2020;4(10):E437–E438. https://doi.org/10.1016/S2542-5196(20)30177-7
6. Chancel L. Global carbon inequality over 1990–2019. Nat Sustain. 2022;5(11):931–938. https://doi.org/10.1038/s41893-022-00955-z
7. Solomon S, Intergovernmental Panel on Climate Change, editors. Climate change 2007: The physical science basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge / New York: Cambridge University Press; 2007.
8. Swinburn BA, Kraak VI, Allender S, Atkins VJ, Baker PI, Bogard JR, et al. The Global Syndemic of obesity, undernutrition, and climate change: The Lancet Commission report. The Lancet. 2019;393(10173):791–846. https://doi.org/10.1016/S0140-6736(19)30310-1
9. Crippa M, Solazzo E, Guizzardi D, Monforti-Ferrario F, Tubiello FN, Leip A. Food systems are responsible for a third of global anthropogenic GHG emissions. Nat Food. 2021;2(3):198–209. https://doi.org/10.1038/s43016-021-00225-9
10. HLPE. Nutrition and food systems. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security. Report no. 12. Rome: Committee on World Food Security; 2017. Available from: https://www.fao.org/fileadmin/user_upload/hlpe/hlpe_documents/HLPE_Reports/HLPE-Report-12_EN.pdf
11. Jamieson L, Berry L, Lake L, editors. South African Child Gauge 2017. Cape Town: Children's Institute, University of Cape Town; 2017. Available from: http://www.ci.uct.ac.za/sites/default/files/image_tool/images/367/Child_Gauge/South_African_Child_Gauge_2017/Child_Gauge_2017_lowres.pdf
12. Boachie MK, Thsehla E, Immurana M, Kohli-Lynch C, Hofman KJ. Estimating the healthcare cost of overweight and obesity in South Africa. Glob Health Action. 2022;15(1), Art. #2045092. https://doi.org/10.1080/16549716.2022.2045092
13. Shayegh S, Manoussi V, Dasgupta S. Climate change and development in South Africa: The impact of rising temperatures on economic productivity and labour availability. Clim Dev. 2021;13(8):725–735. https://doi.org/10.1080/17565529.2020.1857675
14. Parliament of the Republic of South Africa. South African Economic Reconstruction and Recovery Plan. 2020. p. 38.
15. Chersich M, Wright C, Venter F, Rees H, Scorgie F, Erasmus B. Impacts of climate change on health and wellbeing in South Africa. Int J Environ Res Public Health. 2018;15(9):1884. https://doi.org/10.3390/ijerph15091884
16. Wolski P. Twice the global rate [webpage on the Internet]. c2019 [cited 2022 Jun 13]. Available from: https://www.csag.uct.ac.za/2019/09/25/twice-the-global-rate/
17. Roffe SJ, Fitchett JM, Curtis CJ. Classifying and mapping rainfall seasonality in South Africa: A review. South Afr Geogr J. 2019;101(2):158–174. https://doi.org/10.1080/03736245.2019.1573151
18. South African Department of Agriculture, Land Reform and Rural Development. Economic review of the South African agriculture [document on the Internet]. c2020 [cited 2022 Sep 09]. Available from: https://www.dalrrd.gov.za/Portals/0/Statistics%20and%20Economic%20Analysis/Statistical%20Information/Economic%20Review%202020.pdf
19. GCRF Agricap. South Africa's agriculture and food system: A scenarios analysis [document on the Internet]. Pretoria: GCRFAFRICAP; 2018 [cited 2022 Sep 08]. Available from: https://www.namc.co.za/gcrf-africap-participatory-scenarios-workshop-report-rsa-2018/
20. Stats SA. Measuring food security in South Africa: Applying the Food Insecurity Experience Scale [document on the Internet]. c2020 [cited 2022 Sep 05]. Available from: http://www.statssa.gov.za/publications/Report-03-00-19/Report-03-00-192020.pdf
21. National Department of Health, Stats SA, SAMRC. South Africa Demographic and Health Survey 2016 [document on the Internet]. c2016 [cited 2021 Sep 04]. Available from: https://dhsprogram.com/pubs/pdf/FR337/FR337.pdf
22. Jayawardena R, Jeyakumar DT, Misra A, Hills AP, Ranasinghe P. Obesity: A potential risk factor for infection and mortality in the current COVID-19 epidemic. Diabetes Metab Syndr Clin Res Rev. 2020;14(6):2199–2203. https://doi.org/10.1016/j.dsx.2020.11.001
23. Health Promotion Research and Development Unit. The 2nd South African National Youth Risk Behaviour Survey 2008 [document on the Internet]. c2008 [cited 2022 Jun 14]. Available from: https://granthaskin.files.wordpress.com/2012/06/youth-risk-behaviour-survey-2008_final_report.pdf
24. Harika R, Faber M, Samuel F, Mulugeta A, Kimiywe J, Eilander A. Are Low intakes and deficiencies in iron, vitamin A, zinc, and iodine of public health concern in Ethiopian, Kenyan, Nigerian, and South African children and adolescents? Food Nutr Bull. 2017;38(3):405–427. https://doi.org/10.1177/0379572117715818
25. Mugambiwa SS, Tirivangasi HM. Climate change: A threat towards achieving 'Sustainable Development Goal number two' (end hunger, achieve food security and improved nutrition and promote sustainable agriculture) in South Africa. Jàmbá J Disaster Risk Stud. 2017;9(1), Art. #350. https://doi.org/10.4102/jamba.v9i1.350
26. Downs SM, Ahmed S, Fanzo J, Herforth A. Food environment typology: Advancing an expanded definition, framework, and methodological approach for improved characterization of wild, cultivated, and built food environments toward sustainable diets. Foods. 2020;9(4):532. https://doi.org/10.3390/foods9040532
27. Nutrition Division. The double burden of malnutrition. Case studies from six developing countries. Rome: Food and Agriculture Organization of the United Nations; 2006. Available from: https://www.fao.org/3/a0442e/a0442e.pdf
28. Mchiza Z, Steyn N, Hill J, Kruger A, Schönfeldt H, Nel J, et al. A review of dietary surveys in the adult South African population from 2000 to 2015. Nutrients. 2015;7(9):8227–8250. https://doi.org/10.3390/nu7095389
29. Jonah CMP. Evidence of the existence of socioeconomic-related inequality in South African diets: A quantitative analysis of the 2017 General Household Survey. World Nutr. 2019;10(4):27–42. https://doi.org/10.26596/wn.201910427-42
30. Swinburn B, Sacks G, Vandevijvere S, Kumanyika S, Lobstein T, Neal B, et al. INFORMAS (International Network for Food and Obesity/non-communicable diseases Research, Monitoring and Action Support): Overview and key principles: INFORMAS overview. Obes Rev. 2013;14:1–12.
31. Michimi A, Wimberly MC. The food environment and adult obesity in US metropolitan areas. Geospatial Health. 2015;10(2). https://doi.org/10.4081/gh.2015.368
32. Temple NJ, Steyn NP. The cost of a healthy diet: A South African perspective. Nutrition. 2011;27(5):505–508. https://doi.org/10.1016/j.nut.2010.09.005
33. Pietermaritzburg Economic Justice & Dignity Group [PMBEJD]. 2021 Household Affordability Index [webpage on the Internet]. c2021 [cited 2021 Sep 06]. Available from: https://pmbejd.org.za/index.php/household-affordability-index/
34. Ndlovu N, Day C, Sartorius B, Aagaard-Hansen J, Hofman K. Assessment of food environments in obesity reduction: A tool for public health action. South Afr Health Rev. 2018;2018(1):115–123. https://hdl.handle.net/10520/EJC-144dca8702

35. Battersby J, Peyton S. The geography of supermarkets in Cape Town: Supermarket expansion and food access. Urban Forum. 2014;25(2):153–164. https://doi.org/10.1007/s12132-014-9217-5

36. Chandon P, Wansink B. Does food marketing need to make us fat? A review and solutions. Nutr Rev. 2012;70(10):571–593. https://doi.org/10.1111/j.1753-4887.2012.00518.x

37. Wansink B, Sobal J. Mindless eating: The 200 daily food decisions we overlook. Environ Behav. 2007;39(1):106–123. https://doi.org/10.1177/0013916506295573

38. Wilson AL, Buckley E, Buckley JD, Bogomolova S. Nudging healthier food and beverage choices through salience and priming. Evidence from a systematic review. Food Qual Prefer. 2016;51:47–64. https://doi.org/10.1016/j.foodqual.2016.02.009

39. Glanz K, Bader MDM, Iyer S. Retail grocery store marketing strategies and obesity. Am J Prev Med. 2012;42(5):503–512. https://doi.org/10.1016/j.amepre.2012.01.013

40. Erasmus MG. Proposed strategies for South African supermarkets to increase healthier food choices: A literature review. Am J Prev Med. 2019;4:11.

41. Boyland EJ, Halford JCG. Television advertising and branding. Effects on eating behaviour and food preferences in children. Appetite. 2013;62:236–241. https://doi.org/10.1016/j.appet.2012.01.032

42. Erzse A, Christofides N, Stacey N, Lebard K, Foley L, Hofman K. Availability and advertising of sugar sweetened beverages in South African public primary schools following a voluntary pledge by a major beverage company: A mixed methods study. Glob Health Action. 2021;14(1), Art. #1898130. https://doi.org/10.1080/16549716.2021.1898130

43. Moodley G, Christofides N, Norris SA, Achia T, Hofman KJ. Obesogenic environments: Access to and advertising of sugar-sweetened beverages in Soweto, South Africa, 2013. Prev Chronic Dis. 2015;12, Art. #140559. https://doi.org/10.5888/pcd12.140559

44. Lake L, Kroon M, Sanders D, Goga A, Witten C, Swart R, et al. Child health, infant formula funding and South African health professionals: Eliminating conflict of interest. S Afr Med J. 2019;109(12):902. https://doi.org/10.7196/SAMJ.2020.v110i4.14611

45. Yamoah DA, De Man J, Onagbiye SO, Mchiza ZJ. Exposure of children to unhealthy food and beverage advertisements in South Africa. Int J Environ Res Public Health. 2021;18(8):3856. https://doi.org/10.3390/ijerph18083856

46. Abrahams Z, Temple NJ, Mchiza ZJ, Steyn NP. A study of food advertising in magazines in South Africa. J Hunger Environ Nutr. 2017;12(3):429–441. https://doi.org/10.1080/19320248.2016.1227757

47. Igumbor EU, Sanders D, Puoane TR, Tsolekile L, Schwarz C, Purdy C, et al. "Big Food," the consumer food environment, health, and the policy response in South Africa. PLoS Med. 2012;9(7), e1001253. https://doi.org/10.1371/journal.pmed.1001253

48. South African Government. South Africa's Implementation of the 2030 Agenda for Sustainable Development [document on the Internet]. c2019 [cited 2022 Aug 30]. Available from: https://sustainabledevelopment.un.org/content/documents/23402SOUTH_AFRICA_RSA_Voluntary_National_Review_Report_Final__14_June_2019.pdf

49. United Nations. SDG 12 Responsible Consumption and Production [webpage on the Internet]. [cited 2022 Aug 30]. Available from: https://www.un.org/sustainabledevelopment/sustainable-consumption-production/

50. Gouel C, Guimbard H. Nutrition transition and the structure of global food demand. Am J Agric Econ. 2019;101(2):383–403. https://doi.org/10.1093/ajae/aay030

51. Popkin BM. The nutrition transition in the developing world. Dev Policy Rev. 2003;21(5–6):581–597. https://doi.org/10.1111/j.1467-8659.2003.00225.x

52. Godfray HCJ. Food for thought. Proc Natl Acad Sci USA. 2011;108(50):19845–19846. https://doi.org/10.1073/pnas.1118568109

53. Poore J, Nemecek T. Reducing food's environmental impacts through producers and consumers. Science. 2018;360(6392):987–992. https://doi.org/10.1126/science.aaq0216

54. Pye A, Bash K, Joiner A, Beenstock J. Good for the planet and good for our health: The evidence for whole-food plant-based diets. BJPsych Int. 2022;19(4):90–92. https://doi.org/10.1192/bji.2022.7

55. Kabubo-Mariara J, Mulwa RM, Di Falco S. The impact of climate change on food calorie production and nutritional poverty: Evidence from Kenya. Gothenburg: Environment for Development Initiative; 2016. http://www.jstor.org/stable/resrep15061

56. De Guzman C. Climate crisis is driving food nationalism and changing global trade. Time. 12 July 2022 [cited 2022 Jul 20]. Available from: https://time.com/6195984/climate-change-food-security-trade/

57. Campbell R. South African agriculture is dependent on imports for key inputs. Engineering News. 01 May 2020 [cited 2020 Jul 20]. Available from: https://www.engineeringnews.co.za/article/south-african-agriculture-is-dependent-on-imports-for-key-inputs-2020-05-01

58. World Bank. World Bank Climate Change Knowledge Portal [webpage on the Internet]. No date [cited 2022 Aug 30]. Available from: https://climateknowledgeportal.worldbank.org/

59. South African Department of Environment, Forestry and Fisheries (DEFF), CSIR. Food waste prevention & management: A guideline for South Africa. Report no. 1. Pretoria: DEFF, CSIR; 2021.

60. 2018 Global nutrition report: Executive summary. Bristol: Development Initiatives; 2018. Available from: https://globalnutritionreport.org/reports/global-nutrition-report-2018/

61. World Health Organization. Fiscal policies to promote healthy diets: Policy brief [webpage on the Internet]. c2022 [cited 2022 Jun 28]. Available from: https://www.who.int/publications-detail-redirect/9789240049543

62. Grout L, Mizdrak A, Nghiem N, Jones AC, Blakely T, Ni Mhurchu C, et al. Potential effect of real-world junk food and sugar-sweetened beverage taxes on population health, health system costs and greenhouse gas emissions in New Zealand: A modelling study. BMJ Nutr Prev Health. 2022;5(1):19–35. https://doi.org/10.1136/bmjnph-2021-000376

63. Houngbo. Why small farms are key to the future of food - and how we can support them [webpage on the Internet]. c2020 [cited 2022 Sep 03]. Available from: https://www.ifad.org/en/web/latest/-/blog/why-small-farms-are-key-to-the-future-of-food-and-how-we-can-support-them

64. Fardet A, Rock E. Ultra-processed foods and food system sustainability: What are the links? Sustainability. 2020;12(15):6280. https://doi.org/10.3390/su12156280

65. Erzse A, Karim SA, Foley L, Hofman KJ. A realist review of voluntary actions by the food and beverage industry and implications for public health and policy in low- and middle-income countries. Nat Food. 2022;3(8):650–663. https://doi.org/10.1038/s43016-022-00552-5

66. Stacey N, Edoka I, Hofman K, Swart EC, Popkin B, Ng SW. Changes in beverage purchases following the announcement and implementation of South Africa's Health Promotion Levy: An observational study. Lancet Planet Health. 2021;5(4):E200–E208. https://doi.org/10.1016/S2542-5196(20)30304-1

67. Thow AM, Abdool Karim S, Mukanu MM, Ahaibwe G, Wanjohi M, Gaogane L, et al. The political economy of sugar-sweetened beverage taxation: An analysis from seven countries in sub-Saharan Africa. Glob Health Action. 2021;14(1), Art. #1909267. https://doi.org/10.1080/16549716.2021.1909267

68. Abdool Karim S, Kruger P, Hofman K. Industry strategies in the parliamentary process of adopting a sugar-sweetened beverage tax in South Africa: A systematic mapping. Glob Health. 2020;16(1):116. https://doi.org/10.1186/s12992-020-00647-3

69. Kanter R, Vanderlee L, Vandevijvere S. Front-of-package nutrition labelling policy: Global progress and future directions. Public Health Nutr. 2018;21(8):1399–1408. https://doi.org/10.1017/S1368980018000010

AUTHOR:
Ian Moll[1]

AFFILIATION:
[1]Wits School of Education, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Ian Moll

EMAIL:
ian.moll@wits.ac.za

HOW TO CITE:
Moll I. Scientific revolution, industrial revolution, technological revolution or revolutionary technology? A rejoinder to Marwala and Ntlatlapa (S Afr J Sci. 2023;119(1/2)). S Afr J Sci. 2023;119(3/4), Art. #15767. https://doi.org/10.17159/sajs.2023/15767

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
Fourth Industrial Revolution, scientific revolution, technological revolution, information technology, ideology

PUBLISHED:
29 March 2023

Scientific revolution, industrial revolution, technological revolution or revolutionary technology? A rejoinder to Marwala and Ntlatlapa (S Afr J Sci. 2023;119(1/2))

Significance:

Critical responses by Marwala and Ntlatlapa challenged Moll's refutation of a contemporary technological revolution as a necessary but not sufficient stratum of a Fourth Industrial Revolution (S Afr J Sci. 2023;119(1/2)). This rejoinder suggests that they work with loose criteria about what counts as a 'revolution', and therefore confuse the character of industrial, scientific and technological revolutions. Therefore, their defence of the existence of a new, contemporary technological revolution, and a related economic, social and geopolitical revolution, rests on shaky conceptual ground. Neither the pandemic nor an unprecedented fusion of technologies has produced a 'Fourth Industrial Revolution'.

The Commentaries by Marwala[1] and Ntlatlapa[2] on my article[3] are welcome. One unfortunate consequence of 4IR-associated managerialism in universities is the stifling of critical academic debate. As the gigification of universities takes place, knowledge claims or research results are increasingly considered to be measurable performance units on digitised university rating scales. Robust debate of this kind is amongst the casualties.

Ntlatlapa is concerned that the title lacks precision: "[it] leaves the reader with the feeling that … a technological revolution is a fallacy". Of course, he is correct. When I formulated the title, I took it to be implicit that the 'technological revolution' it refers to is a nascent 21st-century technological revolution, a necessary but not sufficient component part of a 4IR. So the title of the article should be "Why there is no new, contemporary technological revolution, let alone a 'Fourth Industrial Revolution'". However, my argument remains that there is no 4IR, including a claim that there is no new, contemporary technological revolution.

Both Marwala and Ntlatlapa charge that to "dismiss [the 4IR] as by-product of these technological changes would be myopic"[1(p.2)]. I agree entirely. Indeed, I make it quite clear in the article that one cannot reduce an industrial revolution to a technological revolution: "An industrial revolution … is the fundamental transformation of every aspect of industrial society, including its geopolitical, cultural, macro-social, micro-social, economic *and technological* strata"[3(p.1)]. Most of my other writings on the 4IR are systematic demonstrations that there is no 4IR in broader social, cultural and geopolitical terms.[4-6] Growing global and national wealth divides, precarity of work for ordinary people, hollowing out of the middle classes, fragmentation of identity and culture, and marginalisation of the South by offshoring, outsourcing and 'onshoring back to the Cloud', are all sustained, deepening aspects of the 3IR. I repeat what I suggested in the article: "it appears increasingly clear that the 'brave new world' of the 4IR is not really happening"[3(p.1)].

However, the explicit delimitation of my article is that no grounds exist to claim that there is a technological revolution of the kind that would be a necessary part of a 4IR. Even though the technological, socioeconomic, sociopolitical and sociocultural mechanisms of an industrial revolution are functionally indivisible, it is quite legitimate to separate out any particular mechanism analytically. My focus on *technological revolution* here does not mean that I believe that an industrial revolution is purely technological. In suggesting this, both critics set my argument up as a straw person.

Ntlatlapa challenges my reading of Schwab and Marwala. Supposedly, I miss Schwab's emphasis in *The Fourth Industrial Revolution*[7] on "the confluence …[and] fusion of technologies. … any authors who challenge the notion of the 4IR as introduced by Schwab must use this as a basis"[2]. My response is twofold:

1. The question of technology convergence is addressed in the article, suggesting that Schwab's sense of a "staggering confluence of emerging technology breakthroughs"[7(p.8)] is overblown. I think that Marwala's talk of "unimaginably rapid, never-seen-before convergence"[8(p.9-10)] is equally exaggerated. Of course, this argument could be more detailed, but then, as now, there were space constraints. All that I can do here is commend Edgerton's demonstration that such "innovation-centric futurism" about technology convergences is historically misleading because it attempts to identify the fusion of new technologies at a single moment in time. In reality, "technologies appear, disappear and reappear, and mix and match across the centuries"[9(p.xii)]. For an extended discussion of this argument, see my *Debunking the Myth of the 4IR*.[6(p.43-51)]
2. Ntlatlapa's recognition that the article is "correct in so far as the technologies he [Moll] chose to analyse"[2(p.1)] is pleasing. At least this deals with popular rhetoric on the 4IR. However, there is a conceptual problem here: most of Schwab's exemplars purporting to ground his proclaimed 4IR – such as robotics and the Internet of Things – are convergences of technology. While they can be described at one level of a taxonomy as "individual advanced technologies", each is a fusion of digital (and sometimes other) technologies. Reading Schwab, one realises that most 'revolutionary' fusions of technology that he discusses could not be more mundane exemplars of digital technological changes in our society, as these examples illustrate:

- pet-tracking implants in human children[7(p.110)]
- "the internet of pipes … employ[ing] sensors in the water system to monitor flows"[7(p.75-76)]

- a shirt that can measure breathing, sweating and heart rate[7(p.116)]
- a robot capable of "picking up a part, holding it in front of an inspection station and receiving a signal to place it in a 'good' or 'not good' pile"[7(p.142)]
- the first commercial drone deliveries[10(p.49)]
- the one terabyte SD memory card[10(p.50)]

One finds the same kind of examples in Marwala's *Closing the Gap*[8], whose '4IR' exemplars include the Mercedes-Benz app warning drivers of fatigue, Facebook notification apps, Instagram shopping features, online bank interest calculators, and the 'tap-and-go' payment facility on the Gautrain[6(p.51)]. None of these are more than current iterations of 3IR technology, more in the order of a gradual evolution of things. If one examines actual technologies discussed by Schwab and Marwala themselves, there is sparse evidence of a contemporary, socially pervasive fusion of new technologies that transcends the digital revolution in some way.

These slips in the meaning of 'technological revolution' suggest little precision in the ways that Marwala and Ntlatlapa use the term 'revolution'. Marwala, in particular, seems to operate with loose criteria for what counts as a revolution. He frequently proclaims that a 4IR is 'tangible' or 'demonstrable' when his only evidence is the existential sense that technology usage is on the rise and our lives are changing because of it. Clearly, this does not establish that there is a 4IR. The point, though, is that not all 'revolutions' are equivalent. The French and Soviet Revolutions, neither of which was an industrial revolution, were not the same kind of social transformation. The former is the archetypal bourgeois political revolution; the latter was, depending on one's perspective, a peasant revolution that established socialism or an armed insurrection that brought a reactionary political elite to power. Similarly, the putative South African 'rainbow revolution', the 'machine tools revolution'[11], the 'fluoride revolution' in toothpaste[12], and the 'ice cream revolution' in which we can choose "between hot fudge … and chunky monkey"[13], are qualitatively different kinds of social transformation. If they wish precisely to determine the existence of a 4IR clearly demarcated from the 3IR, both Marwala and Ntlatlapa need a more rigorous concept of the former qua *revolution*. It is not a matter of saying it looks or feels like a revolution, therefore it must be a 4IR.

This raises another crucial distinction that neither commentator notices, between 'technological revolution' and 'revolutionary technology'.[3(p.5)] Numerous revolutionary technologies appear in scientific research contexts. The article mentions the bionic hand, nanotechnology and autonomous vehicles. Marwala offers us quantum computing and interactive computational forms as examples (note that his 'intelligent automation' refers to the use by skilled people of combinations – not fusions – of digital technologies, not 'a technology' per se). The argument[3(p.4-5)] about why the first three are not technological revolutions is essentially the same for the latter two. Most contemporary quantum computing takes the form of analytic or simulation procedures *in research contexts*, because quantum computers are not widely available.[14] It sits within what Thomas Kuhn[15] terms the *normal science* of a paradigm, rather than a paradigm shift that ruptures the research context. Claims by quantum computing researchers (as opposed to 4IR prophets) are therefore cautionary.[6(p.40),16–18] Take for example this representative opinion in 2021: "research on quantum technologies [has been] performed for decades in [international] partnerships…to push these boundaries further, this collaboration needs to continue"[14(p.19)]. Likewise, interactive computational forms are 'revolutionary' only within research programmes, despite overcoming the limitations of discrete algorithms in modelling interactive systems.[19] For example, in research on autonomous vehicles, driving in traffic-free, geofenced areas can be modelled by discrete algorithms, whereas driving in traffic that "depends on incredibly complex, unpredictable on-line events [is not yet] algorithmically or sequentially describable"[18(p.317)]. Interactive computational forms operate within paradigms.

Marwala's view of the 4IR as a Kuhnian scientific revolution is not convincing, and requires further argument. He takes Kuhn's demarcation of paradigms to imply that "the technologies of the 4IR represent a scientific revolution in itself"[1(p.2)]. However, regarding the *development* of scientific knowledge, Kuhn emphasised *the priority of paradigms* over extended time: "Normal science…is a highly cumulative enterprise, eminently successful in its aim, the steady extension of the scope and precision of scientific knowledge [including scientifically informed technological innovations]"[15(p.52)]. One could just as easily appeal to Kuhn to justify 3IR continuity to the present, as to justify an assumed 4IR rupture. In any case, the comparison of an industrial revolution with a scientific revolution is more by way of analogy than identity: the former is an all-encompassing social transformation, whereas the latter is a response to anomalies within a research programme.

Marwala's contention that the 4IR is like an intelligence revolution is misleading: "We are anticipating that machine intelligence in this era will eventually exceed the intelligence of humans"[1(p.2)]. The history of the cognition-affect dialectic in cognitive psychology[20,21] reveals why this claim is probably wrong, and a deceptive framing of the current era of networked, digital technology. On the one hand, machines have long surpassed the mechanical and computational aspects of our cognition. Adding machines already achieved this in the early 1900s. We routinely offload cognitive tasks onto computers, and there is nothing particularly distinctive about a supposed 4IR in this regard. On the other hand, it seems that even sophisticated computers are incapable of *common sense*, that crucial aspect of cognition that grounds us emotionally in the world[22]: "people are smarter than today's computers because the brain employs a basic computational architecture that is more suited to deal with [this] central aspect of the natural information processing tasks that people are so good at"[23(p.3)].

Finally, a word is necessary about the now tired claim that COVID-19 has accelerated the 4IR. Marwala suggests this is sufficient proof that there is a 4IR. Of course, there is a correlation between the pandemic and the increased use of digitally networked ICTs in government, business and education. The claim that it 'accelerated' the use of digital technology is now a truism. However, the often-repeated claim that the pandemic accelerated a 4IR is simply ideology.

In response to Marwala's and Ntlatlapa's challenges, I suggest the following in this rejoinder. First, there is no clear 'demarcation' that marks the end of the 3IR and a nascent 4IR. Second, the fact that the pandemic hastened the adoption of digital technologies does not warrant a claim that there is a 4IR, nor indeed an attendant technological revolution. Third, it is precisely a careful analysis of Schwab's arguments about a "fusion of technologies across the digital, physical and biological worlds" that leads us to skepticism about a so-called 4IR. We need to modify Marwala's final sentence – we are in the age of the ideology of the 4IR, but there is no associated technological revolution, nor indeed any industrial revolution, to speak of.

Competing interests

I have no competing interests to declare.

References

1. Marwala T. The Fourth Industrial Revolution has arrived. Comments on Moll (S Afr J Sci. 2023;119(1/2), Art. #12916). S Afr J Sci. 2023;119(1/2), Art. #15429. https://doi.org/10.17159/sajs.2023/15429
2. Ntlatlapa N. Defining the Fourth Industrial Revolution. Comments on Moll (S Afr J Sci. 2023;119(1/2), Art. #12916). S Afr J Sci.2023;119(1/2), Art. #15436. https://doi.org/10.17159/sajs.2023/15436
3. Moll I. Why there is no technological revolution, let alone a 'Fourth Industrial Revolution'. S Afr J Sci. 2023;119(1/2), Art. #12916. https://doi.org/10.17159/sajs.2023/12916
4. Moll I. The myth of the fourth industrial revolution. Theoria. 2021;68(167):1–38. https://doi.org/10.3167/th.2021.6816701
5. Moll I. The fourth industrial revolution: A new ideology. Triplec. 2022;20(1):45–61. https://doi.org/10.31269/triplec.v20i1.1297

6. Moll I. Debunking the myth of the Fourth Industrial Revolution. Occasional paper. Johannesburg: Centre for Researching Education and Labour (REAL), University of the Witwatersrand; 2022. https://doi.org/10.54223/uniwitwatersrand-10539-32846
7. Schwab K. The Fourth Industrial Revolution. Geneva: World Economic Forum; 2016.
8. Marwala T. Closing the gap: The Fourth Industrial Revolution in Africa. Johannesburg: Macmillan; 2020. https://doi.org/10.1163/9789004465619_014
9. Edgerton D. The shock of the old: Technology and global history since 1900. London: Profile; 2008.
10. Skilton M, Hovsepian F. The 4th Industrial Revolution: Responding to the impact of artificial intelligence on business. Cham: Palgrave Macmillan; 2018. https://doi.org/10.1007/978-3-319-62479-2
11. Rosenberg N. Technological change in the machine tool industry, 1840-1910. J Econ Hist. 1963;23(4):414–443. https://doi.org/10.1017/S0022050700109155
12. Whelton HP, Spencer AJ, Do LG, Rugg-Gunn AJ. Fluoride revolution and dental caries: Evolution of policies for global use. J Dent Res. 2019;98(8):837–846. https://doi.org/10.1177/0022034519843495
13. Sharpe MJ, Stalnaker T, Schuck NW, Killcross S, Schoenbaum G, Niv Y. An integrated model of action selection: Distinct modes of cortical control of striatal decision making. Annu Rev Psychol. 2019;70:53–76. https://doi.org/10.1146/annurev-psych-010418-102824
14. Möller M, Vuik C. On the impact of quantum computing technology on future developments in high performance scientific computing. Ethics Inf Technol. 2017;19:253–269. https://doi.org/10.1007/s10676-017-9438-0
15. Kuhn T. The structure of scientific revolutions. 2nd edition. Chicago, IL: University of Chicago Press; 1970.
16. Johansson MP, Krishnasamy E, Meyer N, Piechurski C. Quantum computing – A European perspective. Brussels: Partnership for Advanced Computing in Europe; 2021. Available from: https://prace-ri.eu/wp-content/uploads/TR-Quantum-Computing-A-European-Perspective.pdf
17. De Avila AB, Reiser RHS, Pilla ML, Yamin AC. State-of-the-art quantum computing simulators: Features, optimizations, and improvements for D-GM. Neurocomputing. 2020; 393:223-233. https://doi.org/10.1016/j.neucom.2019.01.118
18. Wilkens A. The way I see it: The state of quantum computing. Rice University News. 2021 August 16. Available from: https://news.rice.edu/news/2021/way-i-see-it-state-quantum-computing
19. Wegner P. Interactive foundations of computing. Theor Comput Sci. 1998;192:315–351. https://doi.org/10.1016/S0304-3975(97)00154-0
20. Damasio A. Feeling and knowing: Making minds conscious. New York: Pantheon; 2021. https://doi.org/10.1080/17588928.2020.1846027
21. Lakoff G. Why it matters how we frame the environment. Environ Commun. 2010;4(1):70–81. https://doi.org/10.1080/17524030903529749
22. Mitchell M. Artificial intelligence: A guide for thinking humans. New York: Farrar, Straus & Giroux; 2019.
23. Rumelhart D, McClelland J. Parallel distributed processing: Exploration in the microstructure of cognition. Cambridge, MA: MIT Press; 1986.

AUTHOR:
Jonathan Jansen[1]

AFFILIATION:
[1]Education Policy Studies, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Jonathan Jansen

EMAIL:
jonathanjansen@sun.ac.za

HOW TO CITE:
Jansen J. Learning from COVID-19: A social science perspective on pandemic medicine. Comments on Benatar (S Afr J Sci. 2022;118(11/12)). S Afr J Sci. 2023;119(3/4), Art. #15485. https://doi.org/10.17159/sajs.2023/15485

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
pandemics, social science, medicine, vaccine nationalism, North–South relations

PUBLISHED:
29 March 2023

Learning from COVID-19: A social science perspective on pandemic medicine. Comments on Benatar (S Afr J Sci. 2022;118(11/12))

Significance:

Understanding the public health crisis in the wake of COVID-19 requires the complementary knowledges of the biomedical and the social sciences. It also demands that we examine closely the ways in which health inequalities between the Global North and the Global South are sustained. That said, breakthroughs in African medical science, such as the discovery of the Omicron variant, place South Africa in the contradictory positions of inventor and supplicant (waiting in line for vaccines from rich countries) – something that social science and medicine are only beginning to make sense of.

As a young undergraduate science student, I was once invited to attend a progressive health forum of doctors and dentists against apartheid. It would be a life-changing experience, because up to that point, I had little understanding of the links between medicine and the social world. On the platform were two professorial giants, one Jerry Coovadia and another Solly Benatar. Benatar made a simple, but for me profound, point. Tuberculosis was not simply a disease of the body; it was a consequence of social conditions such as damp housing and poor ventilation. Years later we would think and write together about ethics, medicine and society, while I found myself drawn to fascinating journals such as *Social Science and Medicine*.

In Benatar's two recent papers in this Journal[1,2], he offers us the accumulation of a lifetime of research and activism on global public health; his sheer breadth of understanding and grasp of the issues are astounding. The first paper describes the nature of the problem and the second ventures some solutions. None of the key arguments is contestable. Public health by its very nature constitutes a global crisis, something the pandemic made crystal clear. The health crisis lies at the intersection of other problems such as climate change, environmental degradation and pandemics – all of which are expected to become even worse in the years ahead of us.

What sustains these problems? An unequal world in which the profit motive of an unbridled capitalism once again revealed the sharp divides between the Global North and the Global South during the COVID-19 crisis. Suddenly, new words were popularised in the media around the pandemic crisis, including terms like vaccine nationalism and even vaccine apartheid. The health of a poor villager in rural Africa was much more vulnerable than the health of those in the North where therapeutics and then vaccines became available on demand. Not everybody's public health crisis is the same: it depends on where you are on the planet and, dare one say this, who you are as a race.

Benatar's systematic, if sometimes pertinacious, account of the nature of the crisis is at once accurate and compelling, carrying the urgency of an activist for whom public health and politics is the same thing. Politics, in this sense, is about power and the authority of the privileged, whether within a country or among the wealthy nations to decide, in Laswellian terms, who gets what and when, and on what terms. In the logics of capitalism, even something as *broodnodig* (basic, essential) as a vaccine is subject to stockpiling, intellectual property rights, and international purchasing agreements that favour the rich and leave the poor waiting.

This raises the question of speed. Its pretentions notwithstanding, Albert Bourla's *Moonshot: Inside Pfizer's Nine-Month Race to Make the Impossible Possible*[3], sheds light on an obvious question. What made the production of a vaccine possible in less than a year? There are some useful hints in the book: the pre-existence of vaccine-making capabilities, the (risky) decision to use mRNA technologies, and the determination to produce a vaccine over the constraints of existing protocols. Most of all, the political will was there to make this life-saving vaccine against the odds. The HIV/AIDS research community, still hunting for a vaccine, must have been stunned by this exceptional behaviour on the part of the scientists and governments in the wealthy nations. But what will never again be argued when the next pandemic comes around is that laborious and lengthy trials of vaccines are inevitable.

What the Benatar papers do not explore, however, are the ways in which African medical science revealed itself as a major player in the heat of the crisis to identify variants and vaccines. It is a narrative that needs a place in the telling of the story of how we got through, or are still getting through, the devastation of COVID-19. When our scientists discovered what became known as the Omicron variant in South Africa and Botswana, it stunned the Western world. The first reaction was disbelief, because in the lingering colonial mindset, Africa is the departure point for misery and disease, not a seat of medical ingenuity and discovery; this is a theme to be explored in a forthcoming book, *Racial Logics and the Politics of Knowledge in the Biomedical Sciences in South Africa*[4]. The second reaction was punitive. For the act of discovery and transparency, South Africa was promptly served with flight bans to several Western nations and, to add insult to injury, one of them even required COVID tests be done in a third country. Needless to say, this act of arrogance and ignorance (variant-spread outsprinted travel bans) incensed South Africa's medical science community.

But the discovery and reaction usefully revealed a world in which South Africa finds itself in the contradictory position of being a major contributor in the South to medical science and at the same being a supplicant in the supply chain waiting for much-needed vaccines from a dominant North.[5] For social scientists, this was intellectually fertile soil for re-examining the changing architecture of North–South relations in knowledge production where African agency in medical science alters our understanding of a changing world.

Of course, this capacity for breakthrough medical science is not new in South Africa, only seldom acknowledged. Before and since the first heart transplant to world leadership on poverty and cardiovascular disease, and of course global leadership in HIV prevention among women, South Africa is a major player in the competitive world of medical science research. That capacity has been steadily built up over the years in our leading universities but also through international collaborations and networks that brought high-level expertise back and forth to change the ways in which African science contributes to public health. The homegrown capacity for advanced genomic sequencing composed and built up from the AIDS crisis, for example, positioned South Africa well when the novel coronavirus breached our shores.

All of this means that, in describing the global health crisis, we need to reset our parameters, not only for policy and planning but also for research and analysis, in recognition of the entangled worlds in which knowledge is now produced and from which reality public actions can be pursued. I am not sure that any solution lies with altering "a wicked economic system" as Benatar calls the capitalist world order, but there are other options for deeper understanding and informed actions.

I like very much the notion of starting with Benatar's correct observation of "the multifactorial sociological underpinnings of ongoing global crises" and the corresponding need for transdisciplinary teams to rethink approaches to public health in the light of the pandemic. This is where South African medical science can again provide leadership in the wake of what we learnt from COVID-19.

As already published in the pages of the *South African Journal of Science*, one of the critical mistakes made by the elaborately named National Coronavirus Command Council was that the body was stacked with biomedical experts without vital perspectives from the social sciences and humanities. The result was an authoritarian clampdown in poor communities for not complying with lockdown regulations; there was little grasp of the meaninglessness of the construct of 'social distancing' within cramped settlements. Nor was there an early enough understanding of the economic impacts on street vendors facing the more immediate and felt threat of hunger over that of a deadly but invisible virus. And there was little attention to the need for cultural rituals at funerals or weddings among those removed from the high ideals and sometimes abstruse language of modern medicine. In other words, sociologists, economists, anthropologists and even linguists would have been invaluable in the sensemaking required during the early months of the pandemic.

There is little attention to these complexities in Benatar's papers, and yet they offer an opportunity for South African science and scholarship to advance in these times – not only a more elaborate technological capacity in preparation for the next round of epidemics, but also a more sophisticated conceptual apparatus that brings together the best minds from the social and medical sciences to develop new languages and approaches to public health crises.

What Benatar offers in this regard in these two highly valuable papers, is a solid platform for next-generation research and scholarship that works from a more integrated and efficacious platform for intervention as well as prevention in public health crises.

Competing interests

There are no competing interests to declare.

References

1. Benatar SR. COVID-19, global health and climate change: Causes and convergences. S Afr J Sci. 2022;118(11/12), Art. #13165. https://doi.org/10.17159/sajs.2022/13165
2. Benatar SR. Health in a post-COVID-19 world. S Afr J Sci. 2022;118(11/12), Art. #14995. https://doi.org/10.17159/sajs.2022/14995
3. Bourla A. Moonshot: Inside Pfizer's nine-month race to make the impossible possible. London: HarperCollins; 2022.
4. Jansen JD, Auerbach J, editors. Racial logics and the politics of knowledge in the biomedical sciences in South Africa. Cham: Springer. Forthcoming 2023.
5. Mazzeo V, White A. Racism, colonialism and the structure of medical knowledge in the earliest period of the global COVID-19 vaccination campaign, 2020–2021. In: Jansen JD, Auerbach J, editors. Racial logics and the politics of knowledge in the biomedical sciences in South Africa. Cham: Springer. Forthcoming 2023.

AUTHOR:
Robert Mash[1]

AFFILIATION:
[1]Division of Family Medicine and Primary Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Robert Mash

EMAIL:
rm@sun.ac.za

HOW TO CITE:
Mash R. Climate change, global health and the post-COVID world. Comments on Benatar (S Afr J Sci. 2022;118(11/12)). S Afr J Sci. 2023;119(3/4), Art. #15658. https://doi.org/10.17159/sajs.2023/15658

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
climate change, global warming, planetary health, global health, primary health care

PUBLISHED:
29 March 2023

Climate change, global health and the post-COVID world. Comments on Benatar (S Afr J Sci. 2022;118(11/12))

Significance:

Our world is facing an environmental crisis with major consequences for health. From a health perspective, the choice between development and the environment is a false dichotomy. Humanity can only flourish between a solid social foundation and an ecological ceiling. We are struggling with both objectives. Renewed commitment to the primary health care approach is one sign of hope and the emergence of both Planetary Health and One Health. We need a stronger health voice in the conversation, alongside others, such as our youth and faith communities. The sustainable development goals need inner development goals across the whole of society.

The author of 'COVID-19, global health and climate change: Causes and convergences'[1] and 'Health in a post-COVID-19 world'[2] accurately paints a picture of a world in crisis, at a *kairos* moment of decision-making. Climate change is the 'canary in the coal mine' that has hopefully got our attention, but heralds a more complex and greater potential catastrophe. The environmental crisis is multifaceted, and in addition to climate change, includes a mass extinction of species, pollution of air, ocean acidification, chemical pollution, land conversion, freshwater withdrawals, nitrogen and phosphorus loading, as well as ozone layer depletion.[3] In Africa, countries must continue the quest to improve the social foundation of society, while not developing on a trajectory that further exceeds these planetary environmental limits.[3]

Commentators in South Africa continue to juxtapose the decision-making as a choice between development and the environment. For example, a recent opinion editorial stated "poverty, inequality, and unemployment are what we need to solve first. Environmentalists' needs can't always be above our needs."[4] The environmental crisis, however, is already increasing poverty, inequality and unemployment on the African continent. In the horn of Africa, climate change is causing prolonged drought, locust invasions, loss of livelihoods amongst pastoralists and farmers, and widespread hunger and malnutrition.[5] In Mozambique, cities such as Beira, are recovering from the onslaught of cyclones that have destroyed much of the city.[6] We need to simultaneously build our social foundations at the same time as protecting the planetary boundaries and finding a safe space for humanity to flourish in-between.[3]

The articles[1,2] make for depressing reading. We recently held COP27 in Africa and, as the name suggests, the governments of the world have been meeting to tackle climate change for 27 years. Despite this, the concentration of CO_2 in the atmosphere has risen progressively[7], and so far there has been little substantial change in the global use of fossil fuels relative to alternative sources of energy[8]. These meetings, under the auspices of the United Nations, require a global consensus to make decisions, and governmental delegations are increasingly tainted by lobbyists from the oil and gas industry.[9] While we know that the environmental crisis is reaching irreversible tipping points[10], and that societal collapse is historically possible[11], we continue to pursue what Greta Thunberg has referred to as "blah, blah, blah"[12]. At the same time as we seem unable to tackle the environmental crisis, we continue to struggle with improving the social foundation. Over the last 2 years, "the richest 1% of people have accumulated close to two-thirds of all new wealth created around the world… a billionaire gained roughly $1.7 million for every $1 of new global wealth earned by a person in the bottom 90 percent"[13].

Are there any 'glimmers of hope' as Benatar puts it? His article[1] refers to emerging insights into the crisis, human ingenuity and the development of new worldviews based on cooperation, the common good and changing power dynamics as signs of hope. In the health sector, I see the renewed commitment to primary health care as a sign of hope.[14] Primary health care is ultimately a worldview that embraces social justice, solidarity, equity, and responsiveness to communities. The latest thinking also includes the concept of resilience[15], particularly in relation to challenges such as COVID-19 and climate change, and responsibility for the environmental impact of the health sector[16]. The vision, however, while representing a much needed alternative paradigm, continues to be contested and poorly implemented. In Africa we struggle with multisectoral policy and action as well as community empowerment as part of the primary health care approach.[17] In South Africa, complementary policies such as national health insurance are contested by the private healthcare sector and undermined by concerns about government corruption and capacity.[18] These new worldviews are, however, also highlighted by the emergence of new disciplines such as Planetary Health and One Health. Planetary Health explores the relationship between the environmental crisis and human health[19], while One Health explores the interdependence of human health with that of the other species that share our planet[20].

The health voice has not been a loud one in the environmental crisis, but can bridge the false dichotomy between development and the environment. A failure to build the social foundation of society and a transgression of the planetary boundaries both threaten health, and the factors are interdependent not mutually exclusive. For example, in South Africa, particulate matter in the air is four times the safe level, largely due to the burning of fossil fuels and biomass.[21] This is a major risk factor for non-communicable diseases that impact on working-age adults in poor communities, further increasing poverty. The National Strategic Plan for non-communicable diseases explicitly recognises that environmental protection, economic development and social equity are all required.[21] Society's concern for the health of its people can help us understand that it is not a choice between development and the environment.

There is also increasing hope emerging from our youth, who embrace different worldviews and are motivated by the existential crisis facing their own futures. Younger people are showing leadership and taking action.[22] Faith communities are also an important source of direction and hope in Africa and many traditions have now included 'a season of creation' in their liturgical calendar to address the need to care for all creation and not just humans.[23] Many high-income countries are more secular, but even there one can see the emergence of new 'spiritual capital'[24] through movements such as the inner development goals to mirror the outer Sustainable Development Goals[25]. The inner development goals suggest that "we lack the inner capacity to deal with our increasingly complex environment and challenges" but can develop the skills and abilities needed.[25] They focus on abilities related to the self, thinking, relating, collaborating and acting.

Even within our existing paradigm we see change happening as business recognises that fossil fuels are a cul-de-sac and there is a risk of stranded assets by continuing to invest. Renewable energy is becoming a cheaper and more sustainable investment. Our energy crisis in South Africa is an interesting microcosm of the paradigmatic tensions. We need energy for development and need to transform our reliance on coal as a source of such energy. Is this a choice between development or the environment, or an opportunity for development in a way that fosters new jobs, new sources of energy and a more sustainable future?

Ultimately our hope is not in intergovernmental agreements, politicians or climate scientists, but in the collective action of people, civil society and communities. Prof. Katharine Hayhoe[26] has expressed this well[27]:

> *I don't find hope in the science of climate change – where nearly every time a new study comes out, it shows that climate is changing faster or to a greater extent than we thought. I don't find hope in politics, either – where arguments over the arrangement of deck chairs continue as the Titanic tilts at an ever more dangerous angle.*
>
> *I find hope in recognising that I am not alone; that the giant boulder of climate action isn't sitting at the bottom of a very steep hill with only a few hands on it. In reality, that boulder is already at the top of the hill. It's already rolling down the hill in the right direction. There are millions of hands on it, alongside mine. With more hands, we can make it go faster. And the more of us there are pushing it, the greater the possibility of a better future for us all. That's what gives me hope.*

References

1. Benatar SR. COVID-19, global health and climate change: Causes and convergences. S Afr J Sci. 2022;118(11/12), Art. #13165. https://doi.org/10.17159/sajs.2022/13165
2. Benatar SR. Health in a post-COVID-19 world. S Afr J Sci. 2022;118(11/12), Art. #14995. https://doi.org/10.17159/sajs.2022/14995
3. Raworth K. Doughnut economics: Seven ways to think like a 21st-century economist. London: Cornerstone Digital; 2017.
4. Moorkroft N. Equal access to reliable electricity supply a must. Cape Times. 12 January 2023.
5. Sheriff M, Mash R. Climate change and primary health care in Chakama. Afr J Prim Heal Care Fam Med. 2022;14(1), Art. #3670. https://doi.org/10.4102/phcfm.v14i1.3670
6. Lokotola CL, Uyttersprot T, Felimone P, Scheerens C. How did primary health care in Beira experience Cyclone Idai? African J Prim Heal Care Fam Med. 2022;14(1), Art. #3626. https://doi.org/10.4102/phcfm.v14i1.3626
7. NOAA Climate.gov. Climate change: Atmospheric carbon dioxide [webpage on the Internet]. c2022 [cited 2023 Jan 21]. Available from: https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide
8. Ritchie H, Roser M, Rosado R. Overview of global energy [webpage on the Internet]. c2022 [cited 2023 Jan 21]. Available from: https://ourworldindata.org/energy-overview
9. Michaelson R. 'Explosion' in number of fossil fuel lobbyists at Cop27 climate summit. The Guardian. 10 November 2022. Available from: https://www.theguardian.com/environment/2022/nov/10/big-rise-in-number-of-fossil-fuel-lobbyists-at-cop27-climate-summit
10. Rockstrom J. Tipping points and feedback loops. In: Thunberg G, editor. The climate book. London: Penguin Random House; 2022. p. 32–40.
11. Diamond J. Collapse: How societies choose to fail or survive. London: Penguin Books; 2013.
12. Carrington D. 'Blah, blah, blah': Greta Thunberg lambasts leaders over climate crisis. The Guardian. 28 September 2021. Available from: https://www.theguardian.com/environment/2021/sep/28/blah-greta-thunberg-leaders-climate-crisis-co2-emissions
13. Oxfam International. Richest 1% bag nearly twice as much wealth as the rest of the world put together over the past two years [media release on the Internet]. 16 January 2023 [cited 2023 Jan 21]. Available from: https://www.oxfam.org/en/press-releases/richest-1-bag-nearly-twice-much-wealth-rest-world-put-together-over-past-two-years
14. World Health Organization. Astana Declaration on Primary Health Care [webpage on the Internet]. c2018 [cited 2018 Dec 18]. Available from: https://www.who.int/primary-health/conference-phc/declaration.
15. World Health Organization. Primary health care measurement framework and indicators: Monitoring health systems through a primary health care lens [webpage on the Internet]. c2022 [cited 2022 Mar 13]. Available from: https://www.who.int/publications/i/item/9789240044210
16. Global Green and Healthy Hospitals [homepage on the Internet]. No date [cited 2020 Jun 09]. Available from: https://www.greenhospitals.net/
17. Mash B, Ray S, Essuman A, Burgueño E. Community-orientated primary care: A scoping review of different models, and their effectiveness and feasibility in sub-Saharan Africa. BMJ Glob Heal. 2019;4, e001489. https://doi.org/10.1136/bmjgh-2019-001489
18. Mathew S, Mash R. Exploring the beliefs and attitudes of private general practitioners towards national health insurance in Cape Town, South Africa. Afr J Prim Heal Care Fam Med. 2019;11(1), Art. #2189. https://doi.org/10.4102/phcfm.v11i1.2189
19. Myers S, Frumkin H. Planetary Health: Protecting nature to protect ourselves. Washington DC: Island Press; 2020. https://doi.org/10.5822/978-1-61091-966-1
20. World Health Organization. One Health [webpage on the Internet]. No date [cited 2023 Jan 21]. Available from: https://www.who.int/health-topics/one-health#tab=tab_1
21. South African National Department of Health (DoH). National strategic plan for the prevention and control of non-communicable diseases, 2022–2027. Pretoria: DoH; 2022. Available from: https://www.sancda.org.za/wp-content/uploads/2022/06/NCD-NSP-Final-For-Circulation_June-2022.pdf
22. Thunberg G. The most effective way to get out of this mess is to educate ourselves. In: Thunberg G, editor. The climate book. London: Penguin Random House; 2022. p. 324–327.
23. 2022 Season of creation: 'Listen to the Voice of Creation' [webpage on the Internet]. c2022 [cited 2023 Jan 21]. Available from: https://seasonofcreation.org/2022/04/04/2022-season-of-creation-listen-to-the-voice-of-creation/
24. Zohar D, Marshall I. Spiritual capital: Wealth we can live by. Oakland, CA: Berrett-Koehler Publishers; 2004.
25. Inner Development Goals [homepage on the Internet]. No date [cited 2023 Jan 21]. Available from: https://www.innerdevelopmentgoals.org/
26. Hayhoe K. Saving us: A climate scientist's case for hope and healing in a divided world. New York: Atria/One Signal Publishers; 2021.
27. Dunne D. World Environment Day: What reasons do we have to stay hopeful about our planet's future? Independent. 5 June 2021. Available from: https://www.independent.co.uk/climate-change/news/world-environment-day-2021-positive-future-b1859929.html

AUTHORS:
Hayley C. Cawthra[1]
Martin B.C. Brandt[1]
Nigel Hicks[1]
David Khoza[1]

AFFILIATION:
[1]Council for Geoscience, Pretoria, South Africa

CORRESPONDENCE TO:
Hayley Cawthra

EMAIL:
hcawthra@geoscience.org.za

HOW TO CITE:
Cawthra HC, Brandt MBC, Hicks N, Khoza D. Comments on Singh et al. (2022) 'Marine seismic surveys for hydrocarbon exploration: What's at stake?'. S Afr J Sci. 2023;119(3/4), Art. #14514. https://doi.org/10.17159/sajs.2023/14514

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
just energy transition, CO_2 capture and storage, seismic surveys, climate change, marine life

PUBLISHED:
29 March 2023

Comments on Singh et al. (2022) 'Marine seismic surveys for hydrocarbon exploration: What's at stake?'

Significance:

We write this Commentary as a reply to Singh et al.(S Afr J Sci. 118(3/4), Art. #13420). We found that Singh et al.'s article did not adequately cover a rounded viewpoint on the topic, and we highlight a different perspective, calling for a balanced review in this regard. We base our argument on two premises. First, the literature study is incomplete, which creates a misleading perception that nothing is currently being done in South Africa to transition to a low carbon economy. Second, we comment on the statements made on seismic surveys. Herewith, we request that the authors consider a corrigendum that better reflects this research space, and call for more discussion on this topic.

On the first concern, a statement in the paper[1] suggests that only one programme aims to transition South Africa's economy to a low carbon emission future, and that this is a new initiative. This is not the case. The statement in Singh et al.[1] reads as follows:

> *In November 2021, the governments of South Africa, France, Germany, the United Kingdom and the United States of America, along with the European Union, announced a long-term 'Just Energy Transition Partnership' to support South Africa's decarbonisation efforts. The partnership will mobilise an initial commitment of USD8.5 billion for the first phase of financing through various mechanisms including grants, concessional loans and investments and risk sharing instruments. The Partnership aims to prevent up to 1–1.5 gigatonnes of emissions over the next 20 years and support South Africa's move away from coal and its accelerated transition to a low emission, climate resilient economy.*

Although this is announcing a new project, we are aware of at least four ongoing programmes that focus on this topic. These are discussed below:

1. Funding is provided by the World Bank for research into carbon sequestration and this builds on more than a decade of work, thus far. The Council for Geoscience in collaboration with industry partners and government compiled an atlas in 2010 on the geological storage of carbon dioxide in South Africa[2], which identified possible onshore and offshore repositories within South Africa conforming to the prerequisites for carbon capture and storage. Since the publication of this atlas, research has expanded on three potential storage basins, namely the onshore Zululand and Algoa Basins and the offshore Durban Basin[3-9], with academic research into the viability of the offshore Orange Basin currently ongoing. CO_2 capture and storage is globally recognised as one of the key technologies in a suite of emission reduction strategies to combat anthropogenic climate change.[10,11] CO_2 capture and storage technologies linked with hydrocarbon exploitation is not a new practice – companies such as Statoil in Norway have captured and stored 22 Mt of CO_2 in offshore saline aquifers since 1996[12], largely mitigating the long-term effects of greenhouse gas emissions alluded to by Singh et al.[1] Current research within South Africa[8] indicates that individual sedimentary basins possess multiple storage reservoirs with capacities equivalent to regions of the Rotliegend sandstone in the North Sea[13]. This work is already under way and a next phase of study, or economic studies run in parallel with geological investigations, may investigate the uncertainty surrounding sustainable injection rates and to what extent storage infrastructure is feasible within a balanced energy mix (see for example Lane et al.[14]). The technologies and practices associated with geological CO_2 sequestration are all in current commercial operation, and have been so for a decade to several decades. Such commercial operations include enhanced oil recovery, acid gas (CO_2) injection, natural gas storage and CO_2 pipeline transportation. No major 'breakthrough' technological innovations appear to be required for large-scale CO_2 transportation and storage. There are, however, significant policy, legal and regulatory challenges that must be resolved before CO_2 capture and storage is widely implemented.
2. A newly instated World Bank funded project is in progress in Leandra, Mpumalanga, where the feasibility of injecting between 10 000 and 50 000 metric tons of CO_2 (per year) into underground basaltic formations will be tested in 2023, at a depth of at least 1 km below ground.[15,16] As continental flood basalts represent some of the largest geological structures on the planet, they have the potential to provide large volumes of CO_2 storage capacity to regions such as the Mpumalanga Province in South Africa, where sedimentary storage options are limited. Due to the extensive nature of such geological substrates and their mineral trapping properties, they represent important research focus points for meeting global CO_2 emissions targets, as has been illustrated through the Wallulah Project in the USA and Carbfix in Iceland.[17,18]
3. USAID and Power Africa are building a public–private partnership to improve access to clean electricity and Internet connectivity at health facilities in sub-Saharan Africa, by supporting the development of 3180 megawatts of electricity generation in South Africa through solar and wind power installations.[19]

4. South African banks are also invested in this initiative to consider the just energy transition. Nedbank's funding for renewable energy was established in 2015[20] and Investec's investment in Green Bonds since 2022[21].

Therefore, the message in the Singh et al. article[1], namely that this has not been considered in South Africa, is misleading.

On the matter of seismic surveys, we refer to two recently published papers. In Kavanagh et al.'s[22] 'Seismic surveys reduce cetacean sightings across a large marine ecosystem', they emphasise the importance of timing of seismic surveys to best mitigate against disturbance. These authors provide results on localised avoidance in this regard and we advocate for similar mitigation in planning these surveys in South African waters, before attempting to halt all exploration activities. Additionally, Carroll et al.'s[23] 'A critical review of the potential impacts of marine seismic surveys on fish & invertebrates' talks to the gap in knowledge on sound thresholds and recovery of marine fish and invertebrates. They caution against generalisations about airgun arrays among taxa until more information is available to ensure scientific validity. We underscore the importance of conducting a local study on measured harm or impact that hydrocarbon exploration through seismic surveying imposes on marine life, as this has not yet been done in South Africa. A rising demand for minerals, metals and hydrocarbons, in tandem with a rapid depletion of land-based resources and increasing global population, has led to a surge of interest in blue economies and South Africa is no exception. Therefore, finding a suitable balance between resource extraction and environmental protection is likely a more feasible option than a call for a moratorium on hydrocarbon exploration at this stage. The renewable energy space relies on a different suite of metals, and perhaps because those risks are less well understood, it seems a preferable compromise but requires further research to better constrain the trade-off.

Through this reply, and the two broad points discussed above, we appeal to Singh and colleagues[1] and the South African science community to consider a more representative literature study to present a complete picture of the just transition, and not promote the one specific 'Just Energy Transition Partnership' project. Furthermore, gaining a clearer understanding of risks associated with alternative energy options is timely.

Competing interests

We have no competing interests to declare.

References

1. Singh JA, Roux AL, Naidoo S. Marine seismic surveys for hydrocarbon exploration: What's at stake? S Afr J Sci. 2022;118(3/4), Art. #13420. https://doi.org/10.17159/sajs.2022/13420
2. Cloete M. Atlas on geological storage of carbon dioxide in South Africa. Pretoria: Council for Geoscience; 2010.
3. Chabangu N, Beck B, Hicks N, Viljoen J, Davids S, Cloete M. The investigation of CO_2 storage potential in the Algoa Basin in South Africa. Energy Procedia. 2014;63:2800–2810. https://doi.org/10.1016/j.egypro.2014.11.302
4. Chabangu N, Beck B, Hicks N, Botha G, Viljoen J, Davids S, et al. The investigation of CO_2 storage potential in the Zululand Basin in South Africa. Energy Procedia. 2014;63:2789–2799. https://doi.org/10.1016/j.egypro.2014.11.301
5. Hicks N, Davids S, Beck B, Green A. Investigation of CO_2 storage potential of the Durban Basin in South Africa. Energy Procedia. 2014;63:5200–5210. https://doi.org/10.1016/j.egypro.2014.11.551
6. Vincent CJ, Hicks N, Arenstein G, Tippmann R, Van der Spuy D, Viljoen J, et al. The proposed CO_2 test injection project in South Africa. Energy Procedia. 2013;37:6489–6501. https://doi.org/10.1016/j.egypro.2013.06.579
7. Hicks N, Green A. Sedimentology and depositional architecture of a submarine delta-fan complex in the Durban Basin, South Africa. Mar Pet Geol. 2016;78:390–404. https://doi.org/10.1016/j.marpetgeo.2016.09.032
8. Hicks N, Green A. A first assessment of potential stratigraphic traps for geological storage of CO_2 in the Durban Basin, South Africa. Int J Greenh. 2017;64:73–86. https://doi.org/10.1016/j.ijggc.2017.06.015
9. Davids S, Scharenberg M, Lang A, Hicks N, Smith V, Heinrichs M, et al. How legacy data can open new potential for carbon capture and storage in South Africa. In: Proceedings of the 14th Greenhouse Gas Control Technologies Conference; 2018 October 21–26; Melbourne, Australia. Cheltenham: IEAGHG; 2018. p. 21–26. https://doi.org/10.2139/ssrn.3365942
10. Shukla PR, Skea J, Slade R, Al Khourdajie A, Van Diemen R, McCollum D, et al. Climate change 2022: Mitigation of climate change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 2022;10:9781009157926. https://report.ipcc.ch/ar6wg3/pdf/IPCC_AR6_WGIII_FinalDraft_FullReport.pdf
11. Global CCS Institute (GCCSI). The global status of CCS 2021 [document on the Internet]. c2021 [cited 2022 Aug 04]. Available from: https://www.globalccsinstitute.com/wp-content/uploads/2021/10/2021-Global-Status-of-CCS-Report_Global_CCS_Institute.pdf
12. Ringrose PS. The CCS hub in Norway: Some insights from 22 years of saline aquifer storage. Energy Procedia. 2018;146:166–172. https://doi.org/10.1016/j.egypro.2018.07.021
13. Wilkinson M, Haszeldine RS, Mackay E, Smith K, Sargeant S. A new stratigraphic trap for CO_2 in the UK North Sea: Appraisal using legacy information. Int J Greenh. 2013;12:310–322. https://doi.org/10.1016/j.ijggc.2012.09.013
14. Lane J, Greig C, Garnett A. Uncertain storage prospects create a conundrum for carbon capture and storage ambitions. Nat Clim Change. 2021;11:925–936. https://doi.org/10.1038/s41558-021-01175-7
15. South Africa aims to bring pilot carbon capture project online in 2023. Engineering News. 23 August 2021 [cited 2022 Aug 23]. Available from: https://www.engineeringnews.co.za/article/south-africa-aims-to-bring-pilot-carbon-capture-project-online-in-2023-2021-08-23
16. Dhansay T, Maupa T, Twala M, Sibewu Z, Nengovhela V, Mudau P, et al. CO_2 storage potential of basaltic rocks, Mpumalanga: Implications for the Just Transition. S Afr J Sci. 2022;118(7/8), Art. #12396. https://doi.org/10.17159/sajs.2022/12396
17. McGrail BP, Spane FA, Sullivan EC, Bacon DH, Hund G. The Wallula basalt sequestration pilot project. Energy Procedia. 2011;4:5653–5660. https://doi.org/10.1016/j.egypro.2011.02.557
18. Matter JM, Broecker WS, Gislason SR, Gunnlaugsson E, Oelkers EH, Stute M, et al. The CarbFix Pilot Project – storing carbon dioxide in basalt. Energy Procedia. 2011;4:5579–5585. https://doi.org/10.1016/j.egypro.2011.02.546
19. USAID. Power Africa in South Africa [webpage on the Internet]. c2021 [cited 2022 Aug 04]. Available from: https://www.usaid.gov/powerafrica/south-africa
20. Singh A. Embedded generation: The next vital developments to an energy-secure South Africa [webpage on the Internet]. c2022 [cited 2022 Aug 04]. Available from: https://www.nedbank.co.za/content/nedbank/desktop/gt/en/news/corporate-and-investment-banking-news/press-releases/2022/AEFCIB2022.html
21. Slater D. Investec launches first green bond. Engineering News. 25 April 2022 [cited 2022 Aug 04]. Available from: https://www.engineeringnews.co.za/article/investec-launches-first-green-bond-2022-04-25
22. Kavanagh AS, Nykänen M, Hunt W, Richardson N, Jessopp MJ. Seismic surveys reduce cetacean sightings across a large marine ecosystem. Sci Rep. 2019;9(1), Art. #19164. https://doi.org/10.1038/s41598-019-55500-4
23. Carroll AG, Przeslawski R, Duncan A, Gunning M, Bruce B. A critical review of the potential impacts of marine seismic surveys on fish & invertebrates. Mar Pollut Bull. 2017;114(1):9–24. https://doi.org/10.1016/j.marpolbul.2016.11.038

AUTHORS:
Bernard W.T. Coetzee[1]
Izak P.J. Smit[1,2]
Simone Ackermann[3]
Kevin J. Gaston[3]

AFFILIATIONS:
[1]Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa
[2]Scientific Services, South African National Parks, George, South Africa
[3]Environment and Sustainability Institute, University of Exeter, Penryn, United Kingdom

CORRESPONDENCE TO:
Bernard Coetzee

EMAIL:
bernard.coetzee@up.ac.za

DATES:
Received: 14 July 2022
Revised: 28 Oct. 2022
Accepted: 12 Jan. 2023
Published: 29 Mar. 2023

HOW TO CITE:
Coetzee BWT, Smit IPJ, Ackermann S, Gaston KJ. The impacts of artificial light at night in Africa: Prospects for a research agenda. S Afr J Sci. 2023;119(3/4), Art. #13988. https://doi.org/10.17159/sajs.2023/13988

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITORS:
Jennifer Fitchett
Adriaan van der Walt

KEYWORDS:
ALAN, biodiversity conservation, light pollution, nocturnal ecology, sustainable development

FUNDING:
Jennifer Ward Oppenheimer Research Grant

The impacts of artificial light at night in Africa: Prospects for a research agenda

Artificial light at night (ALAN) has increasingly been recognised as one of the world's most pernicious global change drivers that can negatively impact both human and environmental health. However, when compared to work elsewhere, the dearth of research into the mapping, expansion trajectories and consequences of ALAN in Africa is a surprising oversight by its research community. Here, we outline the scope of ALAN research and elucidate key areas in which the African research community could usefully accelerate work in this field. These areas particularly relate to how African conditions present underappreciated caveats to the quantification of ALAN, that the continent experiences unique challenges associated with ALAN, and that these also pose scientific opportunities to understanding its health and environmental impacts. As Africa is still relatively free from the high levels of ALAN found elsewhere, exciting possibilities exist to shape the continent's developmental trajectories to mitigate ALAN impacts and help ensure the prosperity of its people and environment.

Significance:
We show that the African research community can usefully accelerate work into understudied aspects of ALAN, which demonstrably impacts human and environmental health. Africa presents a unique, and in places challenging, research environment to advance understanding of this global change driver.

Introduction

Since the industrial revolution (post-1750s), global artificial light at night (ALAN) has expanded dramatically.[1,2] ALAN is produced from infrastructure like streetlights, houses, factories, ports, airports, sports stadiums, car parks, billboards and car headlights. This transformation has come with pronounced benefits and costs for society. ALAN has increased work, education and leisure hours.[3] However, there is increasing concern that excessive use of ALAN constitutes a key driver of global environmental change.[4] Alarmingly, a new picture of the breadth of ALAN's potential negative impacts has grown substantially in recent years, and in particular the work has focused on impacts both on human well-being (for an overview see[5-7]) and the wider environment (for an overview see[8-13]). Although its expansion is variable in space and time[14,15], Earth's artificially lit outdoor area is estimated to have grown by 2.2% per year from 2012 to 2016.[14] The power of global satellite observable light emissions increased by at least 49% from 1992 to 2017[2] and 80% of the world's population is now living under skies affected by ALAN[14].

To date, most published studies of the impacts of ALAN have overwhelmingly been biased to Europe and North America.[8] Consequently, there is limited understanding of how ALAN may impact human and environmental health in developing nations and how the expansion of ALAN in these regions may occur. This is particularly true across the African continent.

Here, we report the findings of a literature review of the status of ALAN research across Africa, and discuss the prospects for a research agenda. We did so using expert solicitation of the field (from the authors and our broader network), together with Boolean terms in Google Scholar and SCOPUS ("light pollution AND Africa" and "artificial* + light* AND Africa"), and discarding those not directly discussing or testing for ALAN impacts in Africa – the vast majority. In addition, we conducted direct enquiries to key experts, and ad hoc searching, to identify a broad overview of relevant academic papers, which we summarise below, although we acknowledge that this may not be an exhaustive list.

Using this overview, we elucidate the key topics on which the African research community could usefully focus and accelerate work in this field. We summarise such key themes, across disciplines, to help guide advancement in this field (Table 1). In addition, while the challenges of studying ALAN are cosmopolitan[13], Africa may present unique conditions and opportunities to advance global understanding of this issue. These relate to the quantification of ALAN itself, how development may alter its spread, impacts of ALAN on vector disease transmission, introduction of ALAN in light naïve contexts, and broader influences of ALAN on the environment and on tourism. We also posit that there exists a lack of a culture and capacity in ALAN research in Africa that is hampering progress.

Artificial light at night and the African context

In contrast to much of the rest of the world, much of Africa does not as yet experience significant levels of ALAN, with exceptions being major city centres, most of South Africa, coastal West Africa, and northern Africa (Figure 1). This situation is expected to change rapidly with increasing infrastructure development and urbanisation in Africa as, together with Asia, it is experiencing the fastest rates of urbanisation globally.[32] Even within rural settings, there is increasing interest in the installation of off-grid technology[33], which may potentially increase electrification of sub-Saharan Africa where current rates are the lowest in the world (62.7% in urban areas and 18.9% in rural areas in sub-Saharan Africa, compared to 94.6% and 71.0%, respectively, at a global scale[34]).

Quantification

Quantification of the extent and dynamics of ALAN in Africa is still in its infancy. Indeed, no study has been conducted on how ALAN has changed over Africa's developmental history, perhaps because the relative extent of pristine skies

remaining means this issue has been overlooked by researchers.[14] This lack of information is especially worrying because the major contribution globally to the future growth of ALAN will come from middle-income countries, particularly China, India, and those in South America and Africa.

Three factors complicate the quantification of ALAN in Africa. The first is the quantification of ALAN from satellite data, the second is the lack of local-scale measurements, and the third is failure to measure spectra. We expand on each in turn.

First, while satellite data can provide a continental-scale picture of ALAN, as they do elsewhere, African conditions are often not specifically considered when developing such data sets for analysis. (1) Atmospheric conditions in Africa can differ vastly from those of northern latitudes which have strongly influenced how most of the remote-sensing products are calibrated, processed and applied. In particular, the high frequency of atmospheric dust across most of the continent may complicate accurate assessment of the intensity of lights emitted at night. This may be especially pronounced for regions in northern Africa which is estimated to contribute as much as 65% of annual global aeolian dust emissions[35], and is where much of the current ALAN in Africa originates[14]. (2) Processing of ALAN data from satellites commonly screens out the effects of fires by focusing on persistent lighting and excluding pixels with anomalously high radiance that occurs infrequently[36], although the impact of frequent and undocumented fires and the atmospheric effect of smoke on the calibration of satellite data is not well understood. (3) Many sub-Saharan countries are also beleaguered by frequent and intermittent disruptions to the national electrical grid supply.[37] Generators are often used either as a primary source of electricity, or when conventional electrical grids are interrupted. This may in turn lead to underestimation of artificial light emissions, if generators are not in use later at night when the remote-sensing measurements are actually made during pass-overs of satellites. One of the most prominent remote-sensed products in use today, from the Visible Infrared Imaging Radiometer Suite (VIIRS), is derived from satellite passes across Africa in the early morning, around 01:00 local time, which means light sources that are powered from electrical generators that are switched off when citizens are less active are missed completely. (4) The sparse artificial lighting associated with much of rural Africa may not be well detected by remote-sensing sensors, yet it can have localised effects that can be cumulative over larger landscapes. In sum, these four different effects may mean that documented levels of ALAN emissions across the continent may be marked underestimates.

Second, there is a dearth of local-scale measurements of either artificial light emissions or skyglow (the increased sky brightness that results from ALAN that is emitted or reflected upwards being scattered by water, dust and gas molecules in the atmosphere) across the African continent. Few measurements have been made using on-the-ground sensors, such as the Sky Quality Meter[15]. Indeed, although global networks of such devices have grown rapidly, the contribution of Africa is extremely limited. Mapping of variation in ALAN emissions across towns and cities, whether using on-the-ground[18,19,38], aerial[39] or space-borne approaches[40], has also not been conducted in Africa (excepting those satellite sources with essentially global coverage), although this can be invaluable in identifying key sources and how levels and patterns of emissions are changing.

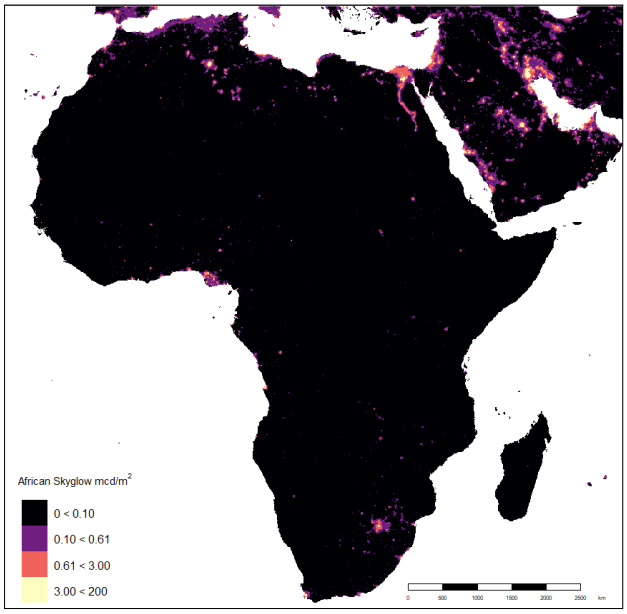
A third challenge to quantifying ALAN across Africa, as it is elsewhere, is the lack of ability to measure its spectral signature.[41] At the time of writing, no publicly readily available satellite colour images for Africa at night presently exist, but techniques to acquire them are available.[40] For Africa, such data have not been obtained, even locally by other means (e.g. night-time aerial flights).

Table 1: Exemplar research themes with key references to advance artificial light at night (ALAN) research in Africa

	Theme	Research area	Key references
1	Quantify the variation in the geography of and timing in ALAN	Measurement	16
2	Addressing atmospheric challenges to quantifying ALAN	Measurement	17
3	ALAN use transitions and trajectories	Measurement	2,17
4	Compare and contrast the remote sensed products that best capture ALAN variation across Africa	Measurement	*research frontier
5	Advancing local scale measurement and quantification of ALAN	Measurement	18,19
6	ALAN mitigation adoption rates in rural areas in Africa	Measurement	*research frontier
7	Relationship between socio-economic context, local economic growth, and ALAN use	Measurement	20
8	Quantifying the change in ALAN across Africa's developmental trajectory	Measurement	*research frontier
9	Accelerating the quantification of spectra in the night environment at ecologically relevant scales	Measurement	19
10	Quantify LED adoption rates across Africa	Measurement	*research frontier
11	Expanding research into the synergetic and/or antagonistic interactions of ALAN and other global change drivers	Measurement / Biological impacts	4
12	Diversify the range of taxa and documented impacts of ALAN on biodiversity, and advance nocturnal ecology	Biological impacts	21 *research frontier
13	Document the link between car headlights and roadkills	Biological impacts	22
14	Interactions of especially African apex predator biology with ALAN	Biological impacts	21,23 *research frontier
15	Advance work in environmental crime and ALAN links	Biological impacts	24 *research frontier
16	Assess how protected areas buffer against ALAN via skyglow	Biological impacts	25
17	Advancing larger scale and longer duration biological experiments	Biological impacts	26
18	Assess the impact of ALAN on species/populations of African taxa, especially those that are light naïve	Biological impacts	*research frontier, examples collated in main text
19	Expanding work on the impacts of ALAN on African marine animals	Biological impacts	27
20	Assess the link between ALAN expansion and astronomy/astrotourism degradation	Astronomy	28
21	Document and alter the lack of culture and capacity impeding ALAN research in Africa	Socio-economic context	*research frontier
22	Assess changes in and application of national/municipal legislation	Socio-economic context	29
23	Assessing the interactions of ALAN with medically important insects, such as disease vectors	Medical impacts	30
24	Impact of ALAN on human physiology and well-being	Medical impacts	7
25	Quantify the relationship between ALAN, human health, and environmental health	Medical impacts	31

**Note that while there is overlap with such work elsewhere in the world, we highlight what we consider to be research frontiers, particularly in Africa, which often reflects a dearth of literature.*

This is important because many of the human and environmental impacts of ALAN are spectrally sensitive, with effects often being exacerbated by emissions at blue wavelengths. Moreover, the spectrum has recently increasingly shifted across much of the world from predominantly narrow wavelength (e.g. from low pressure sodium lamps) to broad wavelength 'white' emissions (e.g. from LED lamps), both as a consequence of retrofitting of older lighting systems and the use of newer technologies in recent developments. Although similar kinds of retrofitting have to some extent occurred in Africa, growth in broad wavelength emissions is likely to occur across much of the continent through the use of such technologies when lighting is first installed in areas.



Source: Data obtained from Falchi et al.[14]

Figure 1. Patterns of skyglow across Africa. The black colour denotes natural night-time illumination levels. The purple colour describes areas in Africa where artificial light at night (ALAN) has polluted the night sky up to the horizon of the landscape but not to the zenith. Areas where skyglow pollution extends to the zenith of the sky are denoted in pink. Lastly, the lightest colour is associated with areas that are polluted to such an extent that no parts of the night sky are visible to human night vision anymore. It is worth drawing attention to the fact that ALAN in Africa is spatially scattered (and thus more widespread than a whole continent map conveys), with large areas of natural night time remaining, and that the highest intensities of skyglow are concentrated around the coastal areas. Given the expansion of ALAN, the figure is an underestimate, but there has been no subsequent calibrated work in Africa since Falchi et al.[14], which may be a fruitful avenue for continued research (see Table 1).

African development and human well-being

The human health consequences of ALAN will be similar to those documented elsewhere[5-7]. However, combined with high poverty rates and lower access to primary health care, the negative impacts of ALAN may exacerbate human health issues across the continent. There is limited understanding of how African countries with different developmental trajectories are adopting new energy portfolios, or how energy expansion will be realised. Such information is critical to further understanding of how ALAN may increase in severity and extent, and is used elsewhere to plan better for, anticipate and mitigate its impact. Estimates made by the International Energy Agency have found that the percentage of the population in sub-Saharan Africa with access to household electrification grew from 20% in 2000 to 43% in 2018 and is projected to reach 66% in 2040 if current stated development policies remain in place.[42] This growth in electrification also considers the projected growth of the human population in sub-Saharan Africa, and means that Africa can expect a substantial increase in ALAN over the next two decades.

The use of newer technologies like LED lights and solar power may reduce adverse human health impacts of lighting with kerosene. However, as is the case in other low-income nations, the increasing adoption of LED lights across Africa is also of concern because the commonly wider spectrum used has been implicated in altering a host of human and ecological processes. While the development of renewable energy is central to the economic development of Africa, careful consideration is needed on how best to mitigate the negative trade-offs that this may bring. The issue is globally pertinent – questionnaires of practitioners in the lighting fraternity have shown that there is little consensus on what the future of 'sustainable LED lighting' should look like, and much work is still required to align such a vision among stakeholders.[43]

Much of the electrification, and by extension the spread, of ALAN that is projected to occur in Africa is expected to take place in rural areas. In such areas the lowest cost option for electrification is quite often 'mini-grid' systems.[44,45] These systems allow for electrification in extremely remote regions completely independently of national grid systems, thereby accelerating the electrification and development of rural communities far beyond the pace that traditional national grid expansion can maintain.[32] This will allow for the penetration of ALAN into virtually completely artificial light naïve landscapes, at a pace that may exceed the global estimates of ALAN growth, and so worsen ALAN impacts in such areas. These mini-grid systems and associated lighting regimes will bring various social and economic benefits to those communities, but possibly also some negative environmental impacts as described above. Lessons from other parts of the world, with a longer history of ALAN, may be useful to inform how to best integrate electricity and lighting into these areas in order to mitigate some of the negative environmental effects that this may bring.

Vector disease transmission

Of particular concern for human health in Africa may be the interactions of disease vectors with ALAN. For example, Africa still has the world's highest rates of malaria, with 213 million cases in 2018, 93% of the global total.[46] Alarmingly, the work to date suggests that ALAN may alter vector borne disease risk (for a review see[47]). Light may modify disease transmission by attracting or repelling vectors, and modulating their biting rates, the long-term survival of vectors, and potentially the success of parasite establishment.[47,48] Some diurnal vector species increase their biting rates when exposed to artificial light (e.g. *Aedes aegypti*[49]), while in other nocturnal species, biting rates are suppressed under artificial light (e.g. *Anopheles gambiae*[50]). This range of responses of mosquitoes to ALAN means the potential interactions between artificial light and disease transmission could usefully be further investigated.[30]

Environmental impacts

Only a handful of studies that we are aware of have directly tested for biological impacts of ALAN in Africa:

1. Lights from sport stadiums were found to benefit urban exploiter bats and allow them to increase their temporal foraging niche.[51]
2. Moth consumption by Cape serotine bats (*Neoromicia capensis*) was found to increase sixfold in experimentally lit conditions, suggesting that specialist moth-eating bats and moths may face increased resource competition and predation risk, respectively.[52]
3. Syntonic bats have also been found to increase their foraging frequency under experimentally lit conditions.[53]
4. Flashing LED light was found to reduce nocturnal livestock predation in Kenya.[54]
5. Manipulating artificial lighting (on and off) in aquatic systems showed that ALAN influenced the behaviour and composition of fish communities in estuarine waters.[55] Artificial light favoured piscivores, presumably through the concentration of prey and improved foraging conditions for predatory fish that use visual cues to hunt.[55]
6. The attraction of fish to light is likely widely true in both the salt- and freshwater bodies of Africa, as illustrated by local fishers on Lake Tanganyika, Tanzania, who use lanterns to improve catch rates by attracting fish to the surface and their nets.[56]

7. Both African sharptooth catfish (*Clarias gariepinus*) and Mozambique tilapia (*Oreochromis mossambicus)* show increased growth rates under artificial light treatments, as the light attracts insects.[57]

8. By applying an innovative transplanting methodology, Foster et al.[58] demonstrated that natural celestial cues are obscured by artificial lighting, with dramatic changes in dung beetle (*Scarabaeus satyrus*) orientation behaviour. For this species, artificial light thus increases individual competition and reduces dispersal efficiency.

Although useful contributions to the global body of evidence, these studies provide limited insight into the diversity and magnitude of biological impacts of ALAN across the African continent.

The potential for negative impacts of ALAN on African species and ecological systems is substantial. Large areas of Africa still possess darker night skies than the global average.[14] Therefore biodiversity in those areas may be more 'light naïve', in that it has not encountered ALAN to the extent and at the intensities to which nature in the Global North has been subjected. Whether 'light naïve' species do indeed suffer greater impact with the addition of ALAN, and what the adaptive capacity is of species to ALAN, remain globally pertinent research questions, and Africa may provide a fruitful testing ground for disentangling such impacts. Because some areas of the continent are intensely artificially lit, opportunities also exist for 'transplant' experiments (see for instance[58]), where taxa may be moved from areas that experience high ALAN to dark skies, and vice versa, to test the influence of altered light regimes or light naïveté on biological responses while accounting for other environmental gradients. In addition, there are opportunities to assess responses to gradients of skyglow, because in many regions dark areas abut intensely lit regions.

We highlight a diverse set of environmental research areas for which Africa might be an ideal living laboratory to advance understanding of impacts of ALAN:

1. Africa arguably hosts the last intact guild of apex predators, most of which are nocturnal[21,23], providing important opportunities to understand the impacts of ALAN upon them. Lions (*Panthera leo*) have been shown to have a higher hunting success during moonless nights, particularly so when attacking humans[59], and given increases in habitat destruction compounded by increased human population increases and expansion, human–lion encounter rates may increase. ALAN, especially through sky glow, may interfere with the moonlight cycles to which African ecosystems have evolved. It is unclear whether light may reduce foraging success, or act as an attractant for activity. LED lights have been used in pilot trials to reduce lion attacks on livestock.[54,60] Other large carnivores like cheetah (*Acinonyx jubatus*) and wild dog (*Lycaon pictus*) are often described as being diurnal to avoid direct competition with more dominant lion. However, evidence suggests that subdominant animals are in fact more nocturnally active than previously thought, particularly so on more moonlit nights.[23]

2. Road networks are expanding on the African continent, and there is a growing realisation that ALAN from cars may act as an overlooked pollutant.[22] Two studies in Tanzania reported that 79% and 63% of all recorded roadkills (excluding birds) were of nocturnal animals[61,62], whilst a study in South Africa reported that 100% and 61% of amphibian and mammal roadkills, respectively, were from nocturnal species[63]. This may be indicative that vehicular movement and 'blinding' by headlights may pose a particular risk for Africa's nocturnal species. This furthermore raises the question of how vehicle headlights (or spotlights on game viewing night drives, a common 'eco-friendly' practice in African protected areas) may impact on the 'fight and flight' responses of nocturnal species, which typically have eyes designed to work optimally under poorly lit conditions and whose vision may be compromised for an unknown period by a sudden exposure to unnatural levels of light.

3. Impacts of ALAN on other major groups of organisms in Africa such as birds, insects and plants are virtually unknown, but may be expected as is the case elsewhere[11,64], and crucially also may alter the ecosystem services they underpin, like pollination[65].

4. The extent to which ALAN also enables environmental crime is unknown and may be significant in Africa. For instance, intertidal poaching for abalone is prevalent across much of the western coast of southern Africa, and skyglow from cities, as well as light emissions from coastal towns, may help orientate poaching activity at night and enable navigation. Similarly, skyglow and point source lights from outside of major protected areas may enable the orientation and navigation for poachers pursuing large game, such as rhinos. In the Kruger National Park, the moon phase plays a role in rhino poaching incidents, with poachers preferring nights with better light conditions.[24]

5. It may be particularly desirable to expand experiments at low latitudes to test the biological impacts of ALAN. This is because of the geographical research bias towards northern latitudes, in regions also already experiencing significant ALAN, and so how such impacts scale to other regions in the globe remains an open question.

6. As is the case elsewhere, protected areas in Africa show increases in ALAN[25], and globally in some protected areas, many species, including mammals, now experience persistent ALAN from skyglow comparable to the light of a full moon.[18] How ALAN, especially where it originates at high levels from urban centres that abut onto protected areas, affects species and ecosystems in regions that are thought to have dark skies, remains a pressing question.

Tourism industry

Tourism makes fundamental contributions to the economies of many African nations. Because ALAN erodes the ability to see celestial objects, its negative impact on imaging astronomy is well appreciated, but less so is its impacts on the bourgeoning astrotourism industry. Astrotourism seeks regions with pristine dark skies in remote locations so that amateur astronomers can view celestial phenomena, especially 'deep sky objects' like nebulae and galaxies, which cannot be detected in more artificially lit areas. It is a rapidly growing tourism sector, and given global declines in the quality of night skies for viewing celestial objects[66], regions which retain dark skies may form focal areas for such activity in future.

A proposed astrotourism route in the southern Northern Cape near Sutherland, South Africa, may increase revenue streams from tourism into a region with few other options for economic development.[67] Indeed, large tracts of Africa are in ideal locations (remote and with dark skies) to promote and develop astrotourism as part of its touristic 'portfolio'. The burgeoning lodge industry, often in wilderness locales, must also focus on maintaining dark skies during its expansion. However, Africa currently only possesses two recognised 'International Dark Sky Reserves' – regions possessing an exceptional or distinguished quality of starry nights and nocturnal environment that is specifically protected for its scientific, natural, educational, cultural, heritage and/or public enjoyment of its unpolluted night skies. Following an intensive review process, to ensure that the dark sky values remain preserved, these areas are designated by the 'International Dark Sky Association' (a global non-governmental organisation mandated to help stop ALAN and protect the night skies).

Culture and capacity

Research into the impacts of ALAN in Africa could usefully be advanced by greater cross-disciplinary work and fostering of appropriate research capacity. The impacts and consequences of ALAN do not yet enter the political and societal discourse in Africa to the extent of other change drivers, such as climate change or the illegal wildlife trade. This is not unique to Africa of course, but certainly the volume of research work across disciplines emerging especially from Asia, North America and Europe, is indicative of a shift in thinking towards the problem. Major funding mechanisms have supported large, long-term projects in the Global North – such as the 'ECOLIGHT' experiment[68], the 'Verlust der Nacht' experiment[69], the 'LightOnNature' experiment[70]. Some government initiatives are opting to change lighting regimes to healthier alternatives and major non-governmental organisations are encouraging the protection of

dark places (e.g. International Dark Sky Association). Industry partners are also striving towards more environmentally friendly lighting alternatives. African research and policy must follow suit in such initiatives and indeed has an opportunity to pre-empt the issues bound to emerge, by being proactive in their planning and implementation, which could reduce both human and environmental impacts (for guidance see[29]).

It is imperative that national lighting regulations are kept well updated and informed by contemporary research. Across the continent, it is unclear to what extent such policies exist, let alone whether they are implemented and frequently revisited. For example, the national lighting regulations in South Africa, set out by the South African Bureau of Standards, were last amended in 2007, which leave them sorely lacking behind the development of new lighting technology and the optimal implementation thereof.

These shortcomings mean that robust human capacity to investigate ALAN must still be developed across the continent, and a greater appreciation is required amongst sectors about ALAN's potentially negative effects, techniques to measure it at various spatial scales, and strategies for mitigation tailored to Africa. Skills across academia, the public sector and industry need to be developed continent-wide and used to inform policymakers and other stakeholders of the importance of mitigating the harmful impacts of ALAN while sustainably optimising its economic development.

Conclusion

Given the realised and potential severity of its human health and environmental impacts, the dearth of studies of ALAN in Africa is a surprising oversight by its research community. If we are to understand the magnitude and the extent of the impacts of ALAN in Africa, we must enhance both the fundamental knowledge of its effects, and develop the technical capacity to do so. Africa has the opportunity to learn lessons from other parts of the world and benefit from technological advances in terms of implementation of ALAN mitigation measures. These mitigation strategies are well known, practical and relatively straightforward to implement. However, how to operationalise such mitigation measures in an African policy context requires greater awareness across societal sectors and the willingness to do so. With currently relatively low levels of ALAN compared to the global situation, many African nations can strategically plan and pro-actively implement mitigation measures without increased risk or increased cost to development. There is real, but diminishing, opportunity to ensure that the expansion of ALAN in Africa does not compromise its human and environmental health.

Acknowledgement

We acknowledge funding from the Jennifer Ward Oppenheimer Research Grant to B.W.T.C.

Competing interests

We have no competing interests to declare.

Authors' contributions

B.W.T.C.: Conceptualisation, methodology, data collection, validation, writing – the initial draft, writing – revisions, project leadership, project management, funding acquisition. I.P.J.S.: Conceptualisation, methodology, writing – revisions, funding acquisition. S.A.: Methodology, data collection, data analysis, validation, writing – revisions. K.J.G.: Conceptualisation, methodology, writing – revisions, student supervision, funding acquisition.

References

1. Gaston KJ, Duffy JP, Gaston S, Bennie J, Davies TW. Human alteration of natural light cycles: Causes and ecological consequences. Oecologia. 2014;176(4):917–931. https://doi.org/10.1007/s00442-014-3088-2
2. Sánchez de Miguel A, Bennie J, Rosenfeld E, Dzurjak S, Gaston KJ. First estimation of global trends in nocturnal power emissions reveals acceleration of light pollution. Remote Sens. 2021;13(16), Art. #3311. https://doi.org/10.3390/rs13163311
3. Gaston KJ, Gaston S, Bennie J, Hopkins J. Benefits and costs of artificial nighttime lighting of the environment. Environ Rev. 2015;23(1):14–23. https://doi.org/10.1139/er-2014-0041
4. Davies TW, Smyth T. Why artificial light at night should be a focus for global change research in the 21st century. Glob Change Biol. 2018;24(3):872–882. https://doi.org/10.1111/gcb.13927
5. Falchi F, Cinzano P, Elvidge CD, Keith DM, Haim A. Limiting the impact of light pollution on human health, environment and stellar visibility. J Environ Manage. 2011;92(10):2714–2722. https://doi.org/10.1016/j.jenvman.2011.06.029
6. Walker WH II, Bumgarner JR, Walton JC, Liu JA, Meléndez-Fernández OH, Nelson RJ, et al. Light pollution and cancer. Int J Mol Sci. 2020;21(24), Art. #9360. https://doi.org/10.3390/ijms21249360
7. Cho YM, Ryu SH, Lee BR, Kim KH, Lee E, Choi J. Effects of artificial light at night on human health: A literature review of observational and experimental studies applied to exposure assessment. Chronobiol Int. 2015;32(9):1294–1310. https://doi.org/10.3109/07420528.2015.1073158
8. Sanders D, Frago E, Kehoe R, Patterson C, Gaston KJ. A meta-analysis of biological impacts of artificial light at night. Nat Ecol Evol. 2021;5:74–81. https://doi.org/10.1146/annurev-ecolsys-110316-022745
9. Davies TW, Duffy JP, Bennie J, Gaston KJ. The nature, extent, and ecological implications of marine light pollution. Front Ecol Environ. 2014;12(6):347–355. https://doi.org/10.1890/130281
10. Gaston KJ, Bennie J, Davies TW, Hopkins J. The ecological impacts of nighttime light pollution: A mechanistic appraisal. Biol Rev. 2013;88(4):912–927. https://doi.org/10.1111/brv.12036
11. Bennie J, Davies TW, Cruse D, Gaston KJ. Ecological effects of artificial light at night on wild plants. J Ecol. 2016;104(3):611–620. https://doi.org/10.1111/1365-2745.12551
12. Duffy JP, Bennie J, Durán AP, Gaston KJ. Mammalian ranges are experiencing erosion of natural darkness. Sci Rep. 2015;5, Art. #12042. https://doi.org/10.1038/srep12042
13. Hölker F, Bolliger J, Davies TW, Giavi S, Jechow A, Kalinkat G, et al. 11 Pressing research questions on how light pollution affects biodiversity. Front Ecol Evol. 2021;9, Art. #767177. https://doi.org/10.3389/fevo.2021.767177
14. Falchi F, Cinzano P, Duriscoe D, Kyba CCM, Elvidge CD, Baugh K, et al. The new world atlas of artificial night sky brightness. Sci Adv. 2016;2(6), e1600377. https://doi.org/10.1126/sciadv.1600377
15. Kyba CCM, Tong KP, Bennie J, Birriel I, Birriel JJ, Cool A, et al. Worldwide variations in artificial skyglow. Sci Rep. 2015;5, Art. #8409. https://doi.org/10.1038/srep08409
16. Cox DTC, Sánchez de Miguel A, Dzurjak SA, Bennie J, Gaston KJ. National scale spatial variation in artificial light at night. Remote Sens. 2020;12(10), Art. #1591. https://doi.org/10.3390/rs12101591
17. Cavazzani S, Ortolani S, Bertolo A, Binotto R, Fiorentin P, Carraro G, et al. Satellite measurements of artificial light at night: Aerosol effects. Mon Not R Astron Soc. 2020;499(4):5075–5089. https://doi.org/10.1093/mnras/staa3157
18. Bennie J, Davies TW, Inger R, Gaston KJ. Mapping artificial lightscapes for ecological studies. Methods Ecol Evol. 2014;5(6):534–540. https://doi.org/10.1111/2041-210X.12182
19. Nilsson DE, Smolka J. Quantifying biologically essential aspects of environmental light. J R Soc Interface. 2021;18(177), Art. #20210184. https://doi.org/10.1098/rsif.2021.0184
20. Kumar P, Rehman S, Sajjad H, Tripathy BR, Rani M, Singh S. Analyzing trend in artificial light pollution pattern in India using NTL sensor's data. Urban Clim. 2019;27:272–283. https://doi.org/10.1016/j.uclim.2018.12.005
21. Gaston KJ. Nighttime ecology: The "Nocturnal Problem" revisited. Am Nat. 2019;193(4):481–502. https://doi.org/10.1086/702250
22. Gaston KJ, Holt LA. Nature, extent and ecological implications of night-time light from road vehicles. J Appl Ecol. 2018;55(5):2296–2307.
23. Cozzi G, Broekhuis F, McNutt JW, Turnbull LA, Macdonald DW, Schmid B. Fear of the dark or dinner by moonlight? Reduced temporal partitioning among Africa's large carnivores. Ecology. 2012;93(12):2590–2599. https://doi.org/10.2307/41739617

24. Koen H, De Villiers JP, Roodt H, De Waal A. An expert-driven causal model of the rhino poaching problem. Ecol Model. 2017;347:29–39. https://doi.org/10.1016/j.ecolmodel.2016.12.007

25. Gaston KJ, Duffy JP, Bennie J. Quantifying the erosion of natural darkness in the global protected area system. Conserv Biol. 2015;29(4):1132–1141. https://doi.org/10.1111/cobi.12462

26. Sanders D, Kehoe R, Cruse D, Van Veen FJF, Gaston KJ. Low levels of artificial light at night strengthen top-down control in insect food webs. Curr Biol. 2018;28(15):2474–2478.e3. https://doi.org/10.1016/j.cub.2018.05.078

27. Tidau S, Smyth T, McKee D, et al. Marine artificial light at night: An empirical and technical guide. Methods Ecol Evol. 2021;12(9):1588–1601. https://doi.org/10.1111/2041-210X.13653

28. Green RF, Luginbuhl CB, Wainscoat RJ, Duriscoe D. The growing threat of light pollution to ground-based observatories. Astron Astrophys Rev. 2022;30(1), Art. #1. https://doi.org/10.1007/s00159-021-00138-3

29. Bará S, Falchi F, Lima RC, Pawley M. Keeping light pollution at bay: A red-lines, target values, top-down approach. Environ Chall. 2021;5, Art. #100212. https://doi.org/10.1016/j.envc.2021.100212

30. Coetzee BWT, Gaston KJ, Koekemoer LL, Kruger T, Riddin MA, Smit IPJ. Artificial light as a modulator of mosquito-borne disease risk. Front Ecol Evol. 2022;9, Art. #768090. https://doi.org/10.3389/fevo.2021.768090

31. Dominoni DM, Borniger JC, Nelson RJ. Light at night, clocks and health: From humans to wild organisms. Biol Lett. 2016;12(2), Art. #20160015. https://doi.org/10.1098/rsbl.2016.0015

32. Sulemana I, Nketiah-Amponsah E, Codjoe EA, Andoh JAN. Urbanization and income inequality in sub-Saharan Africa. Sustain Cities Soc. 2019;48(April), Art. #101544. https://doi.org/10.1016/j.scs.2019.101544

33. Ulsrud K, Winther T, Palit D, Rohracher H. Accelerating access to electricity services through a socio-technical design in Kenya. Energy Res Soc Sci. 2015;5:34–44. https://doi.org/10.1016/j.erss.2014.12.009

34. Pillot B, Muselli M, Poggi P, Dias JB. On the impact of the global energy policy framework on the development and sustainability of renewable power systems in sub-Saharan Africa: The case of solar PV [preprint]. arXiv: 1704.01480 [physics.soc-ph]. 2019. https://doi.org/10.48550/arXiv.1704.01480

35. Ginoux P, Prospero JM, Gill TE, Hsu NC, Zhao M. Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. Rev Geophys. 2012;50(3), RG3005. https://doi.org/10.1029/2012RG000388

36. Elvidge CD, Baugh K, Zhizhin M, Hsu FC, Ghosh T. VIIRS night-time lights. Int J Remote Sens. 2017;38(21):5860–5879. https://doi.org/10.1080/01431161.2017.1342050

37. Blimpo MP, Cosgrove-Davies M. Electricity access in sub-Saharan Africa: Uptake, reliability, and complementary factors for economic impact [document on the Internet]. c2019 [2023 Feb 14]. Available from: https://www.perlego.com/book/1483669/electricity-access-in-subsaharan-africa-uptake-reliability-and-complementary-factors-for-economic-impact-pdf

38. Jechow A, Ribas SJ, Domingo RC, Hölker F, Kolláth Z, Kyba CCM. Tracking the dynamics of skyglow with differential photometry using a digital camera with fisheye lens. J Quant Spectrosc Radiat Transf. 2018;209:212–223. https://doi.org/10.1016/j.jqsrt.2018.01.032

39. Hale JD, Davies G, Fairbrass AJ, Matthews TJ, Rogers CDF, Sadler JP. Mapping lightscapes: Spatial patterning of artificial lighting in an urban landscape. PLoS ONE. 2013;8(5), e61460. https://doi.org/10.1371/journal.pone.0061460

40. Sánchez de Miguel A, Kyba CCM, Aubé M, et al. Colour remote sensing of the impact of artificial light at night (I): The potential of the International Space Station and other DSLR-based platforms. Remote Sens Environ. 2019;224:92–103. https://doi.org/10.1016/j.rse.2019.01.035

41. Levin N, Kyba CCM, Zhang Q, Zamorano J, Cardiel N, Tapia C, et al. Remote sensing of night lights: A review and an outlook for the future. Remote Sens Environ. 2020;237, Art. #111443. https://doi.org/10.1016/j.rse.2019.111443

42. International Energy Agency. World energy outlook 2022. Paris: IEA; 2022. Available from: https://www.iea.org/reports/world-energy-outlook-2022

43. Schulte-Römer N, Meier J, Söding M, Dannemann E. The LED paradox: How light pollution challenges experts to reconsider sustainable lighting. Sustainability. 2019;11(21), Art. #6160. https://doi.org/10.3390/su11216160

44. Mentis D, Howells M, Rogner H, et al. Lighting the world: The first application of an open source, spatial electrification tool (OnSSET) on sub-Saharan Africa. Environ Res Lett. 2017;12(8), Art. #085003. https://doi.org/10.1088/1748-9326/aa7b29

45. Azimoh CL, Klintenberg P, Wallin F, Karlsson B, Mbohwa C. Mini-grid solution for rural electrification in South Africa. Energy Convers Manag. 2016;110:268–277. https://doi.org/10.1016/j.enconman.2015.12.015

46. World Health Organization (WHO). World malaria report 2021. Geneva: WHO; 2021. Available from: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021

47. Wilson R, Wakefield A, Roberts N, Jones G. Artificial light and biting flies: The parallel development of attractive light traps and unattractive domestic lights. Parasit Vectors. 2021;14(1), Art. #28. https://doi.org/10.1186/s13071-020-04530-3

48. Rund SSC, O'Donnell AJ, Gentile JE, Reece SE. Daily rhythms in mosquitoes and their consequences for malaria transmission. Insects. 2016;7(2), Art. #14. https://doi.org/10.3390/insects7020014

49. Rund SSC, Labb LF, Benefiel OM, Duffield GE. Artificial light at night increases *Aedes aegypti* mosquito biting behavior with implications for arboviral disease transmission. Am J Trop Med Hyg. 2020;103(6):2450–2452. https://doi.org/10.4269/ajtmh.20-0885

50. Sheppard AD, Rund SSC, George GF, Clark E, Acri DJ, Duffield GE. Light manipulation of mosquito behaviour: Acute and sustained photic suppression of biting activity in the *Anopheles gambiae* malaria mosquito. Parasit Vectors. 2017;10(1), Art. #255. https://doi.org/10.1186/s13071-017-2196-3

51. Schoeman MC. Light pollution at stadiums favors urban exploiter bats. Anim Conserv. 2016;19(2):120–130. https://doi.org/10.1111/acv.12220

52. Minnaar C, Boyles JG, Minnaar IA, Sole CL, McKechnie AE. Stacking the odds: Light pollution may shift the balance in an ancient predator-prey arms race. J Appl Ecol. 2015;52(2):522–531. https://doi.org/10.1111/1365-2664.12381

53. Bailey LA, Brigham RM, Bohn SJ, Boyles JG, Smit B, Bailey LA. An experimental test of the allotonic frequency hypothesis to isolate the effects of light pollution on bat prey selection. Oecologia. 2019;190(2):367–374. https://doi.org/10.1007/s00442-019-04417-w

54. Lesilau F, Fonck M, Gatta M, Musyoki C, Van 't Zelfde M, Persoon GA, et al. Effectiveness of a LED flashlight technique in reducing livestock depredation by lions (*Panthera leo*) around Nairobi National Park, Kenya. PLoS ONE. 2018;13(1), e0190898. https://doi.org/10.1371/journal.pone.0190898

55. Becker A, Whitfield AK, Cowley PD, Järnegren J, Næsje TF. Potential effects of artificial light associated with anthropogenic infrastructure on the abundance and foraging behaviour of estuary-associated fishes. J Appl Ecol. 2013;50(1):43–50. https://doi.org/10.1111/1365-2664.12024

56. Mgana H, Kraemer BM, O'Reilly CM, et al. Adoption and consequences of new light-fishing technology (LEDs) on Lake Tanganyika, East Africa [preprint]. bioRxiv: 619007. 2019. https://doi.org/10.1101/619007

57. Rapatsa MM, Moyo NAG. The potential role of night-time lighting in attracting terrestrial insects as food for *Oreochromis mossambicus* and *Clarias gariepinus*. Trop Zool. 2017;30(4):156–169. https://doi.org/10.1080/03946975.2017.1362841

58. Foster JJ, Tocco C, Smolka J, Khaldy L, Baird E, Byrne MJ, et al. Light pollution forces a change in dung beetle orientation behavior. Curr Biol. 2021;31:3935–3942. https://doi.org/10.1016/j.cub.2021.06.038

59. Packer C, Swanson A, Ikanda D, Kushnir H. Fear of darkness, the full moon and the nocturnal ecology of African lions. PLoS ONE. 2011;6(7), e22285. https://doi.org/10.1371/journal.pone.0022285

60. Okemwa B, Gichuki N, Virani M, Kanya J, Kinyamario J, Santangeli A. Effectiveness of LED lights on bomas in protecting livestock from predation in southern Kenya. Conserv Evid. 2018;15:39–42.

61. Njovu HK, Kisingo AW, Hesselberg T, Eustace A. The spatial and temporal distribution of mammal roadkills in the Kwakuchinja Wildlife Corridor in Tanzania. Afr J Ecol. 2019;57(3):423–428. https://doi.org/10.1111/aje.12608

62. Kioko J, Kiffner C, Jenkins N, Collinson WJ. Wildlife roadkill patterns on a major highway in Northern Tanzania. Afr Zool. 2015;50(1):17–22. https://doi.org/10.1080/15627020.2015.1021161

63. Collinson WJ, Parker DM, Bernard RTF, Reilly BK, Davies-Mostert HT. An inventory of vertebrate roadkill in the Greater Mapungubwe Transfrontier Conservation Area, South Africa. Afr J Wildl Res. 2015;45(3):301–311. https://doi.org/10.3957/056.045.0301

64. Owens ACS, Cochard P, Durrant J, Farnworth B, Perkin EK, Seymoure B. Light pollution is a driver of insect declines. Biol Conserv. 2020;241, Art. #108259. https://doi.org/10.1016/j.biocon.2019.108259

65. Knop E, Zoller L, Ryser R, Gerpe C, Hörler M, Fontaine C. Artificial light at night as a new threat to pollination. Nature. 2017;548(7666):206–209. https://doi.org/10.1038/nature23288

66. Pásková M, Budinská N, Zelenka J. Astrotourism-exceeding limits of the earth and tourism definitions? Sustainability. 2021;13(1), Art. #373. https://doi.org/10.3390/su13010373

67. Jacobs L, Du Preez EA, Fairer-Wessels F. To wish upon a star: Exploring astrotourism as vehicle for sustainable rural development. Dev South Afr. 2020;37(1):87–104. https://doi.org/10.1080/0376835X.2019.1609908

68. Bennie J, Davies TW, Cruse D, Inger R, Gaston KJ. Cascading effects of artificial light at night: Resource-mediated control of herbivores in a grassland ecosystem. Philos Trans R Soc B Biol Sci. 2015;370(1667), Art. #20140131. https://doi.org/10.1098/rstb.2014.0131

69. Hölker F, Wurzbacher C, Weißenborn C, Monaghan MT, Holzhauer SIJ, Premke K. Microbial diversity and community respiration in freshwater sediments influenced by artificial light at night. Philos Trans R Soc B Biol Sci. 2015;370(1667), Art. #20140130. https://doi.org/10.1098/rstb.2014.0130

70. Spoelstra K, Van Grunsven RHA, Donners M, et al. Experimental illumination of natural habitat—an experimental set-up to assess the direct and indirect ecological consequences of artificial light of different spectral composition. Philos Trans R Soc B Biol Sci. 2015;370(1667), Art. #20140129. https://doi.org/10.1098/rstb.2014.0129

AUTHORS:
Werner P. Strümpher[1]
Clarke H. Scholtz[2]
Thomas Schlüter[3]

AFFILIATIONS:
[1]Ditsong National Museum of Natural History, Pretoria, South Africa
[2]Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa
[3]Department of Geography, Environmental Science and Planning, University of Eswatini, Kwaluseni, Eswatini

CORRESPONDENCE TO:
Werner Strümpher

EMAIL:
strumpher@ditsong.org.za

DATES:
Received: 02 Sep. 2022
Revised: 26 Dec. 2022
Accepted: 03 Jan. 2023
Published: 29 Mar. 2023

HOW TO CITE:
Strümpher WP, Scholtz CH, Schlüter T. An Eocene fossil scarab beetle (Coleoptera: Scarabaeoidea) from Tanzania. S Afr J Sci. 2023;119(3/4), Art. #14681. https://doi.org/10.17159/sajs.2023/14681

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITOR:
Jemma Finch

KEYWORDS:
Africa, Lagerstätte, Mahenge deposit, maar lake, new genus, new species, taxonomy

FUNDING:
None

An Eocene fossil scarab beetle (Coleoptera: Scarabaeoidea) from Tanzania

A fossil scarabaeoid (Coleoptera: Scarabaeoidea) *Mahengea mckayi* new genus, new species – the only well-preserved insect fossil and one of only a few insects found at the site – is described from an otherwise rich Eocene (~ 45.6 mya) maar Lagerstätte at Mahenge in central northern Tanzania. Numerous fossil fishes and plants have been recovered from this site and described. The dearth of insect fossils is surprising considering their richness in other deposits of similar origin and age. We suggest that the rich fish fauna present in the oxygen-rich parts of the water in the former volcanic crater lake may have scavenged most of the terrestrial insects that fell into the water. Although the fossil described here is undoubtedly that of a member of the Scarabaeoidea, the family placement remains unsure.

Significance:
We describe the first Eocene fossil scarab from Africa. It is one of only a few scarab fossils from the continent and one of the best-preserved insect (invertebrate) fossils from the deposit.

Introduction

Unusually rich and well-preserved fossil assemblages are generally known under the term 'Lagerstätte' because the information they provide is not only concerned with details of the extinct animals and plants that once lived there, thus an indication of differing levels of diversity in the past, but because it is characteristic of them that more complete records of communities are preserved. Lagerstätte additionally afford insights into palaeoecological and evolutionary relationships in the geological context.

Classic Eocene deposits in Germany, like the famous Eckfelder Maar, are considered to be Lagerstätte, but these types of fossiliferous deposits have been largely limited to the northern hemisphere. The Orapa kimberlite pipe in north-central Botswana is one of the few maar deposits in the southern hemisphere (located in eastern Botswana), and is famous for its exceptionally well-preserved compression fossils of insects and other organisms in fine-grained mudstone.[1,2] Numerous maar deposits are known from central northern Tanzania (Figure 1), and at least one of them, the Mahenge maar deposit, has yielded various fossils.[3-5]

Beetles in general, and Scarabaeoidea in particular, usually yield only a few useful characters when fossilised. Scarabaeoids are mostly compact and evenly and strongly sclerotised and when compressed and fossilised, usually only the body outline and a few legs are recognisable. In less compact beetles often only elytra become fossilised. As a consequence, well-preserved beetle fossils in sediments are relatively rare.

Fossil beetles can only be reliably identified as belonging to Scarabaeoidea (other than Lucanidae and Passalidae) by at least one of four character suites: antennae with lamellate club; pronotum and fore legs adapted for burrowing (procoxae enlarged, pronotum enlarged to include enlarged coxal musculature, protibiae dilated apically, usually with teeth on their outer margin), albeit that similar adaptations also occur in some other burrowing beetles; reduced wing venation and intrinsic spring mechanism for folding wings; prosternal intercoxal process widened apically behind procoxae.[6]

The fossil record of Scarabaeoidea and a comprehensive molecular study of the origin of major beetle groups hypothesised a Triassic origin for the group[6,7], postulated the establishment of most extant families during the Jurassic and suggested extensive radiation of groups within families during the Cretaceous, especially of phytophagous taxa in co-radiation with the expanding angiosperm flora. Yet, despite the presence of well-placed Scarabaeoidea fossils from as far back as the Middle Jurassic[8], a total of only about 230 species have been described[6,9]; this from one of the largest beetle groups with about 30 000 extant species and probably many thousand more extinct. Furthermore, besides a detailed study of a late Cretaceous (~ 91 mya) Melolonthinae (Scarabaeidae) species[10], only a few Scarabaeidae of Palaeocene[9], seven of Eocene[9,11], another few of Oligocene age[9], and two from the Pliocene[12] are known. Most of the fossil species of all scarabaeoid groups are from the Miocene and younger.[6,9] The Miocene was also the period when grasslands and herbivorous mammals radiated in Africa, which is when and where dung beetles (Scarabaeinae) co-radiated.[13] Three Miocene (13–12 mya) fossil species purportedly belonging to extant dung beetle genera have been described from Kenya.[14]

In spite of several well-studied and extensive Eocene fossil sites (in the northern hemisphere), only a few have produced scarabaeoid fossils, and then in very low numbers: three Aphodiinae (one each from Menat, France; Baltic amber; Bognor Regis, UK) and three Melolonthinae (two species from Geiseltal, Germany and one from the Green River Formation in USA). The one described in this paper is the only Eocene scarabaeoid known from Africa.

Scarabaeoid fossils are rare in African fossil records and have only been recorded from the very rich Cretaceous deposit at Orapa in Botswana (one species: *Ceafornotensis archratiras* Wooley, 2016)[10], a Miocene deposit in Kenya (three species)[14], and a Pliocene site in northern Tanzania (two species)[12]. The specimen described in this paper lies chronologically about midway between the Orapa and Kenyan fossils, the Eocene. Earlier deposits such as the Triassic Molteno Formation yielded many insect fossils, including beetle fossils, but no scarabaeoids.[15,16]

The Orapa fossil belongs to an extinct genus and species (*Ceafornotensis archratiras*) of the scarabaeid subfamily Melolonthinae[10], and the three Kenyan fossils are members of three extant genera of dung beetles (Scarabaeinae); *Anachalcos* Hope, 1837 (the name subsequently synonymised with *Chalconotus* Dejean, 1833), *Copris* Geoffroy, 1762 and *Metacatharsius* Montreuil, 1998[14]. The two Pliocene fossils from Tanzania (Laetoli) are assigned to two different scarabaeid subfamilies – Dynastinae (of the extinct genus *Calcitoryctes* Krell, 2011) and Melolonthinae (one unnamed species of extant melolonthine tribe Schizonychini).[12]

Material and methods

Geographical position

The Mahenge deposit (4°47'50.2"S, 34°15'54.5"E) is located close to the village of Mwaru, about 65 km west of the town Singida and east of the Wembere Steppe (Figure 1).[5]

Geology and palaeoenvironmental reconstruction

Geologists prospecting for diamonds in the area of Singida in the early 1930s recorded the occurrence of superbly preserved fossils contained in lake beds overlying the diamondiferous kimberlite pipes.[2] The fossils were found in sediments largely consisting of shales and mudstones, originally deposited in the former crater lake. One of many in the area, the Mahenge maar deposit has yielded various fossils.[3-5] The Mahenge deposit has been interpreted as a small, previously roughly circular lake, about 400 m in diameter, which was formed in a kimberlite intrusion.[17,18]

Generally, the formation of a kimberlite pipe begins after its eruption, when the overlying basement rock material is scattered around and a crater produced, which is subsequently surrounded by a tuffaceous cone of largely fine-grained ashes. In Mahenge, this steep-sided cone created a rather small shoreline for the original lake, made of the sediments from the surrounding pyroclastic kimberlitic rocks.[19] Thus, the centre of the Mahenge palaeolake is now characterised by well-stratified, micro-laminated shales and mudstones in which the majority of the fossils are found (Figure 2).[3,18,19]

The age of the fossiliferous beds at Mahenge has been determined using radiometric methods. Originally, based on a comparison of elements in the fossil ichthyofauna, a Miocene age was suggested[20], but this changed later to Oligocene[21]. Kimberlites from elsewhere in the Singida region (Figure 1) have been dated using U-Pb and fission track radiometric methods, giving an age between 51 and 54 Ma for some kimberlites (Paleocene to Lower Eocene). A single zircon crystal, collected in 1996 from the stream bed at Mahenge and linked by its origin to the Mahenge diatreme, was dated at 45.83±0.17 Ma, based on 206Pb/238U radiometric measurements (Middle Eocene = Lutetian).[3] Consequently, it is assumed to be only slightly older than the palaeolake at Mahenge, because it is well known from comparable maar lakes in Europe and Africa that lacustrine sediments accumulate soon after formation of the original lake.[22] The age of the Mahenge Lagerstätte can therefore be assumed to be about 45.6 Ma old. Estimates of the sedimentation rate at Mahenge indicate that the fossiliferous deposits in the crater represent an interval ranging from 8000 to 22 700 years.[3]

From palaeobotanical evidence, it has been suggested that the palaeoflora of Mahenge structurally resembled 'miombo' savanna woodlands still widespread in much of southern and East Africa, including Tanzania, and that the region was subjected to an overall dry climate with pronounced seasonality.[5,23]

All the fossils are embedded as compressions in fine-laminated, pale beige, clayey-calcareous lake sediments. The rocky material seems to be extremely hard. The surfaces of the fossils are characterised at least partly by a darker film of possibly organic (coaly) or iron-containing material, varying in colour from reddish, brownish to blackish. Fine details to a size of less than 1 mm can often clearly be seen under a microscope, but processes of increasing weathering are also indicated by subsequent exuding of hardened small bubbles of iron or manganese along the body surfaces. In total, 1900 fossils have been recovered; fish make up the largest part (51%), followed by plant remains (36%) while only nine insect fossils have been identified, most of which are poorly preserved.[5]

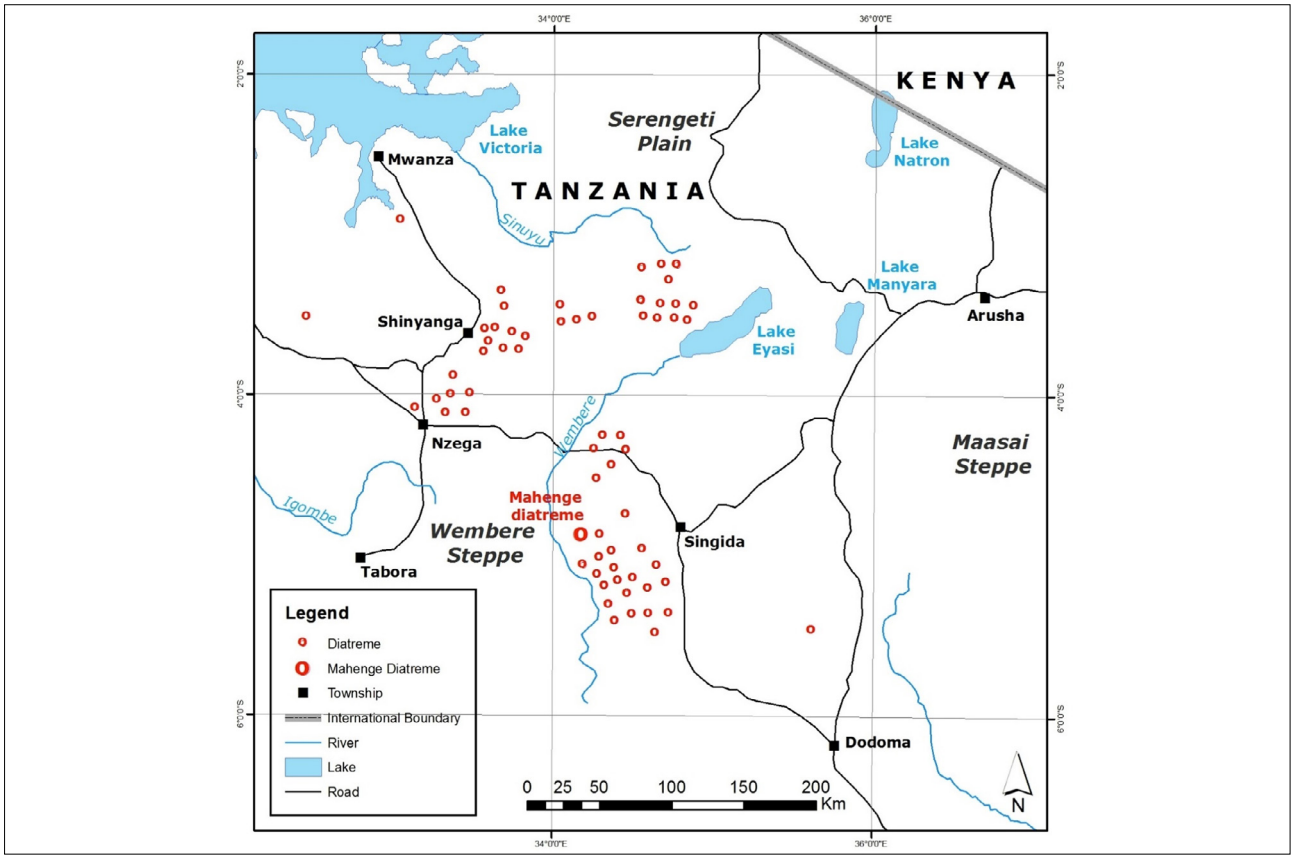


Figure 1: Overview of the Mahenge area in northern Tanzania. Potentially similar kimberlite deposits are indicated as red circles (modified after Kaiser et al.[4]).

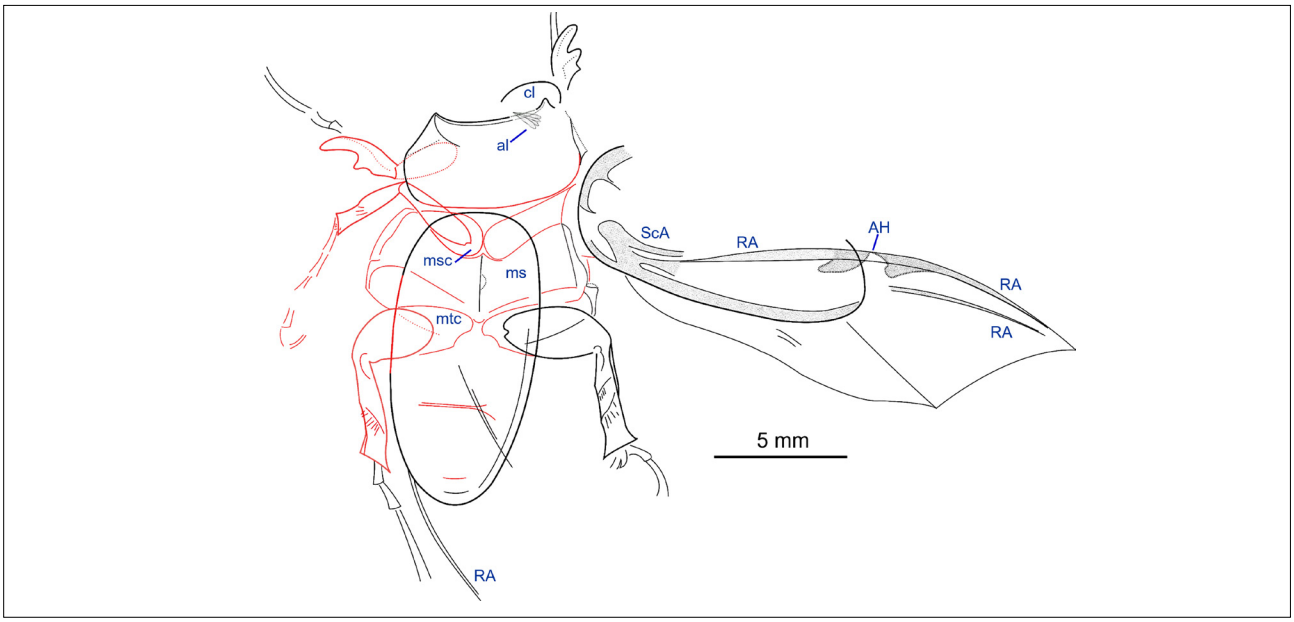


Figure 2: Composite line drawings of *Mahengea mckayi* Strümpher, Scholtz & Schlüter, new genus, new species. Black lines, plate. Red lines, counter plate. Abbreviations: AH, apical hinge; al, antenna lamellae; cl, clypeus; ms, metasternum; msc, mesocoxa; mtc, metacoxa; hindwing venation: RA, radius anterior veins; ScA, subcosta anterior vein.

Examined material

The fossil specimen can be examined and interpreted from both plate and counter plate. On both pale-coloured plate sides it is shown as a dorso-ventral compression. The specimen was examined dry and wet (submerged in water) using a LEICA MZ12.5 dissecting microscope; measurements were made using digital electronic callipers and are here given in millimetres. The plates were photographed with a Canon EOS 5D camera equipped with a 100 mm macro lens. The specimen was illustrated from the photographs with the aid of vector graphics software.[24] The adult specimen is nearly complete, but parts of the head and legs are poorly preserved (Figures 2–4). The body has a cylindrical form. The fore legs are well-visible, distally broadened, and thus apparently suitable for digging. One partial outstretched metathoracic wing is visible (Figures 2, 3).

Taxonomy

Order: **Coleoptera** Linnaeus, 1758[25]

Superfamily: **Scarabaeoidea** Latreille, 1802[26]

Family: ***incertae sedis*** **[uncertain]**

Mahengea Strümpher, Scholtz & Schlüter, **new genus**

(Figures 2–4)

Type species: *Mahengea mckayi*, here designated.

Generic diagnosis

This genus can be diagnosed as follows: *Body* moderately large (~18 mm long), somewhat robust, cylindrical form, broadly oval. *Head*. Clypeal margin rounded. Antenna with lamellate club. *Pronotum*. About twice as wide as long, widest at middle. Shape subquadrate, anterior margin of pronotum concave, basal margin relatively straight, only slightly rounded; lateral margin strongly curved, with basal half rounded, anterior half straight. Anterior angles acute, posterior angles broadly rounded. *Elytra*. Shape parabolic, about twice as long as wide medially, indistinct striae. *Metathoracic wings*. Fully developed, apical hinge discernible. *Legs*. Protibiae widening apically, tridentate, mesotibiae and metatibiae slender, each with medial transverse carinae. Mesocoxa near contiguous at midline. Metafemora large, broadly ovate.

Etymology

The generic name *Mahengea* is based on the name of the palaeolake, Mahenge, in Tanzania from which the fossil originates.

Mahengea mckayi Strümpher, Scholtz & Schlüter, **new species**

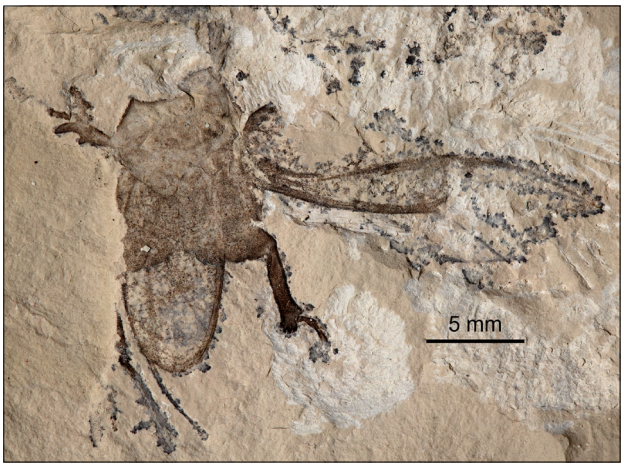


Figure 3: Habitus of *Mahengea mckayi* Strümpher, Scholtz & Schlüter, new genus, new species, holotype specimen (plate).

Holotype

Sex unknown, plate and counter plate. The specimen is deposited in the collection of the Ditsong National Museum of Natural History (TMSA; Coll. No. 811) (Figures 3 and 4).

Description

Broadly oval, elongate, cylindrical form; body length: about 18 mm, body width: 11 mm. *Head*. Clypeus partly visible, anterior margin rounded. Antennal club lamellate with at least three lamellae discernible, lamellae elongate and narrow (Figure 2). *Pronotum*. length 3.8 mm, width about 7.5 mm. Shape and form as in genus. Scutellum not preserved. *Elytra and metathoracic wing*. Elytron: length 12.5 mm, width 6.2 mm. Shape parabolic, about twice as long as wide medially, widest at middle, tapering slightly posteriorly, rounded apically, without discernible striae or tubercles

on surface. Metathoracic wings: Fully developed. *Legs*. Protibiae with three teeth on the outer margin. Mesocoxae close, round; mesofemora and mesotibia slender, latter not strongly dilated at the apex, not wider than mesofemur; mesotarsi distinctly longer than mesotibia. Metacoxae transverse, narrowly separated; metafemora broad, ovate, widest medially; metatibiae slightly dilated apically. Transverse carinae present on the mesotibia and metatibiae, with presence of spine-like setae.

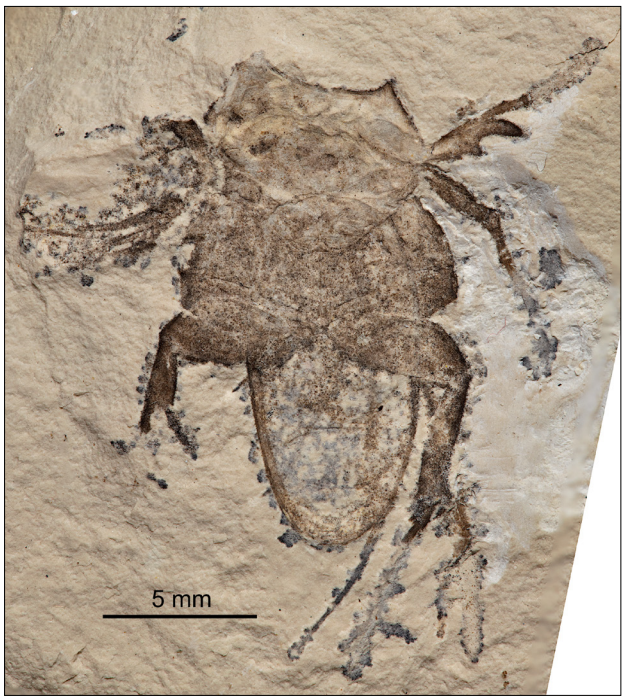


Figure 4: Habitus of *Mahengea mckayi* Strümpher, Scholtz & Schlüter, new genus, new species, holotype specimen (counter plate).

Diagnosis

Same as the genus.

Etymology

This species is named in honour of Ian Mckay (1963–2022) who died suddenly and unexpectedly recently – he was an insect fossil curator, palaeoentomologist and geoscience outreach educationist extraordinaire at the University of the Witwatersrand, Johannesburg, South Africa.

Age and significance of the Mahenge fossil

Based on the geological and palaeontological evidence now known from the Mahenge site, it can be concluded that the maar and related deposits in this area represent one of the very few sampled temporal and geographical windows of freshwater and terrestrial ecosystems in sub-Saharan Africa. Although an Eocene age is not of particular significance because of the undisputed older origin of scarabaeoids, its significance lies in the fact that few scarab fossils of older than Oligocene age have been described, and only one (Cretaceous) from Africa. It is also the only Eocene fossil scarab from Africa and one of only a few worldwide. Furthermore, its description brings to seven the total number of scarabaeoid fossils from the continent.

Possible family placement

All currently recognised families of the Scarabaeoidea (and their subfamilies) are undoubtedly older than the Eocene and the distinguishing characters outlined above could equally apply to most of these taxa. The environment was postulated to be 'savanna', so conditions conducive to habitation by almost all of the known higher taxa would have been possible and none can be excluded on ecological grounds. Consequently, we are left with little option other than to conclude that, although the fossil undoubtedly is that of a scarabaeoid species, the family should be considered *incertae sedis*.

Acknowledgements

We are grateful to Ruth Muller (Pretoria, South Africa) for photographing the fossil specimen for this publication. We thank the anonymous reviewers for their valuable comments that have improved this article.

Competing interests

We have no competing interests to declare.

Authors' contributions

T.S. was responsible for the conception of this study, provided the fossil scarab described herein and reviewed drafts of the paper. C.H.S. and W.P.S. wrote and prepared the various drafts leading up to the submission of this article. W.P.S examined and described the new scarab fossil, and prepared figures. All authors read and approved the final manuscript.

References

1. McKay IJ, Rayner RJ. Cretaceous fossil insects from Orapa, Botswana. J Entomol Soc South Afr. 1986;49:7–17.
2. Rayner RJ, Waters SB, McKay IJ, Dobbs PN, Shaw AL. The mid-Cretaceous palaeoenvironment of central southern Africa (Orapa, Botswana). Palaeogeogr Palaeoclimatol Palaeoecol. 1991;88(1–2):147–156. https://doi.org/10.1016/0031-0182(91)90020-R
3. Harrison T, Msuya CP, Murray A, Jacobs BF, Baez AM, Ludwig KR, et al. Paleontological investigations at the Eocene locality of Mahenge in north-central Tanzania, East Africa. In: Gunnell G, editor. Eocene biodiversity: Unusual occurrences and rarely sampled habitats. New York: Kluwer/Plenum; 2001. p. 39–74. https://doi.org/10.1007/978-1-4615-1271-4_2
4. Kaiser TM, Alberti G, Bullwinkel V, Michalik P, Msuya C, Schulz E. Mahenge – Ein Fenster zum Mitteleozän Afrikas [Mahenge - A window to the Middle Eocene of Africa]. Nat.wiss Rundsch. 2003;56:540–546. German.
5. Kaiser TM, Ansorge J, Arratia G, Bullwinkel V, Gunnell GF, Herendeen PS, et al. The maar lake of Mahenge (Tanzania) – unique evidence of Eocene terrestrial environments in sub-Saharan Africa. Z Dtsch Ges Geowiss. 2006;157(3):411–431. https://doi.org/10.1127/1860-1804/2006/0157-0411
6. Krell F-T. The fossil record of Mesozoic and Tertiary Scarabaeoidea (Coleoptera: Polyphaga). Invertebr Taxon. 2000;14:871–905. https://doi.org/10.1071/IT00031
7. Zhang SQ, Che LH, Li Y, Liang D, Pang H, Ślipiński A, et al. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. Nat Commun. 2018;9, Art. #205. https://doi.org/10.1038/s41467-017-02644-4
8. Bai M, Ahrens D, Yang X-K, Ren D. New fossil evidence of the early diversification of scarabs: *Alloioscarabaeus cheni* (Coleoptera: Scarabaeoidea) from the Middle Jurassic of Inner Mongolia, China. Insect Sci. 2012;19(2):159–171. https://doi.org/10.1111/j.1744-7917.2011.01460.x
9. Krell F-T. Catalogue of fossil Scarabaeoidea (Coleoptera: Polyphaga) of the Mesozoic and Tertiary – Version 2007. Denver Mus Nat Sci Rep. 2007–8:1–79.
10. Woolley C. The first scarabaeid beetle (Coleoptera, Scarabaeidae, Melolonthinae) described from the Mesozoic (Late-Cretaceous) of Africa. Afr Invertebr. 2016;57:53–66. https://doi.org/10.3897/AfrInvertebr.57.8416
11. Ratcliffe B, Smith D, Erwin D. *Oryctoantiquus borealis*, new genus and species from the Eocene of Oregon, U.S.A., the world's oldest fossil dynastine and largest fossil scarabaeid (Coleoptera: Scarabaeidae: Dynastinae). Coleopt Bull. 2005;59:127–135. https://doi.org/10.1649/0010-065X(2005)059[0127:OBNGAS]2.0.CO;2
12. Krell F-T, Schawaller W. Beetles (Insecta: Coleoptera). In: Harrison T, editor. Paleontology and geology of Laetoli: Human evolution in context. Fossil hominins and the associated fauna. Vol. 2. Dordrecht: Springer; 2011. p. 535–548. https://doi.org/10.1007/978-90-481-9962-4_19
13. Mlambo S, Sole CL, Scholtz CH. A molecular phylogeny of the African Scarabaeinae (Coleoptera: Scarabaeidae). Arthropod Syst Phylogeny. 2015;73(2):303–321.
14. Paulian R. Three fossil dung beetles (Coleoptera: Scarabaeidae) from the Kenya Miocene. J East Afr Nat Hist Soc Natl Mus. 1976;31(158):1–4.
15. Anderson JM, Kohring R, Schlüter T. Was insect biodiversity in the Triassic akin to today? – A case study from the Molteno Formation (South Africa). Entomol Gener. 1998;23(1/2):15–26. https://doi.org/10.1127/entom.gen/23/1998/15

16. Anderson HM, Anderson JM. Molteno sphenophytes: Late Triassic biodiversity in southern Africa. Palaeontol Afr. 2018;53(SI):1–391.

17. Mannard GW. The geology of the Mckayi Kimberlite Pipes, Tanganyika [PhD thesis]. Montreal: McGill University; 1962.

18. Murray AM. Eocene cichlid fishes from Tanzania, East Africa. J Vertebr Paleontol. 2000;20(4):651–664. https://doi.org/10.1671/0272-4634(2000)020[0651:ECFFTE]2.0.CO;2

19. Schlüter T. Eocene insects from a Maar Lagerstätte at Mahenge, northern Tanzania. Entomol Gener. 2018;37(3/4):375–392. https://doi.org/10.1127/entomologia/2017/0653

20. Greenwood PH. Fossil denticipitid fishes from East Africa. Bull Br Mus Nat Hist Geol. 1960;5:1–11.

21. Greenwood PH, Patterson C. A fossil osteoglossoid fish from Tanzania (E. Africa). Zool J Linnean Soc. 1967;47(311):211–223. https://doi.org/10.1111/j.1096-3642.1967.tb01404.x

22. Rayner RJ, McKay IJ. The treasure chest of the Orapa diamond mine. Botsw Notes Rec. 1986;18:55–61.

23. Jacobs BF, Herendeen PS. Eocene dry climate and woodland vegetation in tropical Africa reconstructed from fossil leaves from northern Tanzania. Palaeogeogr Palaeoclimatol Palaeoecol. 2004;213:115–123. https://doi.org/10.1016/j.palaeo.2004.07.007

24. InkScape version 1.1.1. Boston, MA: Free Software Foundation, Inc. Available from: https://inkscape.org/

25. Linnaeus C, Salvius L. Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis [System of nature through the three kingdoms of nature, according to classes, orders, genera, species, with characters, differences, synonyms, places]. 10th ed. Vol. 1. Stockholm: Laurentii Salvii; 1758. Latin. https://doi.org/10.5962/bhl.title.542

26. Latreille PA. Histoire naturelle, générale et particuliére des Crustacés et des Insectes. Ouvrage faisant suite aux oeuvres de Leclerc de Buffon, et partie du cours complet d'Histoire naturelle rédigé par C. S. Sonnini, membre de plusieurs sociétés savantes. Familles naturelles des genres [Natural, general and particular history of crustaceans and insects. Book following the works of Leclerc de Buffon, and part of the complete Natural History course written by C. S. Sonnini, member of several learned societies. Natural families of genera]. Vol. 3. Paris: F. Dufart; 1802. French. https://doi.org/10.5962/bhl.title.15764

AUTHORS:
Susan Taljaard[1,2]
Lara van Niekerk[1,2]
Janine B. Adams[2,3]
Taryn Riddin[2,3]

AFFILIATIONS:
[1]Coastal Systems and Earth Observations Research Group, Council for Scientific and Industrial Research (CSIR), Stellenbosch, South Africa
[2]Institute for Coastal and Marine Research, Nelson Mandela University, Gqeberha, South Africa
[3]DSI/NRF Research Chair in Shallow Water Ecosystems, Nelson Mandela University, Gqeberha, South Africa

CORRESPONDENCE TO:
Susan Taljaard

EMAIL:
staljaar@csir.co.za

DATES:
Received: 06 July 2022
Revised: 14 Dec. 2022
Accepted: 05 Jan. 2023
Published: 29 Mar. 2023

HOW TO CITE:
Taljaard S, Van Niekerk L, Adams JB, Riddin T. Advancing ecosystem accounting in estuaries: Swartkops Estuary case study. S Afr J Sci. 2023;119(3/4), Art. #14303. https://doi.org/10.17159/sajs.2023/14303

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Sydney Moyo

KEYWORDS:
estuary-level, extent and condition accounts, pressure accounts, physical ecosystem service accounts, management

FUNDING:
South African National Research Foundation, South African Department of Science and Innovation

Advancing ecosystem accounting in estuaries: Swartkops Estuary case study

Rapid degradation of ecosystems and loss of ecosystem services have sparked interest in developing approaches to report and integrate such change with socio-economic information systems, such as the System of National Accounts. Here we describe an approach and application of ecosystem accounting for individual estuaries, building on approaches previously applied at national and bay levels. Using the Swartkops Estuary as a case study, the focus is on physical accounts for ecosystem extent and condition, as well as accounts for two important ecosystem services (carbon sequestration and recreational use). Pressure accounts are also introduced to demonstrate the value of identifying key areas for management and restoration interventions in response to changes in extent and/or condition accounts. Greater resolution in these account reports, achieved through zoning, provides spatially explicit information on ecosystem assets and their services within an estuary to also inform management decision-making at local level. Further, these accounts can also inform local restoration prioritisation, in support of the UN Decade on Ecosystem Restoration (2021–2030), for example offsetting irreversibly degraded areas in one zone with restoration or maintenance of similar habitats in another.

Significance:

- This study is the first to apply the ecosystem accounting approach at the individual estuary level.
- We provide spatially explicit information on ecosystem assets and their services in support of resource management.
- Physical accounts include extent and condition, as well as ecosystem service and pressure accounts.
- These accounts inform estuary management and restoration at the local governance level.

Introduction

Rapid degradation of ecosystems has sparked interest in developing information systems able to evaluate and report change, as well as the loss of valuable services they provide to people.[1-3] From this stemmed the concept of ecosystem accounting (also referred to as natural capital accounting), a systematic approach by which to trace the contribution of natural capital to the productive system over time and space in an accounting structure.[2,4] Hein et al.[2] reflected on the diverse applications of these type of accounts, including the provision of a comprehensive overview of ecosystem assets and their use, measurement of sustainable use of ecosystems within the context of the sustainable development goals (SDGs), and provision of spatially explicit information in support of resource management.

There is no single, universal method for ecosystem or natural capital accounting, although at its core is the measurement of the extent, condition and value of natural capital assets, as well as the services and benefits derived from them.[5,6] In turn, this information, provides the basis for the evaluation of environmental management and development options against pre-defined objectives.[6] Internationally, emerging global ecosystem account approaches include the Global Ocean Accounts Partnership's framework for ocean accounting and the UN's System for Environmental-Economic Accounting (SEEA).[7-9] The primary aim of the SEEA is to gather and organise environmental information consistently and enable its integration with socio-economic information, such as the System of National Accounts.[10] The SEEA Ecosystem Accounting (SEEA-EA) poses five ecosystem accounts, including three physical accounts and two monetary accounts.[11] The physical accounts comprise ecosystem extent, ecosystem condition, and ecosystem services supply and use, while the monetary accounts include ecosystem services supply and use (expressed in monetary terms) as well as ecosystem monetary asset accounts. How ecosystem accounts are presented mimics the internationally accepted accounting concepts of gathering and organising information consistent with the socio-economic components of the System of National Accounts to ensure comparability. Wang et al.[12] posed designs for marine ecosystem accounts, aligned with the SEEA-EA approach, while others demonstrated the application in marine and coastal ecosystems at national level and at bay level.[13-15]

Most ecosystem and natural capital accounting approaches have been designed for application at national level (e.g. SEEA-EA).[4,11] However, application at the local scale is more challenging.[2] For example, while ecosystem service and natural capital assessments are effective in facilitating collaboration among financiers, policymakers, conservationists and stakeholders to drive investment in natural assets[16], this is still acknowledged as an obstacle at the local scale[17]. To this end, Burdon et al.[18] demonstrated the use of participatory mapping to enable engagement among local stakeholders. Another common obstacle facing ecosystem accounting at the local scale is the lack of quality place-based natural capital data and environmental baselines, especially accurate ecological baselines of coastal and estuarine areas.[17-19] In management interventions, reporting against baselines is especially pertinent for habitat restoration.[17,20] In such instances, baselines can assist with the identification of scale and nature of investment required to restore or improve natural habitats.[17] Thus the application of ecosystem accounting as a monitoring and reporting tool in restoration management is especially relevant in this UN Decade on Ecosystem Restoration (2021–2030).[21,22]

Estuarine ecosystems provide disproportionate value to society in comparison to most other habitats, and as a result, they are highly utilised.[17] Poor management of anthropogenic impacts and ineffective conservation approaches, together with global economic changes, have resulted in many estuaries becoming both economically and environmentally degraded.[17] The ecosystem services provided by estuaries often are essential to those relying on them, and with most reliant communities being under increasing economic stress, the pressures on these valuable systems are unlikely to decline.[17] Estuaries, therefore, are viewed as ideal ecosystems for the implementation of natural capital approaches paving the way to green economic growth and improved well-being.[17]

Since 2014, South Africa has made significant progress in the development of ecosystem accounting as a subset of natural capital accounting. Formal structures and partnerships in government have been formed between Statistics South Africa (Stats SA), the South African National Biodiversity Institute (SANBI) and the Department of Forestry, Fisheries and the Environment (DFFE) to co-ordinate and lead the development of ecosystem accounts. In consultation with a range of national and sub-national stakeholders, they pilot tested SEEA Ecosystem Accounting in South Africa. Similar structures have yet to be formally proposed for ocean accounting, but it is envisaged to be the same partnerships with possible shifts in roles and responsibilities. In 2019, the Council for Scientific and Industrial Research (CSIR), in collaboration with DFFE and the Nelson Mandela University (NMU), in consultation with SANBI and Stats SA, undertook a project to develop ecosystem accounting methods for estuaries in South Africa, and prepared the first country-level physical accounts for estuaries.[14] They largely used the country-level information generated as part of the National Biodiversity Assessment 2018 (NBA 2018) to populate accounts and to demonstrate the interlinkages between the NBA and Ecosystem Accounting.[23]

We investigate an approach and application of ecosystem accounting at the local estuary level using the Swartkops Estuary as a case study, focusing on physical accounts.[8] Specifically, we demonstrate methods by which to collect and present place-based extent and condition data, identified as one of the obstacles facing ecosystem and natural capital approaches in local applications.[17-19] Further, we build on the approach previously applied in the country-level accounting for South African estuaries[14] primarily to enable seamless alignment between local to national assessments. As with the country-level accounting, we also introduce a local pressure account, and demonstrate the construct of two important local ecosystem services physical accounts. This is the first reported study for South Africa in which ecosystem accounting was investigated in detail at the local estuary level.

Approach and methods

Study approach

Ecosystem accounting relies on environmental monitoring data, often the same data that are applied in other environmental management domains, albeit presented in a different format. In this study, we use existing data on the Swartkops Estuary to construct the estuary-level ecosystem accounts, primarily derived from the NBA 2018[23], but refined with information from more recent high-resolution studies[24,25].

To ensure comparability and integration across resource management initiatives[2], it is important for ecosystem accounting approaches to align with other existing resource management tools. Therefore, in our approach and method for the estuary-level accounts developed in this study, we drew strongly on those applied in the country-level estuarine ecosystem accounts[14], as well as key estuary-specific approaches adopted in South Africa's estuarine environmental flow requirement method[26-29]. Previous methods adopted the estuarine functional zone (EFZ) (demarcating the official spatial boundaries of estuaries) as the basic spatial unit (BSU). In the estuary-level accounts, the EFZ also demarcates the estuary boundaries, but further spatial resolution is introduced through the demarcation of estuary zones (i.e. smaller BSUs). As with the environmental flow requirement methods, the estuarine ecosystem is resolved into the key abiotic components (hydrology, hydrodynamics, salinity, water quality, and physical habitat) and biotic components (microalgae, macrophytes, invertebrates, fish and birds) for the estuary-scale extent and conditions accounts.

Extent accounting

Ecosystem extent accounts typically reflect change in the area cover of habitats, and abundance or biomass of associated biota. For the study, extent accounts distinguish between abiotic and biotic habitats within the EFZ, disaggregated into the smaller BSUs. Unfortunately, insufficient data were available to prepare extent accounts for other biotic components such as microalgae, invertebrates, fish, and birds. Based on the earliest available historical data on the extent of habitats in the Swartkops Estuary, the opening account dates from 1942.

Condition accounting

Condition accounts primarily reflect change in the health of ecosystems. In South Africa, the term 'estuary health' is used to describe an estuary's condition – measured as the degree to which the present condition deviates from its pristine condition.[26,27,30] The Estuarine Health Index (EHI) of South Africa was developed and reflects change as a percentage similarity (0–100%) to a defined natural state (referred to as the 'Reference Condition'). Condition is estimated for both abiotic (hydrology, hydrodynamics, salinity, water quality and physical habitat) and biotic indicators (micro-algae, macrophytes, invertebrates, fish and birds) derived from various data and information sources. Ratings for indicators are weighted (25% for each abiotic and 20% for each biotic component) and aggregated (50:50) to provide an overall percentage deviation from natural.[31] These percentage values are then translated into six ecological condition categories, ranging from natural (A) to critically modified (F). These categories also represent declining functionality in process and pattern, from natural to little remaining. To ensure alignment with existing health assessment processes for estuaries in South Africa, the construct of ecosystem condition accounts at the estuary level also adopted the EHI to express condition.[26,30] In the case of the Swartkops Estuary, ecological condition data were not available for 1942. For this reason, the initial opening account assumes a pre-industrial natural, roughly estimated as 1750s in line with other national assessments.[14]

Pressure accounting

Pressure accounts are useful to understand the impacts that contribute to change in ecosystem extent and condition. They can also highlight where pressures have increased and where management intervention is needed, including more frequent monitoring and assessment. Six main pressure categories have been identified for South African estuaries, namely: flow modification, land use and development, exploitation of living resources (fishing), pollution (wastewater disposal), inlet manipulation, and biological invasion (plants and fish).[32] To estimate the magnitude of pressure in these accounts, the method of Van Niekerk et al.[32] was applied, using various indicators in the EHI as proxies (Table 1).

Ecosystem service accounting (physical)

In essence, ecosystem service accounts express the 'amount' of a service that can be delivered by specific environmental components, depending on both the extent and condition of the component. In this study, we investigated the construct of physical accounts for two important ecosystem services provided by estuaries, that is, carbon sequestration and recreational use.

In the case of carbon sequestration, the account requires information on ecosystem delivery capacity, e.g. *carbon sequestration (C t/a) = Salt marsh extent (ha) * f (dependent on condition).* Such information can then be used to translate the 'amount' of ecosystem service flows into a monetary value (if relevant), e.g. m*onetary value (R 000'K) = carbon sequestration (C t/a) * Unit market value.* Carbon sequestration accounts distinguish between sediment and aboveground biomass.[33] Matching the biotic extent account, the initial opening account for carbon sequestration dates from 1942.

For recreational use, the account evaluates suitability of use by comparing microbiological indicator data against water quality guidelines for recreation, as summarised in Table 2.[34,35]

Table 1: Indicators of magnitude and proxies for extent of impact of key pressures in estuaries (adapted from Van Niekerk et al.[32])

Pressure	Magnitude of pressure	Proxy for extent of impact
Flow modification	Modification in mean annual run-off	'Hydrology' in Estuarine Health Index (EHI), where 'low' >75% (similar to natural, Categories A–B), 'medium' 75–60% (Category C), 'high' 6040% (Category D), 'very high' <40% (Categories E–F)
Land use and development	Visual examination using Google Earth™	'Physical habitat' in EHI, where 'low' >75%, 'medium' 75–60%, 'high' 60–40%, 'very high' <40%
Exploitation of living resources (fishing)	Available information on fish catch (tons), as well as illegal catch records	'Fish' in EHI, as well as fishing effort, catch composition and size, age distribution of fish in sample data sets, where 'none' = no fish present, 'low' = little fishing pressure, 'medium' = medium level of recreational/subsistence exploitation, 'high' = high recreational or subsistence exploitation, 'very high' = gillnetting or fish traps
Pollution (wastewater)	Daily volume of municipal and industrial wastewater discharges	'Water quality' (excluding salinity) in EHI, where 'low' >75%, 'medium' = 75–60%, 'high' 60–40%, 'very high' <40%
Inlet manipulation	Extent of artificial manipulation of inlet	'Hydrodynamic' in EHI, 'low' >75%, 'medium' = 75–60%, 'high' 60–40%, 'very high' <40%
Biological invasions	Alien plants: visual examination of extent along estuary	Three-tier scale, where 'low' = alien coverage <5% of estuarine functional zone (EFZ), 'medium' = alien coverage 5–15% of EFZ, 'high' = alien coverage >15% of EFZ
	Alien or extralimital fish: estimated from species lists	Number and type of fish species present, where 'low' = 1–2 species, 'medium' = 3–4 species, 'high' = 5–6 species, or 1–2 predatory species, or sharptooth catfish *Clarias gariepinus* present (if translocated), 'very high' >7 species, or 3–4 predatory species

Table 2: Criteria for recreational use of coastal waters in South Africa[34]

STATUS	Enterococci (colony forming units/100 mL)	*Escherichia coli* (colony forming untis/100 mL)
Excellent	≤ 100 (95th percentile)	≤ 250 (95th percentile)
Good	≤ 200 (95th percentile)	≤ 500 (95th percentile)
Fair (minimum)	≤ 185 (90th percentile)	≤ 500 (90th percentile)
Poor	> 185 (90th percentile)	> 500 (90th percentile)

These are typically based on 95/90 percentile values calculated over a 12-month period (sampled twice monthly). The 'fair' category is accepted as the minimum requirement for contact recreational use in South Africa. For recreational use, the initial opening accounts assumes the pre-industrial natural conditions, roughly estimated as the 1750s.[14]

Study area

The Swartkops Estuary (Figure 1) was selected as a case study to demonstrate the application of ecosystem accounting at the individual estuary level based on the availability of extent (spatial) and condition data recently generated as part of a local flow requirement and restoration project[24], thus providing a sound foundation for the development of a range of accounts. The estuary is a permanently open system located in the warm temperate region of the Eastern Cape on the outskirts of Gqeberha (previously called Port Elizabeth) with a catchment area of 1390 km^2. The estuary's length spans approximately 16.4 km, with the upper reaches being about 90 m wide with steep banks and winding channels becoming shallower, wider, and straighter towards the mouth. The lower reaches have extensive intertidal mudflats, islands, salt marshes and sandbanks. The estuary is relatively shallow with water depth generally ranging between 1 m and 2 m.[24] The boundaries of the EFZ of the Swartkops Estuary, comprising an area of 2 861 ha, are defined by[23]:

- Downstream boundary – 33°51'53.14"S, 25°37'57.41"E (estuary mouth)
- Upstream boundary – 33°48'45.20"S, 25°31'29.20"E (at Perseverance)
- Lateral boundaries – 5 m contour above mean sea level along each bank

The zonation of the estuary used in the flow requirement and restoration study[24] was also adopted as the three BSUs for the Swartkops Estuary ecosystem accounts (Figure 1). The zones were based on general homogeneity in bathymetry, salinity patterns and water retention. Zone A comprises the larger and deeper lower reaches (1103 ha), Zone B is the middle reaches (1015 ha), and Zone C (743 ha) is the shallower upper reaches.

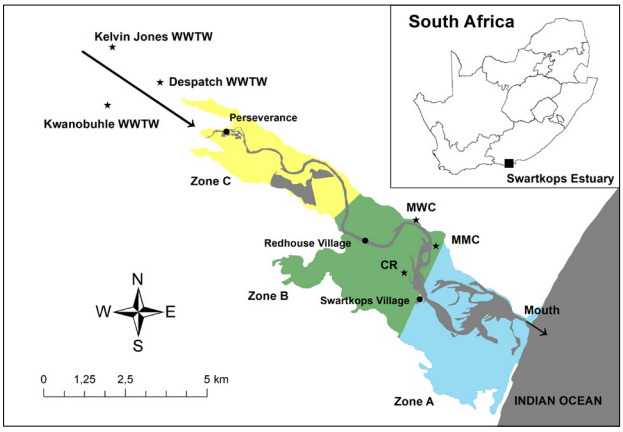


Figure 1: Map of the Swartkops Estuary, showing the location of upstream wastewater treatment works (WWTWs), inflows from Motherwell Canal (MWC), Markman Canal (MMC), Chatty River (CR), recreation water quality sites (Perseverance, Redhouse Village and Swartkops Village), as well as zonation of the estuary (Zones A–C).

Swartkops is a highly urbanised estuary with large areas transformed by industrial and residential development. Effluent from three wastewater treatment works – namely Kelvin Jones (21 120 m^3/day), KwaNobuhle (1980 m^3/day) and Despatch (4410 m^3/day) – enters just upstream of the estuary.[24] The estuary also receives contaminated urban run-off through the Chatty River, draining highly populated townships, the Markman Canal, draining industrial areas, and the Motherwell Canal that drains the township of Motherwell.[35] As a result of these effluent discharges and increased stormwater run-off from hardened urban areas, the present mean annual run-off is approximately 41% higher than the natural mean annual run-off. Construction of roads, bridges, railways, and saltpans in the floodplain also contributed to habitat loss.[24,36] The system is also heavily utilised by local communities for fishing and bait collection.[24]

Four abiotic habitats were considered, namely: open water areas, mud and sand flats, beach and dunes, and floodplains (distinguishing between undeveloped and developed floodplains).[14] Floodplain habitat included all dynamic areas influenced by long-term estuarine sedimentary processes, as captured within the EFZ. Open water areas are highly dynamic and ephemeral, changing in response to numerous factors such as mouth state, river inflow conditions and state of the tide.

To account for biotic habitat extent, we used the key estuarine vegetation types as defined in Adams et al.[37] and updated in the NBA 2018[23] and Adams et al.[24], that is: intertidal salt marsh, supratidal salt marsh,

submerged macrophytes, and reeds and sedges. Unlike estuarine abiotic habitats, biotic habitats do not occur continuously within the EFZ. Rather they are a mosaic of biotic habitats 'superimposed' onto abiotic habitats, and as a result total biotic habitat area does not necessarily add up to 'total EFZ area'. For the Swartkops Estuary, aerial photographs of 50 cm resolution were used to manually digitise change in the extent of the key biotic habitats. This was done using ArcGIS 10.6.1 on an average scale of 1:2000. This fine-scale mapping is preferred over the supervised and unsupervised classification methods using satellite imagery because habitat in estuaries is often only a few square metres in extent. Present extent was verified with ground-truthing. Older aerial imagery and historical data were used to determine change in extent over time. Aerial photographs of 50 cm resolution were obtained from the Chief Directorate: National Geo-Spatial Information (http://www.cdngiportal.co.za/cdngiportal/; their images date back to the 1940s.

For the carbon sequestration accounts, available data were sourced from Van Niekerk et al.[23], Adams et al.[24] and Raw et al.[25] The ecosystem account for recreation was constructed using available microbiological indicator data (Swartkops Conservancy, unpublished data) together with national water quality guidelines for recreation, evaluating suitability based on estimated gastrointestinal and febrile respiratory health risks.[33]

Results

Extent accounts

Table 3 presents the abiotic habitat extent account for the Swartkops Estuary, reflecting change from 1942 to 2018, and then 2021, while Figure 2 shows the geo-referenced maps from where data were derived.[23,24]

Table 3: Extent account for abiotic habitat in the Swartkops Estuary

	TOTAL (ha)	HABITAT EXTENT (expressed in ha)		
		Zone A	Zone B	Zone C
OPENING STOCK (1942):				
Open water area	**409.1**	225.1	107.3	76.7
Mud/sand banks	**120.9**	101.8	19.1	0.0
Beach and dunes	**32.3**	32.3	0.0	0.0
Floodplain	**2316.2**	761.3	888.6	666.3
Developed	**37.5**	37.5	0.0	0.0
Increase/decrease:				
Open water area	***0***	0	0	0
Mud/sand banks	***0***	0	0	0
Beach and dunes	***0.1***	+0.1	0	0
Floodplain	***-844.8***	-224.3	-393.2	-227.4
Developed	***+844.8***	+224.2	+393.2	+227.4
CLOSING STOCK (2018):				
Open water area	**409.1**	225.1	107.3	76.7
Mud/sand banks	**120.9**	101.8	19.1	0.0
Beach and dunes	**32.4**	32.4	0.0	0.0
Floodplain	**1471.3**	537.0	495.4	438.9
Developed	**882.3**	261.7	393.2	227.4
Reappraisal increase/decrease:				
Open water area	***0***	*0*	*0*	*0*
Mud/sand banks	***0***	*0*	*0*	*0*
Beach and dunes	***0***	*0*	*0*	*0*
Floodplain	***0***	*0*	*0*	*0*
Developed	***0***	*0*	*0*	*0*
CLOSING STOCK (2021):				
Open water area	**409.1**	225.1	107.3	76.7
Mud/sand banks	**120.9**	101.8	19.1	0.0
Beach and dunes	**32.4**	32.4	0.0	0.0
Floodplain	**1471.3**	537.0	495.4	438.9
Developed	**882.3**	261.7	393.2	227.4

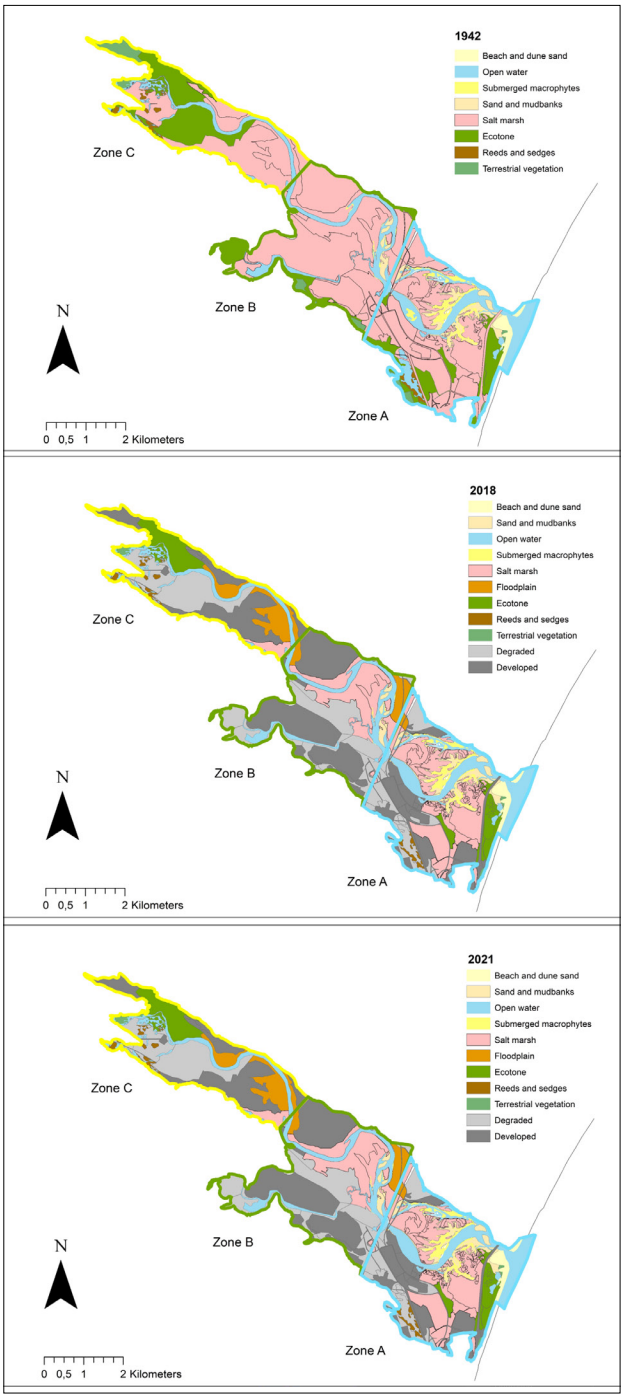


Figure 2: Geo-referenced maps depicting change in abiotic and biotic habitat extent in the Swartkops Estuary between 1942, 2018 and 2021.

Floods play a key role in maintaining sediment processes and the long-term equilibrium of the sediment dynamics in the Swartkops system. Because floods have not been affected significantly by water resource developments in the catchments, the EFZ extent and associated abiotic habitats in the Swartkops Estuary did not change markedly between 1942 and 2018, except in the floodplains. By 1942, about 37.5 ha of the floodplain had already been developed, mostly in the lower reaches (Zone A). Between 1942 and 2018, another 844.8 ha of floodplain was lost due to development, with most of this occurring in Zone B (393.2 ha). This loss was mainly attributed to residential and industrial development, stemming from housing, bridges, roads, railways, mining, and saltpans.[38,39] No additional losses were evident in the 2021 reappraisal, possibly owing to the relatively short time between 2018 and 2021.

Table 4 presents the biotic habitat extent account for the Swartkops Estuary, reflecting change from 1942 to 2018, and to 2021, while Figure 2 displays the spatial changes between the different periods.[23,24] The estuary has extensive areas of supratidal and intertidal salt marsh and large stands of the endangered seagrass *Zostera capensis.* The latter occurs in the lower intertidal zone and is abundant in the middle and lower estuary reaches. An additional seagrass species *Halophila ovalis* is also present in subtidal areas, but its precise distribution is unknown.

Table 4: Extent account for biotic habitat in the Swartkops Estuary

	TOTAL (ha)	HABITAT EXTENT (expressed in ha)		
		Zone A	Zone B	Zone C
OPENING STOCK (1942):				
Intertidal saltmarsh	**536.9**	230	260.4	46.5
Supratidal saltmarsh	**642.6**	359	196	87.6
Submerged macrophytes	**53.6**	49.3	4.3	0.0
Reeds and sedges	**21.8**	10.3	0.0	11.5
Floodplain/ecotone (>2.5 m)	**881.4**	95.3	90.5	695.6
Terrestrial	**248.5**	62.0	62.0	124.5
Increase/decrease:				
Intertidal saltmarsh	***-344.5***	*-79.5*	*-226*	*-39*
Supratidal saltmarsh	***-283.6***	*-180.4*	*-33.2*	*-70*
Submerged macrophytes	***0***	*0*	*0*	*0*
Reeds and sedges	***0***	*0*	*0*	*0*
Floodplain/ecotone (>2.5 m)	***-670.1***	*-89.0*	*-46.6*	*-534.5*
Terrestrial	***-237.2***	*-61.5*	*-61.1*	*-114.5*
CLOSING STOCK (2018):				
Intertidal saltmarsh	**192.4**	150.5	34.4	7.5
Supratidal saltmarsh	**358.9**	178.6	162.8	17.6
Submerged macrophytes	**53.6**	49.3	4.3	0.0
Reeds and sedges	**21.8**	10.3	0.0	11.5
Floodplain/ecotone (>2.5 m)	**307.4**	6.3	43.9	161.1
Terrestrial	**11.7**	0.5	0.9	10.0
Reappraisal increase/decrease:				
Intertidal saltmarsh	***0***	*0*	*0*	*0*
Supratidal saltmarsh	***0***	*0*	*0*	*0*
Submerged macrophytes	***0***	*0*	*0*	*0*
Reeds and sedges	***0***	*0*	*0*	*0*
Floodplain/ecotone (>2.5 m)	***0***	*0*	*0*	*0*
Terrestrial	***0***	*0*	*0*	*0*
CLOSING STOCK (2021):				
Intertidal saltmarsh	**192.4**	150.5	34.4	7.5
Supratidal saltmarsh	**358.9**	178.4	162.8	17.6
Submerged macrophytes	**53.6**	49.3	4.3	0.0
Reeds and sedges	**21.8**	10.3	0.0	11.5
Floodplain/ecotone (>2.5 m)	**307.4**	6.3	43.9	161.1
Terrestrial	**11.7**	0.5	0.8	10.0

The dominant species in the supratidal salt marsh include *Salicornia pillansii, Suaeda fructicosa, Sporobolus virginicus* and *Disphyma crassifolium*. Important intertidal species are *Spartina maritima*, *Salicornia tegetaria*, *Triglochin* spp. and *Limonium* spp. *Cotula coronopifolia* occurs amongst the salt marsh habitat in the middle reaches where the water is more brackish. Floodplain species such as *Lycium cinereum* and *Suaeda fructicosa* occur together with *Stenotaphrum secundatum* and *S. pillansii* in the upper reaches of the Swartkops Estuary.[40] Supratidal salt marsh species such as *Disphyma crassifolium* grow best at a salinity of 1–19 under dry conditions.[41] Reeds and sedges mainly occur in the upper reaches and at freshwater seepage sites but are not extensive due to the disturbance of the banks. The common reed *Phragmites australis* is dominant. Macroalgae (seaweeds) are common in the saline areas of the estuary and include *Codium tenue* (Kutzing) Kutzing, *Ulva rigida* C. Agardh, *Gracillaria gracilis* and *Hypnea viridis* Papenfuss.

The biotic extent account indicates that, overall, there has been a 344.5 ha decrease in intertidal salt marsh and a 283.6 ha decrease in supratidal salt marsh area, with most losses occurring in Zones A and B. There was also a 670.1 ha and 237.2 ha loss in floodplain/ecotone and terrestrial habitat, respectively, but mostly in Zone C. As with changes in abiotic floodplain area, land use development was largely responsible for losses in biotic habitats, with most of the development occurring in the floodplain/ecotone and terrestrial habitats.[39]

Condition account

The condition account for the Swartkops Estuary is provided in Table 5, with Figure 3 providing a graphic overview of the changes from ~1750 (natural), 2018 and 2021. At the estuary level, condition accounts should ideally also distinguish between different zones in a system. However, available data on estuary condition did not disaggregate ratings into the three zones, except for 'salinity', 'water quality' (including dissolved oxygen, turbidity, nutrients, and toxic substances) and 'microalgae' in the 2021 account. Notwithstanding, to demonstrate the construct of a condition account at the estuary level, the overall condition ratings for various indicators were used as proxies for condition in each zone, except for the above three indicators in the 2021 account. In the case of the Swartkops Estuary, data on the ecological condition of the system were not available for 1942. As a result, the opening account was equated to the pre-industrial natural, which has been assumed as ~1750. Comparing the opening account with the 2018 closing account, the condition of the Swartkops Estuary declined by 47%, with the deterioration in biotic condition (52%) being higher than the decline in abiotic condition (42%).

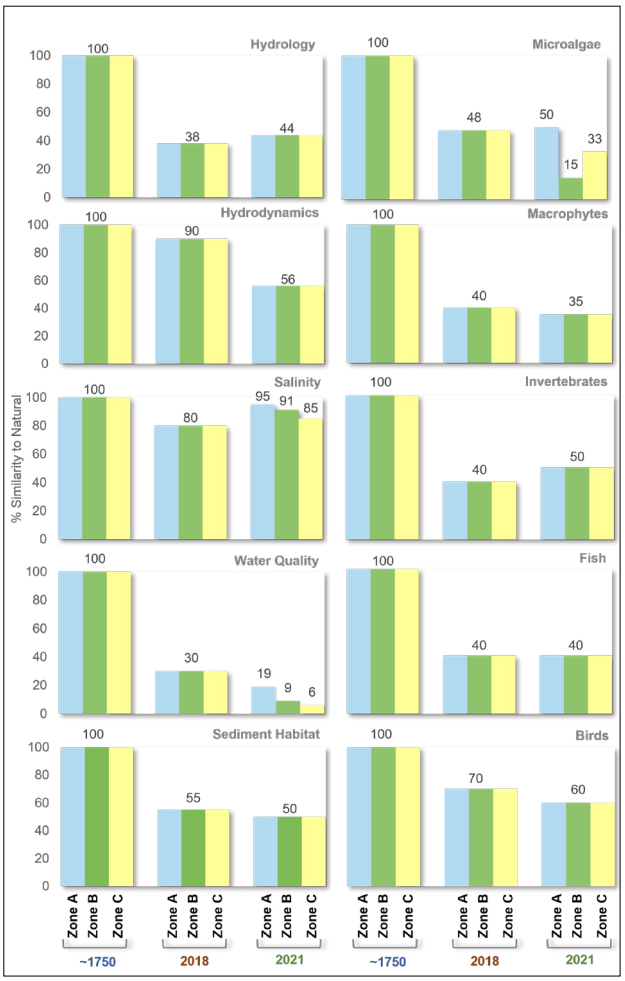


Figure 3: Change in condition in selected abiotic and biotic indicators across Zones A, B and C in the Swartkops Estuary between ~1750 (natural), 2018 and 2021.

Following a reappraisal in 2021, the estuary's condition across Zones A, B and C deteriorated further by 5%, 9% and 8%, respectively. Here the additional deterioration was most evident in the abiotic components (8%, 10% and 11%, respectively), driven by shifts in hydrodynamics and water quality indicators (Table 5) due to an increase in input from wastewater treatment works. The biotic condition across Zones A, B and C also deteriorated by a further 1%, 8% and 4%, respectively. This was largely associated with a deterioration in microalgae condition, linked to the deterioration in water quality (increased nutrient enrichment) during this period.

Pressure account

The pressure account for the Swartkops Estuary is provided in Table 6 and depicts the change from natural to 2018, and then 2021. As available data on pressures were not disaggregated into the smaller BSUs (or zones), the pressure account reverts to the EFZ as its BSU, and therefore reflects an overall pressure rating for the Swartkops Estuary. We focused on five of the six key pressure categories identified for South African estuaries, and encountered in the Swartkops, that is: flow modification, land use and development, exploitation of living resources (fishing), pollution (wastewater disposal), and biological invasion (plants and fish).[32] The Swartkops Estuary is a permanently open system and, as yet, mouth manipulation has not emerged as a key pressure and was therefore not included here.

Flow modification increased freshwater inputs to the system by 41%, mainly from wastewater treatment works and urban stormwater discharges, resulting in very high pressure on the system. Similarly, land use and development within the EFZ have resulted in a loss of 847.7 ha, emerging as a very high pressure. Fishing pressures are also very high in this system with up to 47 tons of fish harvested annually, mostly consisting of undersized juveniles and/or threatened species, and through the use of unsustainable practices such as gill netting. Approximately 27 510 m^3 wastewater is discharged daily into the system, resulting in very high pollution pressures.

The Swartkops is also under high pressure from a combination of invasive aquatic and terrestrial plant species. Noteworthy is the high abundance of alien aquatic species such as *Azolla filliculoides*, *Salvinia molesta* and *Eicchornia crassipes,* especially in the upper reaches (Zone C). Two carnivorous invasive fish species – largemouth bass *Micropterus salmoides* and smallmouth bass *Micropterus dolomieu* –, recorded in the estuary contribute to the high pressure from such alien invasion, especially in the upper fresh and more brackish regions (Zones B and C), thereby also affecting connectivity between the estuary and river.

Table 5: Ecosystem condition account for the Swartkops Estuary based on selected abiotic and biotic indicators (derived from Van Niekerk et al. and Adams et al.[23,24])

	ECOLOGICAL CONDITION (expressed as % similarity to natural)		
	Zone A	Zone B	Zone C
OPENING STOCK (~1750):			
OVERALL	**100**	**100**	**100**
ABIOTIC CONDITION INDICATOR	**100**	**100**	**100**
Hydrology	100	100	100
Hydrodynamics	100	100	100
Salinity	100	100	100
Water quality	100	100	100
Sedimentary habitat	100	100	100
BIOTIC CONDITION INDICATOR	**100**	**100**	**100**
Microalgae	100	100	100
Macrophytes	100	100	100
Invertebrates	100	100	100
Fish	100	100	100
Birds	100	100	100
Increase/decrease:			
OVERALL	*-47*	*-47*	*-47*
ABIOTIC CONDITION INDICATOR	*-42*	*-42*	*-42*
Hydrology	*-62*	*-62*	*-62*
Hydrodynamics	*-10*	*-10*	*-10*
Salinity	*-20*	*-20*	*-20*
Water quality	*-75*	*-75*	*-75*
Sedimentary habitat	*-45*	*-45*	*-45*
BIOTIC CONDITION INDICATOR	*-52*	*-52*	*-52*
Microalgae	*-52*	*-52*	*-52*
Macrophytes	*-60*	*-60*	*-60*
Invertebrates	*-60*	*-60*	*-60*
Fish	*-60*	*-60*	*-60*
Birds	*-30*	*-30*	*-30*
CLOSING STOCK (2018):			
OVERALL	**53**	**53**	**53**
ABIOTIC CONDITION INDICATOR	**58**	**58**	**58**
Hydrology	38	38	38
Hydrodynamics	90	90	90
Salinity	80	80	80
Water quality	30	30	30
Sedimentary habitat	55	55	55
BIOTIC CONDITION INDICATOR	**48**	**48**	**48**
Microalgae	48	48	48
Macrophytes	40	40	40
Invertebrates	40	40	40
Fish	40	40	40
Birds	70	70	70
Reappraisal increase/decrease:			
OVERALL	*-5*	*-9*	*-8*
ABIOTIC CONDITION INDICATOR	*-8*	*-10*	*-11*
Hydrology	*6*	*6*	*6*
Hydrodynamics	*-34*	*-34*	*-34*
Salinity	*15*	*11*	*5*
Water quality	*-11*	*-21*	*-24*
Sedimentary habitat	*-5*	*-5*	*-5*
BIOTIC CONDITION INDICATOR	*-1*	*-8*	*-4*
Microalgae	*2*	*-33*	*-15*
Macrophytes	*-5*	*-5*	*-5*
Invertebrates	*10*	*10*	*10*
Fish	*0*	*0*	*0*
Birds	*-10*	*-10*	*-10*
CLOSING STOCK (2021):			
OVERALL	**48**	**44**	**45**
ABIOTIC CONDITION INDICATOR	**50**	**48**	**47**
Hydrology	44	44	44
Hydrodynamics	56	56	56
Salinity	**95**	**91**	**85**
Water quality	**19**	**9**	**6**
Sedimentary habitat	50	50	50
BIOTIC CONDITION INDICATOR	**47**	**40**	**44**
Microalgae	**50**	**15**	**33**
Macrophytes	35	35	35
Invertebrates	50	50	50
Fish	40	40	40
Birds	60	60	60

Table 6: Pressure account for the Swartkops Estuary

	PRESSURE MAGNITUDE	DEGREE OF IMPACT ON ESTUARY (expressed as % of estuarine area affected)			
		Low	Medium	High	Very high
OPENING STOCK (Pre-industrial natural):					
Flow modification	–	**0**	**0**	**0**	**0**
Land use and development	–	**0**	**0**	**0**	**0**
Pollution: wastewater disposal	–	**0**	**0**	**0**	**0**
Exploitation of living resources: fishing	–	**0**	**0**	**0**	**0**
Invasive plants	–	**0**	**0**	**0**	**0**
Alien or extralimital fish	–	**0**	**0**	**0**	**0**
Increase/decrease:					
Flow modification		*0*	*0*	*0*	*+100*
Land use and development		*0*	*0*	*0*	*+100*
Pollution: wastewater disposal		*0*	*0*	*0*	*+100*
Exploitation of living resources: fishing		*0*	*0*	*+100*	*0*
Invasive plants		*0*	*0*	*+100*	*0*
Alien or extralimital fish		*0*	*0*	*+100*	*0*
CLOSING STOCK (2018):					
Flow modification (%MAR change)	**41%⇧**	**0**	**0**	**0**	**100**
Land use and development (ha)	**847.7**	**0**	**0**	**0**	**100**
Pollution: wastewater disposal (m3/d)	**27 510**	**0**	**0**	**0**	**100**
Exploitation of living resources: fishing (t/a)	**46.7**	**0**	**0**	**100**	**0**
Invasive plants (# species)	**13**	**0**	**0**	**100**	**0**
Alien or extralimital fish (# species)	**2**	**0**	**0**	**100**	**0**
Reappraisal increase/decrease:					
Flow modification		*0*	*0*	*0*	*0*
Land use and development		*0*	*0*	*0*	*0*
Pollution: wastewater disposal		*0*	*0*	*0*	*0*
Exploitation of living resources: fishing		*0*	*0*	*0*	*0*
Invasive plants		*0*	*0*	*0*	*0*
Alien or extralimital fish		*0*	*0*	*0*	*0*
CLOSING STOCK (2021):					
Flow modification (%MAR change)	41%⇧	**0**	**0**	**0**	**100**
Land use and development (ha)	847.7	**0**	**0**	**0**	**100**
Pollution: wastewater disposal (m^3/d)	27 510	**0**	**0**	**0**	**100**
Exploitation of living resources: fishing (t/a)	46.7	**0**	**0**	**100**	**0**
Invasive plants (# species)	13	**0**	**0**	**100**	**0**
Alien or extralimital fish (# species)	2	**0**	**0**	**100**	**0**

MAR, mean annual run-off

Ecosystem service accounts

Carbon sequestration

Table 7 presents the physical ecosystem service account for the carbon sequestration of the Swartkops Estuary, based on the extent of important blue and teal carbon vegetation types (see Table 4), with Figure 4 graphically showing change between 1942, 2018 and 2021. The data on storage per unit area were sourced from Raw et al.[25] While results between 2018 and 2021 did not show any marked change in carbon sequestration potential, a nett loss of 266 356.8 Mg (sediment plus biomass [above]) occurred between 1942 and 2018, largely attributed to a loss in intertidal and supratidal salt marsh habitat (Table 4). The highest loss in carbon sequestration potential was encountered in Zone B (109 900.8 Mg), followed by Zone C (46 216.0 Mg) and then Zone A (10 197.6 Mg) (Table 4).

Recreational use

Table 8 demonstrates a physical ecosystem service account for recreational use in the Swartkops Estuary. Drawing on the extent account (Table 3), the open water area potentially available for water recreational activities may not have changed over time, although this does not mean that the entire area is suitable for recreational activities. To reflect change in condition, focusing on human health, *Escherichia coli* data – collected by the Swartkops Conservancy at popular recreation spots at Settlers Bridge and Swartkops Village (Zone A) and Perseverance (Zone C) (Swartkops Conservancy, unpublished data) – were consulted. However, only Swartkops Village (Zone A) and Redhouse Village (Zone B) had sufficient data over a one-year period (2019/2020) to demonstrate the application of the recreational guidelines to ecosystems accounting (using a percentile calculation approach developed globally – see Table 2) (Table 8).

Based on 2019/20 data, Swartkops Village (Zone A) was categorised as 'poor' (Table 2 – not suitable for recreational use), representing a marked deterioration from natural when it was assumed that conditions would have been representative of 'excellent' without any human bacterial contamination. Conditions at Redhouse Village (Zone B) also showed deterioration but are still categorised as 'fair' (Table 2), meeting the minimum requirements for recreational use.

Table 7: Ecosystems services account for carbon sequestration potential in Swartkops Estuary

	TOTAL	STORAGE (expressed as Mg)							
		Intertidal saltmarsh		Supratidal saltmarsh		Submerged macrophytes		Reeds and sedges	
		Sediment	Biomass (above)	Sediment	Biomass (above)	Sediment	Biomass (above)	Sediment	Biomass (above)
Storage/unit area (Mg/ha)		**169**	**255**	**169**	**255**	**125**	**108**	n.d.	**100**
OPENING STOCK (1930/40s):									
	514786.8	**90736.1**	**136909.5**	**108599.4**	**163863.0**	**6700.0**	**5788.8**	n.d.	**2190.0**
Zone A	**262252.9**	38870.0	58650.0	60671.0	91545.0	6162.5	5324.4	n.d.	1030.0
Zone B	**194515.5**	44007.6	66402.0	33124.0	49980.0	537.5	464.4	n.d.	0.0
Zone C	**58018.4**	7858.5	11857.5	14804.4	22338.0	0.0	0.0	n.d.	1160.0
Gains/losses:									
	-266356.8	***-58186.7***	***-87796.5***	***-47945.3***	***-72343.5***	***0.0***	***0.0***	***n.d.***	***0.0***
Zone A	*-10197.6*	*-13435.5*	*-20272.5*	*-30487.6*	*-46002.0*	*0.0*	*0.0*	*n.d.*	*0.0*
Zone B	*-109900.8*	*-38194.0*	*-57630.0*	*-5610.8*	*-8466.0*	*0.0*	*0.0*	*n.d.*	*0.0*
Zone C	*-46216.0*	*-6591.0*	*-9945.0*	*-11830.0*	*-17850.0*	*0.0*	*0.0*	*n.d.*	*0.0*
CLOSING STOCK (2018):									
	248430.0	**32515.6**	**49062.0**	**60654.1**	**91519.5**	**6700.0**	**5788.8**	**n.d.**	**2190.0**
Zone A	**152055.3**	25434.5	38377.5	30183.4	45543.0	6162.5	5324.4	n.d.	1030.0
Zone B	**84614.7**	5813.6	8772.0	27513.2	41514.0	537.5	464.4	n.d.	0.0
Zone C	**11802.4**	1267.5	1912.5	2974.4	4488.0	0.0	0.0	n.d.	1160.0
Gains/losses:									
	0	***0***	***0***	***0***	***0***	***0***	***0***	***n.d.***	***0***
Zone A	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*n.d.*	*0*
Zone B	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*n.d.*	*0*
Zone C	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*n.d.*	*0*
CLOSING STOCK (2021):									
	248430.0	**32515.6**	**49062.0**	**60654.1**	**91519.5**	**6700.0**	**5788.8**	**n.d.**	**2190.0**
Zone A	**152055.3**	25434.5	38377.5	30183.4	45543.0	6162.5	5324.4	n.d.	1030.0
Zone B	**84614.7**	5813.6	8772.0	27513.2	41514.0	537.5	464.4	n.d.	0.0
Zone C	**11802.4**	1267.5	1912.5	2974.4	4488.0	0.0	0.0	n.d.	1160.0

n.d., no data

Table 8: Ecosystems services account for recreational use in Swartkops Estuary

	AREA & SUITABILITY FOR CONTACT RECREATION (based on *E. coli* counts – see Table 2)		
	Zone A (Swartkops Village)	Zone B (Redhouse Village)	Zone C (Perseverance)
OPENING STOCK (PRE-INDUSTRIAL NATURAL):			
Open water area (ha)	225.1	107.3	76.7
90%ile for *E. coli* (counts per 100 mLl)	0	0	0
95%ile for *E. coli* (counts per 100 mL)	0	0	0
Suitability for recreational use	**Excellent**	**Excellent**	**Excellent**
Increase/Decrease:			
Open water area (ha)	*0*	*0*	*0*
90%ile for E. coli *(counts per 100 mL)*	*+290*	*+250*	*n.d.*
95%ile for E. coli *(counts per 100 mL)*	*+2468*	*+898*	*n.d.*
Suitability for recreational use	⇩⇩	⇩	*n.d.*
CLOSING STOCK (2019/2020):			
Open water area (ha)	225.1	107.3	76.7
90%ile for *E. coli* (counts per 100 mL)	754	282	n.d.
95%ile for *E. coli* (counts per 100 mL)	2255	817	n.d.
Suitability for recreational use	**Poor**	**Fair**	n.d.

n.d., no data

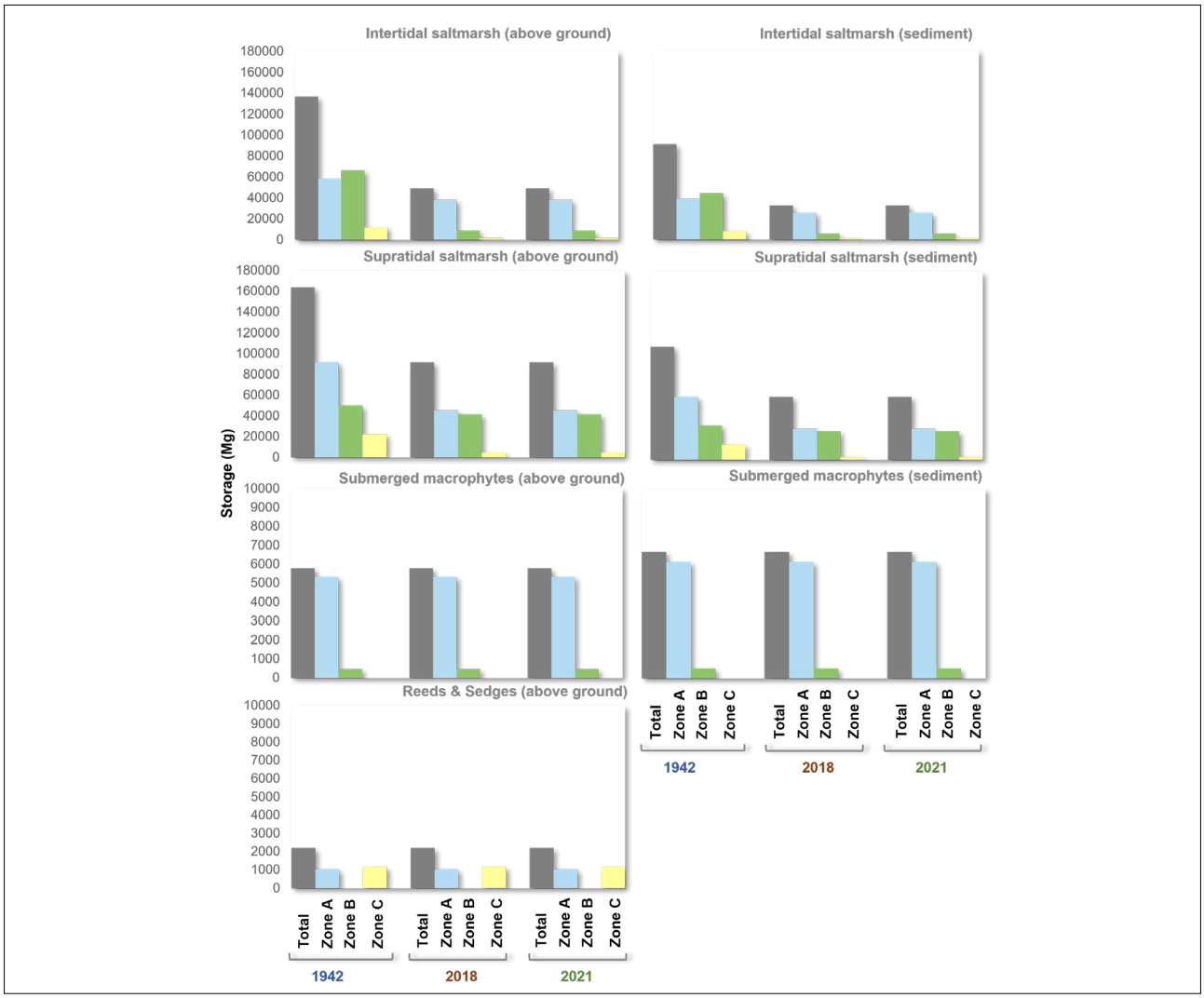


Figure 4: Change in carbon sequestration potential (aboveground biomass and in below ground sediment) across Zones A, B and C in in the Swartkops Estuary between 1942, 2018 and 2021.

Discussion

Globally, most natural capital or ecosystem accounts amalgamate estuaries either into wetland accounts or coastal accounts. However, estuaries offer disproportionally higher socio-economic benefits to society per unit area compared to other natural systems (e.g. nursery areas for important fisheries, carbon sequestration).[17]

Therefore, in South Africa, these important transitional water ecosystem types have been elevated through the provision of separate estuarine ecosystem accounts.[14] This allows for reporting of change at meaningful scales, which is not possible if estuaries are combined in ecosystem accounting with freshwater wetlands or other marine habitats.

In this study, we built on the country-level ecosystem account previously developed for South African estuaries[11] by developing an approach for physical ecosystem accounts at the local estuary level, addressing one of the key obstacles previously identified for application at this scale[17-19]. In so doing, we are introducing the advantages of ecosystem accounting for decision-making to local governance mechanisms.[17] Further, we demonstrated alignment of this method with other existing resource management tools in South Africa, such as the ecological flow requirement methods for estuaries[26-29], the national Estuarine Health Index[26,30] and the National Biodiversity Assessment[23].

The method was then applied to the Swartkops Estuary, depicting ecosystem changes, drawing on available data and information but using an accounting format comparable with socio-economic information systems, such as the System of National Accounts. The abiotic habitat extent account showed a significant additional loss in intact floodplain (844.9 ha) between 1942 and 2018, especially in Zone B (393.2 ha loss). Over the same period, 344.5 ha of intertidal salt marsh and 283.6 ha of supratidal salt marsh were lost, mainly from Zones A and B. The floodplain/ecotone and terrestrial habitats also decreased by 670.1 ha and 237.2 ha, respectively. No measurable habitat losses were evident between 2018 and the 2021 reappraisal. The condition account showed that the Swartkops Estuary declined about 47% in overall condition by 2018 compared with pre-industrial natural, with a further loss of 5%, 9% and 8% in Zones A, B and C, respectively, by 2021. The pressure account demonstrates a simple and practical manner to reflect possible causes of change in estuarine extent and condition, showing that, in the Swartkops flow modification, land use and development and pollution contribute most to the decline in ecosystem health. Because the pressure account largely categorises pressures on a sectoral basis it can easily inform sectoral management planning. For example, in South Africa, responsibility for flow modification management lies with the department responsible for water, while the department responsible for fisheries is mandated to address the exploitation of living resources (fisheries). Pollution (wastewater disposal) is a joint responsibility of the departments responsible for water and the environment. The ecosystem service account for carbon sequestration showed a nett loss of 266 356.8 Mg in carbon sequestration potential between 1942 and 2018, mostly driven by losses in intertidal and salt tidal saltmarsh.

The account for recreational use showed that, in 2019/2020, water quality in the estuary had deteriorated significantly from natural across most zones.

While this study successfully demonstrated the application of ecosystem accounting at the estuary level, it can certainly benefit from incremental improvement in future, following an adaptive management approach. Specifically, to better resolve spatial resolution in the condition accounting, future research should focus on the disaggregation of condition ratings of ecosystem components into the different zone within an estuary. Greater spatial resolution in condition accounting will assist in trade-off interventions across zones within an estuary. Also, future research should investigate approaches by which to disaggregate conditions of abiotic and biotic components, requiring refinement of the EHI method. For carbon sequestration accounting, further research is needed on the refinement of carbon storage values for South African estuaries. Preliminary data suggest that storage values may be lower than those reported globally due to the more arid nature of South Africa's estuaries. While the ecosystem service account for recreation does demonstrate the application of such an account, it is acknowledged that results are based on single data points within zones and may not represent conditions across the entire zones. Therefore, for future accounting, it is recommended that monitoring sites be extended to all important recreational sites within zones to better reflect broader zonal conditions. Also, it is recommended that microbiological data (*E. coli*, preferably also Enterococci) be collected routinely at bi-weekly intervals in alignment with the national guideline requirements.[33]

Another critically important ecosystem service of estuaries is nursery function; it is therefore recommended that future research also focuses on the development thereof. Such accounts should consider fish lifecycle requirements and the location and condition of key habitats that support these. Internationally, ecosystem service accounts for fishing reflect only stock abundance as a measure of extent. However, fish stocks have geographical ranges which can expand, shrink or split through direct pressures (e.g. over-exploitation of fishing resources) or indirect causes (climate change). Therefore, it is important that the spatial element of a stock extent also be captured in these ecosystem accounts to reflect underlying ecosystem shifts in resource abundance. Further, given the socio-economic value of estuaries, and to reflect the opportunity cost of declining extent and condition at the estuary level, future monetary accounts for key ecosystem services such as carbon sequestration, recreational use, and nursery function should build and align with the physical ecosystem accounts presented here.

Finally, reflecting on the three primary applications previously identified for ecosystem accounting[2], these accounts also can play a key role in informing decision-making in restoration prioritisation within individual estuaries in support of the UN Decade on Ecosystem Restoration (2021–2030)[21,22], for example, by assisting in the identification of the scale and nature of investment required for restoration[17], or informing options for offsetting irreversibly degraded areas in one zone with restoration or maintenance of similar habitats in another. Many large-scale donor-funded projects require strict project monitoring and reporting of both project efforts and project outcomes. Given that, for example, the restoration of blue carbon habitats may far exceed project lifespans, it is critically important that a standardised long-term reporting framework be introduced that can measure both short-term efforts (e.g. extent of restored land) and long-term benefits (e.g. increased carbon stock over time or contribution to climate mitigation). Ecosystems accounts applied at the local level provide such a standardised approach. The restoration of estuaries is urgent considering the increasing pressures and loss of ecosystem services such as nursery habitat provision, coastal protection, blue carbon storage and contributions towards climate change mitigation. Restoration of estuarine ecosystem services would contribute billions to the national economy.

Acknowledgements

The DSI/NRF Community of Practice in the Western Indian Ocean: Ocean Accounts Framework (CoP-OAF) and the Council for Scientific and Industrial Research (CSIR), through the Department of Science and Innovation's Parliamentary Grant, are thanked for funding S.T. and L.v.N. The National Research Foundation of South Africa through the support of the DSI/NRF Research Chair in Shallow Water Ecosystems supported J.B.A. and is thanked for funding (UID 84375). Dr Stephen Lamberth (DFFE) is thanked for his input on invasive alien fish, fishing pressure and nursery function in the Swartkops. Dr Monique Nunes (NMU) is thanked for preparing the study site map.

Competing interests

We have no competing interests to declare.

Authors' contributions

S.T.: Formulation of overarching construction and content of manuscript, analysis of available data and information, preparation and creation of manuscript. L.v.N.: Formulation of overarching construction and content of manuscript, analysis of available data and information, significant contribution to preparation and creation of manuscript. J.B.A.: Analysis of available data and information, significant contribution to preparation and creation of manuscript. T.R.: Analysis of available data and information.

References

1. Eden B, Hein L. Towards a consistent approach for ecosystem accounting. Ecol Econ. 2013;90:41–52. https://doi.org/10.1016/j.ecolecon.2013.03.003
2. Hein L, Obst C, Edens B, Remme RP. Progress and challenges in the development of ecosystem accounting as a tool to analyse ecosystem capital. Curr Opin Environ Sustain. 2015;14:86–92. https://doi.org/10.1016/j.cosust.2015.04.002
3. Farrell CA, Coleman L, Kelly-Quinn M, Obst CG, Eigenraam M, Norton D, et al. Applying the System of Environmental-Economic Accounting – Ecosystem Accounting (SEEA-EA) framework at catchment scale to develop ecosystem extent and condition accounts. One Ecosyst. 2021;6, e65582. https://doi.org/10.3897/oneeco.6.e65582
4. Grilli G, Ferrini S, Luisetti T, Turner RK. The role of choice experiments in natural capital accounting approaches: Fast track versus simulated exchange value in the Deben Estuary saltmarshes. J Environ Plan Manag. 2022;65(7):1281–1300. https://doi.org/10.1080/09640568.2021.1957794
5. Natural Capital Coalition. Natural Capital Protocol Principles and Framework [webpage on the Internet]. c2016 [cited 2022 Dec 01]. Available from: https://naturalcapitalcoalition.org/natural-capital-protocol/
6. Hooper T, Börger T, Langmead O, Marcone O, Rees SE, Rendon O, et al. Applying the natural capital approach to decision making for the marine environment. Ecosyst Serv. 2019;38:100947. https://doi.org/10.1016/j.ecoser.2019.100947
7. Global Ocean Accounts Partnership (GOAP). Technical guidance on ocean accounting for sustainable development. Sydney: GOAP; 2019. Available from: https://www.oceanaccounts.org/technical-guidance-on-ocean-accounting-2/
8. United Nations (UN). System of Environmental-Economic Accounting 2012: Central framework. New York: UN; 2014. Available from: https://seea.un.org/content/seea-central-framework
9. United Nations (UN). System of Environmental-Economic Accounting 2012: Experimental ecosystem accounting. New York: UN; 2014. Available from: https://seea.un.org/ecosystem-accounting
10. Chenoweth JJ, Anderson AR, Kumar P, Hunt WF, Chimbwandira SJ, Moore TLC. The interrelationship of green infrastructure and natural capital. Land Use Policy. 2018;75:137–144. https://doi.org/10.1016/j.landusepol.2018.03.021
11. United Nations (UN). Technical recommendations in support of the System of Environmental-Economic Accounting 2012: Experimental ecosystem accounting. New York: UN; 2019. Available from: https://seea.un.org/ecosystem-accounting
12. Wang T, He GS, Zhou QL, Gao JZ, Deng LJ. Designing a framework for marine ecosystem assets accounting. Ocean Coast Manag. 2018;163:92–100. https://doi.org/10.1016/j.ocecoaman.2018.05.019
13. Schenau S, Rietveld H, Bosch D. Natural capital accounts for the Dutch North Sea 2019 [webpage on the Internet]. c2019 [cited 2022 Dec 05]. Available from: https://www.cbs.nl/en-gb/background/2019/51/natural-capital-accounts-for-the-dutch-north-sea-2019

14. Van Niekerk L, Taljaard S, Adams JB, Lamberth SJ, Weerts SP. Experimental ecosystem accounts for South Africa's estuaries: Extent, condition and ecosystem services accounts. Report no. CSIR/SPLA/SECO/IR/2020/0026/A. Stellenbosch: Council for Scientific and Industrial Research; 2020. Available from: https://seea.un.org/content/experimental-ecosystem-accounts-south-africas-estuaries-extent-condition-and-ecosystem

15. Eigenraam M, McCormick F, Contreras Z. Marine and coastal ecosystem accounting: Port Phillip Bay. Report to Commissioner for Environmental Sustainability, Department of Environment, Land, Water and Planning. Melbourne: State of Victoria; 2016. Available from: https://seea.un.org/content/marine-and-coastal-ecosystem-accounting-port-phillip-bay

16. Hinson C, O'Keeffe J, Mijic A, Bryden J, Van Grootveld J, Collins AM. Using natural capital and ecosystem services to facilitate participatory environmental decision making: Results from a systematic map. People Nat. 2022;4:652–668. https://doi.org/10.1002/pan3.10317

17. Dowdall L, Bowe C, Kirby JR. The Natural Capital approach: A framework for the successful recovery of nature in estuaries. Ecol Solut Evid. 2022;3, e12161. https://doi.org/10.1002/2688-8319.12161

18. Burdon D, Potts T, Barnard S, Boyes SJ, Lannin A. Linking natural capital, benefits and beneficiaries: The role of participatory mapping and logic chains for community engagement. Environ Sci Policy. 2022;134:85–99. https://doi.org/10.1016/j.envsci.2022.04.003

19. Cosgrove P. Suffolk marine pioneer: Lessons and recommendations for applying the natural capital approach in England [document on the Internet]. c2020 [cited 2022 Dec 05]. Available from: https://www.suffolkcoastandheaths.org/wp-content/uploads/2021/01/2020-Suffolk-Marine-Pioneer-Applying-Natural-Capital-Approach-Gudiance.pdf

20. Barbier EB. The protective value of estuarine and coastal ecosystem services in a wealth accounting framework. Environ Resour Econ. 2016;64:37–58. https://doi.org/10.1007/s10640-015-9931-z

21. United Nations (UN). The UN Decade on Ecosystem Restoration 2021–2030. Prevent, halt and reverse the degradation of ecosystems worldwide. New York: UN; 2020. Available from: https://www.decadeonrestoration.org/

22. Adams JB, Whitfield AK, Van Niekerk L. A socio-ecological systems approach towards future research for the restoration, conservation and management of southern African estuaries. Afr J Aquat Sci. 2020;45(1–2):231–241. https://doi.org/10.2989/16085914.2020.1751980

23. Van Niekerk L, Adams JB, Lamberth S, MacKay F, Taljaard S, Turpie J, et al. South African National Biodiversity Assessment 2018: Technical report. Volume 3: Estuarine realm. Report number CSIR/SPLA/EM/EXP/2019/0062/A. Stellenbosch: Council for Scientific and Industrial Research; 2019. Available from: http://opus.sanbi.org/handle/20.500.12143/6373

24. Adams JB, Hughes D, James N, Kibble R, Lemley D, Rishworth G, et al. Swartkops Estuary: Present ecological status and future restoration scenarios. Re-evaluating the present ecological status and recommended ecological category, as well as consequences of future climate change, urbanisation and restoration scenarios. Institute for Coastal and Marine Research report no. 48. Gqeberha: Nelson Mandela University; 2021.

25. Raw J, Tsipa V, Banda S, Riddin T, Van Niekerk L, Adams JB. Scoping study: A blue carbon sinks assessment for South Africa. Project 83360258 funded by GIZ: Deutsche Gesellschaft für Internationale Zusammernarbeit GmbH for the Department of Environment Forestry and Fisheries, South Africa. Cape Town: South African Department of Environment Forestry and Fisheries; 2021.

26. South African Department of Water Affairs and Forestry (DWAF). Resource directed measures for protection of water resources: Methodology for the determination of the ecological water requirements for estuaries. Version 2.1. Pretoria: DWAF; 2008.

27. Van Niekerk L, Taljaard S, Adams JB, Lamberth SJ, Huizinga P, Turpie JK, et al. An environmental flow determination method for integrating multiple scale ecohydrological and complex ecosystem processes in estuaries. Sci Total Environ. 2019;56(15):482–494. https://doi.org/10.1016/j.scitotenv.2018.11.276

28. Van Niekerk L, Adams JB, Allen D, Taljaard S, Weerts S, Louw D, et al. Assessing and planning future estuarine resource use: A scenario-based regional scale freshwater allocation approach. Sci Total Environ. 2019:657(20):1000–1013. https://doi.org/10.1016/j.scitotenv.2018.12.033

29. Taljaard S, Lemley DA, Van Niekerk L. A method to quantify water quality change in data limiting estuaries. Estuar Coast Shelf Sci. 2022;272, Art. #107888. https://doi.org/10.1016/j.ecss.2022.107888

30. Turpie JK, Taljaard S, Van Niekerk L, Adams JB, Wooldridge T, Cyrus D, et al. The Estuary Health Index: A standardised metric for use in estuary management and the determination of ecological water requirements. WRC report no. 1930/1/12. Pretoria: Water Research Commission; 2012.

31. Van Niekerk L, Adams JB, Bate GC, Forbes N, Forbes A, Huizinga P, et al. Country-wide assessment of estuary health: An approach for integrating pressures and ecosystem response in a data limited environment. Estuar Coast Shelf Sci. 2013;130:239–251. https://doi.org/10.1016/j.ecss.2013.05.006

32. Van Niekerk L, Taljaard S, Lamberth SJ, Adams JB, Weerts SP, MacKay FC. Disaggregation and assessment of estuarine pressures at a country-level to better inform management and resource protection. Afr J Aquat Sci. 2022;47(2):127–148. https://doi.org/10.2989/16085914.2022.2041388

33. Adams JB, Pretorius L, Snow JC. Deterioration in the water quality of an urbanised estuary with recommendations for improvement. Water SA. 2019;45(1):86–96. https://doi.org/10.4314/wsa.v45i1.10

34. South African Department of Environmental Affairs (DEA). South African water quality guidelines for coastal marine waters. Volume 2: Guidelines for recreational use. Cape Town: DEA; 2012. Available from: https://www.dffe.gov.za/sites/default/files/legislations/water_qualityguidelines_0.pdf

35. Council of European Community (CEC). Council Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC. Published in Official Journal of the European Communities [webpage on the Internet]. c2006 [cited 2022 Dec 05]. Available from: http://europa.eu/legislation_summaries/consumers/consumer_safety/l28007_en.htm\

36. Baird D, Hanekom NM, Grindley JR. Estuaries of the Cape: Part 2: Synopsis of available information on individual systems – Report no. 23 – Swartkops. CSIR report no. 422. Pretoria: Council for Scientific and Industrial Research; 1986.

37. Adams JB, Veldkornet D, Tabot P. Distribution of macrophyte species and habitats in South African estuaries. S Afr J Bot. 2016;107:5–11. https://doi.org/10.1016/j.sajb.2016.08.001

38. Adams JB, Riddin T. State of knowledge: Conservation and management of the Swartkops Estuary. Benguela Current Convention Report September 2020. Institute for Coastal and Marine Research. Gqeberha: Nelson Mandela University; 2020.

39. Bornman TG, Schmidt J, Adams JB, Mfikili AN, Farre RE, Smit AJ. Relative sea-level rise and the potential for subsidence of the Swartkops Estuary intertidal salt marshes, South Africa. S Afr J Bot. 2016;107:91–100. https://doi.org/10.1016/j.sajb.2016.05.003

40. Schmidt J. Influence of predicted sea level rise on the salt marsh of the Swartkops, Kromme and Knysna estuaries [MSc thesis]. Gqeberha: Nelson Mandela University; 2013.

41. Matabane R. Tolerance of salt marsh ecotone species to salinity and inundation [MSc thesis]. Gqeberha: Nelson Mandela University; 2018.

AUTHORS:
Tinotendashe Pori[1,2]
Mduduzi Ndlovu[1,3]
Miles B. Markus[1,4]

AFFILIATIONS:
[1]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]School of Life Sciences, University of Warwick, Coventry, United Kingdom
[3]School of Biology and Environmental Sciences, University of Mpumalanga, Mbombela, South Africa
[4]Wits Research Institute for Malaria, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Mduduzi Ndlovu

EMAIL:
mduduzindlovu@gmail.com

DATES:
Received: 09 Mar. 2021
Revised: 09 Dec. 2022
Accepted: 05 Jan. 2023
Published: 29 Mar. 2023

HOW TO CITE:
Pori T, Ndlovu M, Markus MB. Influence of season and other factors on avian *Trypanosoma* spp. and microfilarial prevalence in the Lowveld, South Africa. S Afr J Sci. 2023;119(3/4), Art. #10358. https://doi.org/10.17159/sajs.2023/10358

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Sydney Moyo

KEYWORDS:
Afrotropical, birds, microfilariae, South Africa, *Trypanosoma*

FUNDING:
South African National Research Foundation; University of the Witwatersrand

Influence of season and other factors on avian *Trypanosoma* spp. and microfilarial prevalence in the Lowveld, South Africa

To comprehend the effects of emerging infectious diseases on both human and animal health, it is necessary to understand the ecology of pathogens that have wildlife reservoirs. In this study, we determined the prevalence of the parasites *Trypanosoma* spp. and filarial nematodes in the bloodstream of birds in and around the Kruger National Park, South Africa, partly to test the hypothesis that season influences parasitaemia. Other factors considered were foraging habits, gregariousness or solitariness, and whether location might facilitate contact between birds and parasite vectors. Microscopy was used to screen stained blood smears prepared from 685 captured birds of 87 species. It was found that 3.9% of the birds were infected with filarial nematodes (as reflected by the presence of microfilariae) and 3.1% with *Trypanosoma* spp. No cases of co-infection with both types of parasite were encountered. Ground-foraging and solitary birds had the highest parasite prevalences compared to other birds. Infections were recorded throughout the year at all six sites. The respective percentages of birds harbouring the two parasite types in the dry season were the same (both 2.3%), whereas microfilariae dominated in the wet season (6.9%) and the prevalence of *Trypanosoma* spp. then was 1.4%. These findings represent new knowledge concerning avian haemoparasite prevalence in an Afrotropical setting – something that has so far been poorly studied.

Significance:

- The determination by microscopy of the prevalence of microfilariae of filarial nematodes (3.9%) and *Trypanosoma* spp. (3.1%) in the peripheral blood of 685 birds of 87 species provides new knowledge on birds in Africa.
- Unexpectedly, ground-foraging and solitary birds had the highest parasite prevalences.
- The possibility of human infection with these two types of avian parasites is considered.

Introduction

Wildlife species serve as reservoirs for the transmission of more than 60% of emerging infectious diseases.[1] We therefore need to know more about the ecology of pathogens that have wildlife reservoirs, particularly at human–wildlife interfaces where susceptibility to disease spill-over can be high. Disease ecologists have thus increased their surveillance of hosts in wildlife populations so as to help predict and manage zoonotic disease outbreaks.[2] Surveillance is also needed for organisms which are not necessarily zoonotic but might be harmful within wildlife populations themselves.

In this study, we investigated the prevalence of *Trypanosoma* spp. and microfilariae of filarial nematodes in the peripheral blood of wild birds. These two organisms, which are referred to here as haemoparasites, also occur widely in vertebrates in general. However, few studies have focused on bird hosts in subtropical African regions. The pathogenicity of avian trypanosomes and filarial helminths for birds may have been underestimated, and birds found dead should accordingly be examined.[3] The two groups of parasites considered here have not been prioritised in past research, yet they can be useful as models for host–parasite interactions.[4] The ecology of their vectors, the drivers of infection, and the exact timing and location of the acquisition of infection within the host lifespan still need to be thoroughly documented.[5,6]

Trypanosomes are common protozoan parasites that are transmitted to vertebrates by arthropods or leeches (apart from the sexually transferred *Trypanosoma equiperdum*). Serious disease is caused by particular *Trypanosoma* species in humans and domestic livestock, but in wild birds, trypanosomes are considered to be far less pathogenic or even non-pathogenic. Avian trypanosomes belong to three different groups and several lineages.[4,7] They have occasionally been found to give rise to clinical signs[8,9] and can have a wide range of bird hosts[10]. Recent molecular evidence indicates that some avian trypanosomes might infect wild mammals.[11] Experimentally, the mammalian *Trypanosoma evansi* has been transmitted to nestling pigeons.[12]

Microfilariae are larval helminth stages in the bloodstream or skin of vertebrates and are produced by adult female filarioid nematodes living in the body. Although by no means a commonly diagnosed zoonosis[13], human infections with filarial nematodes that normally parasitise other vertebrates are nevertheless often reported from around the world. It is likely that filarial parasites of vertebrates in the wild frequently develop (undiagnosed) in humans to some extent.[13] As an example, avian *Pelecitus* has been associated with human eye infection.[14,15]

The vectors involved in filarial nematode transmission are haematophagous arthropods. Microfilariae are sucked up by the vector when it ingests blood from an infected host. The microfilariae then develop into infective larvae in the vector and can in due course be transferred to another host through feeding by the vector. The adult filarial nematodes occur in parts of the body which are not accessible if the host is alive and thus cannot be detected. The generic identity of adult helminths from birds as well as of some avian microfilariae can be determined relatively easily on the basis of their morphological characteristics, but identification of microfilariae to the species level can

be more difficult or impossible.[16,17] In the current state of knowledge, it is in fact frequently not feasible to identify even genera on the basis of the morphology of microfilariae from birds.[18] Some microfilariae found during avian haematozoan surveys will be those of previously undescribed taxa.

It is generally assumed that the prevalence of haemoparasites is determined inter alia by weather changes and other environmental variables. The weather is thought to be particularly influential[19] because the biology of parasite vectors such as biting midges and mosquitoes is strongly influenced by seasonal fluctuations in temperature and rainfall. Vector efficacy is usually higher during summer and suppressed in winter, as temperatures become less favourable.

Haemoparasite infection in the avian host is affected by both intrinsic factors, such as the host's age, physiological and immune status, and extrinsic factors, such as habitat type and the quality and abundance of food.[20] It is also thought to be related to the density of the host population and to the frequency of host exposure to infective parasite stages. Gregarious social behaviour may sometimes increase exposure and, consequently, parasite prevalence and load.[21,22] Foraging height is another contributing factor. As an example, a study by Astudillo et al.[23] investigated the probability of occurrence of haemoparasite infections in birds of different foraging guilds. The middle-upper vertical feeding stratum was dominated by *Trypanosoma* spp., whereas the frequency with which microfilariae were encountered was similar across different foraging guilds.

Parasite prevalence information for wild birds is useful for risk mapping of infections that could be transmitted to domestic poultry or pet birds in surrounding human settlements, although the origin of infections in such hosts can be uncertain.[24,25] Given that avian parasitism is driven by environmental as well as intrinsic factors[26], we hypothesised that trypanosome and microfilarial parasite prevalence is affected by both. We attempted to explore how seasonal changes (wet vs dry), study areas, and land use type (inside vs outside the Park) affected prevalence. We also looked at how avian parasite prevalence differs by host species, and whether it seems to be affected by traits such as foraging height and gregariousness. We hypothesised that haemoparasites might be more prevalent in birds at sites in the southern region of the Park, where the higher annual rainfall would have a direct impact on potential vector establishment.

To obtain a representative indication of the occurrence of haemoparasites across the study area, the primary aim was to collect blood from as many bird species as possible with different habits and in different localities. Our assumption was that birds foraging near water bodies would have a higher haematozoan prevalence because they would spend a significant part of the day and night closer to where some vectors breed, thereby inadvertently increasing their chances of becoming infected. Birds that forage in tree canopies, on the other hand, are expected to have a lower overall prevalence of haematozoan infection because the risk of transmission of some parasites decreases with increasing height.[23]

The focus of the present study was limited to the occurrence of *Trypanosoma* spp. and filarial nematodes, although other avian haematozoa occur in the study area as well.

Methods

Study sites

Birds were sampled at four Lowveld sites in the Kruger National Park: Skukuza (24º99´S, 31º60´E), Satara (24º39´S, 31º77´E), Shangoni (23º45´S, 30º97´E) and Shingwedzi (23º11´S, 31º43´E); and at two sites outside the Park: Phalaborwa (23º94´S, 31º14´E) and Mkhuhlu (24º59´S, 31º14´E). The climate in this whole area is subtropical. Summers are hot and humid with temperatures that can exceed 38 °C, while winters are usually free of frost, the average minimum temperature being around 10 °C. The region receives summer rainfall during the months of September to May, with a rainfall gradient that decreases from the southern to the northern parts of the Park (from 750 mm to 350 mm per annum). Drought commonly occurs in the region.[27]

Sample collection

Birds were caught in 2016 using mist nets and walk-in traps during both dry (April–September) and wet (October–March) seasons at two sampling locations per site. Sampling effort was equivalent across all six sites, with four mist nets and two walk-in traps used at each sampling location. Sampling took place twice at all locations other than the Shangoni and Phalaborwa sites, where difficulties resulted in sampling only once, during the dry season. Trapping of birds was carried out in the morning (06:00–09:00) and late afternoon (16:00–19:00) when temperatures were low, so as to minimise the birds' stress. Mist nets and walk-in traps were in use at the abovementioned times only and were constantly checked for caught birds.

Screening for haemoparasites

The captured birds were individually marked with a metal South African Bird Ringing Unit (SAFRING) band. After blood samples had been taken, birds were released in the same locality where they had been caught. For a separate haemoparasite-related study, body mass and moult status were recorded, as well as standard measurements of tarsus length, and head and culmen lengths. Blood samples were collected by puncturing a blood vessel in the right wing with a sterile 25 G needle and drawing blood into a 75 μL micro-haematocrit capillary tube. Although obtaining blood from the brachial vein like this has been shown to underestimate the prevalence of microfilariae, as compared to blood taken from the pulmonary artery[28], the latter is not possible in live birds. Two blood smears were prepared from each bird sampled. The slides were air dried in the field, and then fixed with absolute methyl alcohol shortly after preparation to preserve the integrity of the cells and increase their rigidity before the slides were transported to the Skukuza Scientific Services Laboratory. Blood smears were stained with 10% Giemsa solution. The duplicate slide, which was not examined, was deposited with Scientific Services for curation purposes.

One blood smear from each bird was screened as follows for the presence of parasites, using an Olympus BX40 compound microscope: an area of the blood film approximately one third from the end of the slide was selected for examination, beginning at low magnification, i.e. x100, followed by x400 and then x1000, using oil immersion. Each slide was screened by moving two fields along the bottom edge, two fields up, two across and then two fields down, etc., and any parasites observed were recorded. The total number of fields covered on each slide at x400 magnification was 20. Photographs of parasites were taken with an Olympus S 30 camera, as well as of any abnormalities, using analySIS getIT software (Version 5.1). Parasites other than microfilariae were identified to genus level and photographs of organisms (Figure 1) were sent to an avian blood parasite expert, Dr Michael Peirce (UK), for confirmation of identification.

Molecular diagnosis was not carried out during this study because of financial constraints.

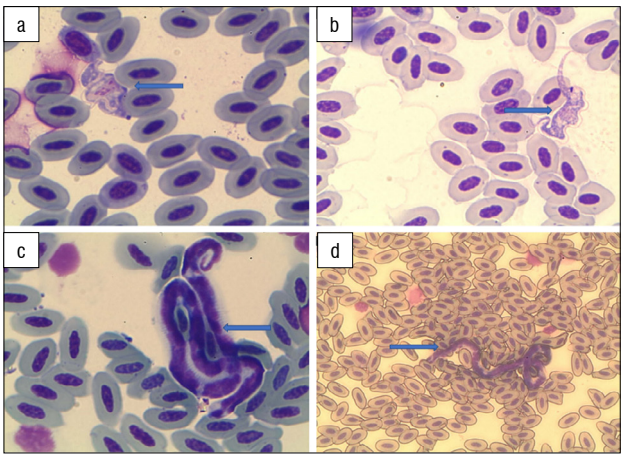


Figure 1: Haemoparasites in avian blood films: *Trypanosoma* sp. in (a) a red-billed hornbill (*Tockus erythrorhynchus*) at Shingwedzi and (b) a yellow-fronted canary (*Serinus mozambicus*) at Skukuza. (c) and (d) Microfilariae in an African mourning dove (*Streptopelia decipiens*) at Shingwedzi.

Data analysis

We first established the presence of the whole range of avian haemoparasites in the birds (regardless of sampling site and seasonal variation). This was then followed by calculating parasite prevalence at each site, taking seasonal variation into account. The prevalence amongst different family groups, foraging guilds and in relation to social behaviour (i.e. solitary or gregarious) was also determined. A Krustal-Wallis test was used to ascertain whether there was any significant variation in infection prevalence between sites. To investigate whether there were any observable seasonal variations in infection, we used a Mann-Whitney U test. A Wilcoxon signed ranks test was also used to analyse the data from the four sites from which there were both wet and dry season samples. Data from Shangoni and Phalaborwa were excluded because for those two sites, there was only dry season information. A Kruskal–Wallis test was also used to test for the significance of infection as a function of foraging guild.

The comparative parasite prevalence was determined for each of the five most dominant family groups of birds (group sample $n \geq 50$). It was decided that a sample of less than 50 would not be adequate for accurate statistical analysis.[29] Infection prevalence was also compared between solitary and gregarious birds using a Wilcoxon signed ranks test. All analyses were tested at the 5% level of significance using the IBM SPSS 23 software.[30]

Ethical clearance

The study was approved by the South African National Parks Board (research permit no: NDLM1262) and the University of the Witwatersrand's Animal Ethics Screening Committee (clearance certificate no: 2015/02/B).

Results

Blood smears were prepared from a total of 685 birds of 87 species. Microfilariae and *Trypanosoma* spp. were detected by microscopy (Figure 1) and 48 birds of 18 species were found to be infected (Table 1). The prevalence of these haematozoa for all sites combined was 3.9% for microfilariae and 3.1% for *Trypanosoma* spp. Greater blue-eared starlings and red-billed hornbills had the highest parasite prevalences (Table 2). No co-infections with both parasite groups were recorded. Detected infections with each parasite were equal (both 2.3%) during the dry season, whereas during the wet season, microfilariae (6.9%) were more often encountered than *Trypanosoma* spp. (1.4%).

Despite equivalent sampling efforts at all six sites and during different seasons, the data were insufficient to determine whether infections in individual host species varied between sites and seasons. This is because the abundance of some bird species differed geographically and seasonally, resulting in various species not being captured at all sites during both seasons. Furthermore, Phalaborwa and Shangoni could only be sampled once, during the dry season. We found no seasonal variations in infection prevalence, at all sites combined, for *Trypanosoma* ($U=8.50$, $p=0.454$) and microfilariae ($U=9.00$, $p=0.495$).

Individual parasite prevalence did not differ significantly between foraging guilds (*Trypanosoma*: $H_3=2.601$, $p=0.457$; microfilariae: $H_3=0.473$, $p=0.925$). In almost all the avian families listed in Table 3, microfilariae were more prevalent than *Trypanosoma* spp. The highest microfilarial prevalence was recorded in Bucerotidae (12.5%) and Columbidae (11.5%), whereas no microfilariae were detected in Hirundinidae (0%). Trypanosomes occurred most frequently in Bucerotidae (7.1%), whilst their lowest prevalence was in Passeridae and Sturnidae (both 1.6%), with no infections seen in Columbidae (Table 3). Solitary birds had the highest number of infections by both filarial nematodes (revealed by the presence of microfilariae in the bloodstream) and *Trypanosoma* spp. (Table 4).

Table 1: Avian species in which *Trypanosoma* spp. and microfilariae were detected

Common name	Species	Haemoparasite present
African jacana	*Actophilornis africanus*	M (1)
African mourning dove	*Streptopelia decipiens*	M (6)
Ashy flycatcher	*Muscicapa caerulescens*	T (1), M (1)
Black-collared barbet	*Lybius torquatus*	M (1)
Blue waxbill	*Uraeginthus angolensis*	T (1)
European swallow	*Hirundo rustica*	T (1)
Greater blue-eared starling	*Lamprotornis chalybaeus*	T (3), M (4)
Green wood-hoopoe	*Phoeniculus purpureus*	M (1)
House sparrow	*Passer domesticus*	T (1), M (1)
Laughing dove	*Streptopelia senegalensis*	T (1), M (2)
Long-billed crombec	*Sylvietta rufescens*	T (1)
Marabou stork	*Leptoptilos crumeniferus*	M (1)
Melba finch	*Pytilia melba*	M (1)
Red-billed hornbill	*Tockus erythrorhynchus*	T (2), M (5)
Southern grey-headed sparrow	*Passer diffusus*	T (2), M (1)
Southern yellow-billed hornbill	*Tockus leucomelas*	T (1), M (2)
Spectacled weaver	*Ploceus ocularis*	T (1)
Yellow-fronted canary	*Serinus mozambicus*	T (6)
TOTALS		**T (21), M (27)**

T, Trypanosoma*; M, microfilaria(e)*

Figures in parentheses are the number of birds infected with that particular parasite out of 685 birds of 87 species from which blood was examined. No cases of co-infection with both types of parasite were found.

Table 2: Birds with the highest infection prevalence, correlated with their habits

Common name	Species	Foraging guild	Social association	Sample size	Haemoparasite present
Greater blue-eared starling	*Lamprotornis chalybaeus*	Ground	Gregarious	170	T (3), M (4)
Red-billed hornbill	*Tockus erythrorhynchus*	Ground	Solitary	35	T (2), M (5)
African mourning dove	*Streptopelia decipiens*	Ground	Solitary	14	M (6)
Yellow-fronted canary	*Serinus mozambicus*	Ground	Solitary	22	T (5)
TOTALS				**241**	**T (10), M (15)**

T, Trypanosoma*; M, microfilaria(e)*

Figures in parentheses represent the number of birds infected by the corresponding parasite. No cases of co-infection with both types of parasite were found.

Table 3: Prevalence of parasites according to Aves families. Only bird families from which more than 50 individuals were examined are included here.

Parasite†	Prevalence (%)				
	Sturnidae	Passeridae	Columbidae	Bucerotidae	Hirundinidae
Trypanosoma	1.60	1.60	0	7.14	1.96
Microfilariae	2.13	1.06	11.48	12.50	0

†No cases of co-infection with both types of parasite were found.

Table 4: Haemoparasite prevalence as a function of social association

Parasite†	Prevalence (%)	
	Gregarious	Solitary
Trypanosoma	1.25	4.38
Microfilariae	2.19	20.00

†No cases of co-infection with both types of parasite were found.

Discussion

The combined parasite prevalence for all sites was low but similar to those that have been recorded elsewhere.[31-33] However, our findings differ from a prevalence of more than 30% in African rainforest birds.[34] It should be noted that there are parasite prevalence determination variables when screening for avian parasites using microscopy. Birds with acute infections may be relatively immobile[35] and hence will not be sampled when mist nets are used for trapping. Another factor is that we recorded infections in birds caught during the day. Other research has demonstrated circadian periodicity for microfilariae, such as a 1.4% prevalence in the American robin *Turdus migratorius* when sampled in daytime and a 11.1% prevalence when roosting robins were sampled at night.[32] During a more recent investigation, the parasitaemia for both trypanosomes and microfilariae was found to peak around midnight[17], perhaps because that is when the particular transmitting vectors concerned are active. The vectors involved in the transmission of avian trypanosomes in general are, however, diverse[36,37], and little is currently known about avian haemoparasite vectors in the Afrotropical region[38].

The similarity in haemoparasite prevalence at the six sites in the Kruger National Park is evidence of the geographically widespread occurrence of avian haematozoa in the region. It was not possible to interpret the patterns of distribution of infections, given that wild birds are highly mobile. As in previous studies[34], we found that the collection site was not an important factor in predicting infection with *Trypanosoma* spp. and filarioid nematodes. Because the sampling sites represented a variety of landscapes (vegetation etc.), our results suggest that the vectors responsible for the transmission of trypanosomes and filarial nematodes are ubiquitous in and around the Park. The altitude at which blood of the various birds was sampled in this study is similar.

Given the ectothermic nature of vectors that are insects, we had assumed that seasonal variations in temperature and rainfall were likely to drive the patterns of prevalence in the environment.[39] However, our findings in the Kruger National Park area do not support this hypothesis. Although temperature is an important variable for predicting the prevalence, distribution and diversity of haemoparasites[40], an explanation for our results could conceivably have been the unusually high winter temperatures during 2016 (mean = 20 °C and maximum = 29.2 °C)[41]. This possibly contributed to the infection of birds during the winter months (a time when the incidence of infection is expected to be low). Alternatively, the infections recorded during that winter might have been long-standing ones. The unusually warm winter could in theory have altered the distribution of vectors or extended their breeding time into the dry season. In fact, global warming is expected to alter wildlife ranges and expand the distribution of vector-borne diseases[19,42], which could result in novel risks[43,44]. Moreover, there is the possibility that under warmer conditions, vectors might facilitate the transmission of parasites throughout the year.

We cannot conclude that foraging in wetlands is a predictor for a higher combined prevalence of haemoparasites of all kinds, as has been found elsewhere[26], because of our small sample size for aquatic bird species (n=12) and detection of infection in only one Egyptian goose *Alopochen aegyptiacus*. Our assumption was that birds foraging near water bodies would have a higher haematozoan prevalence as they would spend a significant part of their day closer to where some vectors breed, thereby inadvertently increasing the chances of becoming infected. By contrast, birds that forage in tree canopies are expected to have a lower overall haematozoan infection prevalence as the risk of transmission of some parasites decreases with vertical distance (elevation).[23] In the present study, birds feeding on the ground had the highest prevalence of infection with *Trypanosoma* spp. and microfilariae. This would seem to be compatible with a finding elsewhere in Africa that birds which follow army ant columns and feed on insects flushed by them have a higher prevalence in their blood of trypanosomes and microfilariae than other birds.[45] It is also interesting to note that the prevalence of both parasites was higher in the present study in solitary birds than in gregarious birds (Table 4). This is contrary to the logic which predicts a higher prevalence in birds that are more social. We therefore suggest that avian social behaviour (gregarious vs solitary) as a prevalence predictor might be heavily dependent on the kind of parasite, and affect bird family groups differently.

The findings here challenge current paradigms in avian parasitology, which predict clear variations in haemoparasite prevalence according to social behaviour and environmental factors, especially location and season. Worldwide, the situation probably varies, however. Our generally low parasite prevalences hindered the application of conventional analysis for elucidating the basis for prevalence patterns. Nevertheless, the results of this study suggest that a host's family group and foraging guild may generally still be a good predictor of prevalence.

Future studies should consider nesting height and nest structure, as infection of nestlings could be important, and because adult birds that are immobile at the nest will no doubt be particularly susceptible to vector bites. When interpreting infection patterns, future research should also take into consideration co-infection with all haematozoa present, because interactions via the immune system could perhaps take place.[46] For instance, the presence of filarial nematodes might predispose to avian malaria.[47]

Acknowledgements

We received funding from the National Research Foundation (South Africa) and the University of the Witwatersrand. We thank Dr Danny Govender, Purvance Shikwambane and Dr Michael Peirce for assistance with identification of parasites on avian blood slides. We are grateful to Alfredo Colina, Miranda Thomas, Damaris Chenoweth, Jackson Seminara and several other undergraduate students from the Organisation for Tropical Studies for their help with fieldwork. SANParks and the Bushbuckridge Local Municipality kindly granted permission for birds to be captured in the Kruger National Park and sites adjacent to the park, respectively.

Competing interests

We have no competing interests to declare.

Authors' contributions

T.P.: Methodology, data collection, data analysis, writing and manuscript revision. M.N.: Conceptualisation, student supervision, methodology, writing and manuscript revision. M.B.M.: Student supervision, methodology, writing and manuscript revision.

References

1. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451:990–993. https://doi.org/10.1038/nature06536
2. Daszak P, Cunningham AA, Hyatt AD. Emerging infectious diseases of wildlife – threats to biodiversity and human health. Science. 2000;287:443–449. https://doi.org/10.1126/science.287.5452.443
3. Himmel T, Harl J, Matt J, Weissenböck H. A citizen science-based survey of avian mortality focusing on haemosporidian infections in wild passerine birds. Malar J. 2021;20:417. https://doi.org/10.1186/s12936-021-03949-y
4. Zídková L, Čepička I, Szabová J, Svobodová M. Biodiversity of avian trypanosomes. Infect Genet Evol. 2012;12:102–112. https://doi.org/10.1016/j.meegid.2011.10.022
5. Valkiūnas G. Avian malaria parasites and other Haemosporidia. Boca Raton, FL: CRC Press; 2005. https://doi.org/10.1201/9780203643792
6. Svobodová M, Zídková L, Čepička I, Oborník M, Lukeš J, Votýpka J. *Sergeia podlipaevi* gen. nov., sp. nov. (Trypanosomatidae, Kinetoplastida), a parasite of biting midges (Ceratopogonidae, Diptera). Int J Syst Evol Microbiol. 2007;57:423–432. https://doi.org/10.1099/ijs.0.64557-0
7. Svobodová M, Volf P, Votýpka J. Trypanosomatids in ornithophilic bloodsucking Diptera. Med Vet Entomol. 2015;29:444–447. https://doi.org/10.1111/mve.12130
8. Larrat S, Dallaire AD, Lair S. Emaciation and larval filarioid infection in boreal owls (*Aegolius funereus*). Avian Pathol. 2012;41:345–349. https://doi.org/10.1080/03079457.2012.688940
9. Soares L, Ellis VA, Ricklefs RE. Co-infections of haemosporidian and trypanosome parasites in a North American songbird. Parasitology. 2016;143:1930–1938. https://doi.org/10.1017/S0031182016001384
10. Galen SC, Borner J, Perkins SL, Weckstein JD. Phylogenomics from transcriptomic "bycatch" clarify the origins and diversity of avian trypanosomes in North America. PLoS ONE. 2020;15, e0240062. https://doi.org/10.1371/journal.pone.0240062
11. Cooper C, Thompson RCA, Botero A, Kristancic A, Peacock C, Kirilak Y, et al. A comparative molecular and 3-dimensional structural investigation into cross-continental and novel avian *Trypanosoma* spp. in Australia. Parasit Vectors. 2017;10:234. https://doi.org/10.1186/s13071-017-2173-x
12. Mandal M, Laha R, Sasmal NK. First report of establishment of *Trypanosoma evansi* infection in pigeon nestlings (*Columba livia*). J Parasitol. 2008;94:1428–1429. https://doi.org/10.1645/GE-1509.1
13. Orihel TC, Eberhard ML. Zoonotic filariasis. Clin Microbiol Rev. 1998;11:366–381. https://doi.org/10.1128/CMR.11.2.366
14. Bain O, Otranto D, Diniz DG, dos Santos JN, de Oliveira NP, de Almeida INF, et al. Human intraocular filariasis caused by *Pelecitus* sp. nematode, Brazil. Emerg Infect Dis. 2011;17:867–869. https://doi.org/10.3201/eid1705.101309
15. Muñoz-García CI, López-Díaz O, Osorio-Sarabia D, Martínez-Hernández F, Villalobos G, Isaak-Delgado AB, et al. New insights into the clinico-histopathological and molecular features of *Pelecitus* (Filarioidea: Onchocercidae) from a raptor bird. Parasitol Res. 2018;117:3319–3325. https://doi.org/10.1007/s00436-018-6009-1
16. Binkienė R, Chagas CRF, Bernotienė R, Valkiūnas G. Molecular and morphological characterization of three new species of avian Onchocercidae (Nematoda) with emphasis on circulating microfilariae. Parasit Vectors. 2021;14:137. https://doi.org/10.1186/s13071-021-04614-8
17. Chagas CRF, Binkienė R, Valkiūnas G. Description and molecular characterization of two species of avian blood parasites, with remarks on circadian rhythms of avian haematozoa infections. Animals. 2021;11:3490. https://doi.org/10.3390/ani11123490
18. Ribeiro PVA, Cury MC, de Melo C. First record of microfilariae in *Antilophia galeata* (Aves: Pipridae). Acta Bras. 2020;4:106–109. https://doi.org/10.22571/2526-4338302
19. Greiman SE, Wilson RE, Sesmundo B, Reakoff J, Sonsthagen SA. Detection of *Splendidofilaria* sp. (Onchocercidae : Splendidofilariinae) microfilaria within Alaskan ground-dwelling birds in the grouse subfamily Tetraoninae using TaqMan probe-based real-time PCR. J Parasitol. 2022;108:192–198. https://doi.org/10.1645/21-101
20. de la Torre GM, Campiáo KM. Bird habitat preferences drive hemoparasite infection in the Neotropical region. Integr Zool. 2021;16:755–768. https://doi.org/10.1111/1749-4877.12515
21. Brown CR, Brown MB. Coloniality in the Cliff Swallow: The effect of group size on social behavior. Chicago, IL: University of Chicago Press; 1996.
22. Hoi H, Darolova A, König C, Kistofík J. The relation between colony size, breeding density and ectoparasite loads of adult European bee-eaters (*Merops apiaster*). Ecoscience. 1998;5:156–163. https://doi.org/10.1080/11956860.1998.11682455
23. Astudillo VG, Hernández SM, Kistler WM, Boone SL, Lipp EK, Shrestha S, et al. Spatial, temporal, molecular, and intraspecific differences of haemoparasite infection and relevant selected physiological parameters of wild birds in Georgia, USA. Intl J Parasitol Parasites Wildl. 2013;2:178–189. https://doi.org/10.1016/j.ijppaw.2013.04.005
24. Huang Y-L, Tsai S-S, Thongchan D, Khatri-Chhetri R, Wu H-Y. Filarial nematode infection in eclectus parrots (*Eclectus roratus*) in Taiwan. Avian Pathol. 2017;46:188–194. https://doi.org/10.1080/03079457.2016.1237014
25. Šupić J, Alić AS, Hasanić M, Goletić S, Duscher GG, Hodžić A, et al. *Eulimdana clava* (Nematoda: Filarioidea) infection in domestic pigeons (*Columba livia domestica*): Molecular characterization and pathological changes. Vet Parasitol. 2018;251:44–49. https://doi.org/10.1016/j.vetpar.2018.01.003
26. Sehgal RNM. Manifold habitat effects on the prevalence and diversity of avian blood parasites. Intl J Parasitol Parasites Wildl. 2015;4:421–430. https://doi.org/10.1016/j.ijppaw.2015.09.001
27. Freitag-Ronaldson S, Venter F. Kruger National Park management plan [document on the Internet]. c2008 [cited 2022 Nov 27]. Available from: www.sanparks.org/assets/docs/conservation/park_man/knp-management-plan1.pdf
28. Holmstad PR, Anwar A, Lezhova T, Skorping A. Standard sampling techniques underestimate prevalence of avian hematozoa in willow ptarmigan (*Lagopus lagopus*). J Wildl Dis. 2003;39:354–358. https://doi.org/10.7589/0090-3558-39.2.354
29. Ganti A. Central limit theorem (CLT): Definition and key characteristics [webpage on the Internet]. c2022 [updated 2022 Jun 28; cited 2022 Nov 27]. Available from: https://www.investopedia.com/terms/c/central_limit_theorem.asp
30. IBM Corporation. IBM SPSS statistics for Windows, version 23.0. New York: IBM Corporation; 2015.
31. Kirkpatrick CE, Smith TB. Blood parasites of birds in Cameroon. J Parasitol. 1988;74:1009–1013. https://doi.org/10.2307/3282224
32. Hamer GL, Anderson TK, Berry GE, Makohon-Moore AP, Crafton JC, Brawn JD, et al. Prevalence of filarioid nematodes and trypanosomes in American robins and house sparrows, Chicago USA. Intl J Parasitol Parasites Wildl. 2013;2:42–49. https://doi.org/10.1016/j.ijppaw.2012.11.005
33. Gimba FI, Ola-Fadunsin SD, Abdullah DA, Konto M, Daudu BB. Prevalence of trypanosome and microfilaria parasites in slaughtered chicken in Jalingo chicken market, Taraba state, North East Nigeria. Alexandria J Vet Sci. 2018;59:148–153. https://doi.org/10.5455/ajvs.867
34. Sehgal RNM, Jones HI, Smith TB. Host specificity and incidence of *Trypanosoma* in some African rainforest birds: A molecular approach. Mol Ecol. 2001;10:2319–2327. https://doi.org/10.1046/j.1365-294x.2001.01339.x
35. Markus MB, Oosthuizen JH. Pathogenicity of *Haemoproteus columbae*. Trans R Soc Trop Med Hyg. 1972;66:186–187. https://doi.org/10.1016/0035-9203(72)90072-7
36. Fialová M, Santolíková A, Brotánková A, Brzoňová J, Svobodová M. Complete life cycle of *Trypanosoma thomasbancrofti*, an avian trypanosome transmitted by culicine mosquitoes. Microorganisms. 2021;9:2101. https://doi.org/10.3390/microorganisms9102101

37. Santolíková A, Brzoňová J, Čepička I, Svobodová M. Avian louse flies and their trypanosomes: New vectors, new lineages and host-parasite associations. Microorganisms. 2022;10:584. https://doi.org/10.3390/microorganisms10030584

38. Hellard E, Cumming GS, Caron A, Coe E, Peters JL. Testing epidemiological functional groups as predictors of avian Haemosporidia patterns in southern Africa. Ecosphere. 2016;7, e01225. https://doi.org/10.1002/ecs2.1225

39. Jones BA, Grace D, Kock R, Alonso S, Rushton J, Said MY, et al. Zoonosis emergence linked to agricultural intensification and environmental change. Proc Natl Acad Sci USA. 2013;110:8399–8404. https://doi.org/10.1073/pnas.1208059110

40. Pérez-Rodríguez A, Fernández-González S, de la Hera I, Pérez-Tris J. Finding the appropriate variables to model the distribution of vector-borne parasites with different environmental preferences: Climate is not enough. Glob Change Biol. 2013;19:3245–3253. https://doi.org/10.1111/gcb.12226

41. Scientific Services. Geographic Information System Laboratory: South African National Parks; 2017.

42. Patz JA, Reisen WK. Immunology, climate change and vector-borne diseases. Trends Immunol. 2001;22:171–172. https://doi.org/10.1016/s1471-4906(01)01867-1

43. Dobson A, Foufopoulos J. 2001. Emerging infectious pathogens of wildlife. Phil Trans R Soc Lond B Biol Sci. 2001;356:1001–1012. https://doi.org/10.1098/rstb.2001.0900

44. McMichael AJ, Campbell-Lendum DH, Corvalán CF, Ebi KL, Githeko AK, Scheraga JD, et al. Climate change and human health: Risks and responses. Geneva: World Health Organization; 2003. https://apps.who.int/iris/handle/10665/42742

45. Peters MK. Ant-following and the prevalence of blood parasites in birds of African rainforests. J Avian Biol. 2010;41:105–110. https://doi.org/10.1111/j.1600-048X.2010.04896.x

46. Markus MB, Fincham JE. Helminthiasis, bystander diseases and vaccines: Analysis of interaction. Trends Parasitol. 2007;23:517–519. https://doi.org/10.1016/j.pt.2007.07.011

47. Clark NJ, Wells K, Dimitrov D, Clegg SM. Co-infections and environmental conditions drive the distributions of blood parasites in wild birds. J Anim Ecol. 2016;85:1461–1470. https://doi.org/10.1111/1365-2656.12578

AUTHORS:
Eman A. Ibrahim[1]
Samah M. El-Sayed[2]
Soha A. Murad[1]
Walid E. Abdallah[3]
Hoda S. El-Sayed[2]

AFFILIATIONS:
[1]Plant Biochemistry Department, National Research Centre, Cairo, Egypt
[2]Dairy Department, National Research Centre, Cairo, Egypt
[3]Chemistry of Medicinal Plants Department, National Research Centre, Cairo, Egypt

CORRESPONDENCE TO:
Samah M. El-Sayed

EMAIL:
samah_mosbah80@yahoo.com

DATES:
Received: 03 Apr. 2022
Revised: 01 Dec. 2022
Accepted: 23 Dec. 2022
Published: 29 Mar. 2023

HOW TO CITE:
Ibrahim EA, El-Sayed SM, Murad SA, Abdallah WE, El-Sayed HS. Evaluation of the antioxidant and antimicrobial activities of fucoxanthin from *Dilophys fasciola* and as a food additive in stirred yoghurt. S Afr J Sci. 2023;119(3/4), Art. #13722. https://doi.org/10.17159/sajs.2023/13722

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☒ All data included
☐ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Pascal Bessong

KEYWORDS:
fucoxanthin, antioxidant, antimicrobial, probiotics, stirred yoghurt, physiochemical analysis, sensory evaluation

FUNDING:
National Research Centre of Egypt (120703)

Evaluation of the antioxidant and antimicrobial activities of fucoxanthin from *Dilophys fasciola* and as a food additive in stirred yoghurt

We investigated the effects of fucoxanthin isolated from the edible macroalga *Dilophys fasciola* on pathogenic microbes and probiotics in vitro and the antioxidant activity of fucoxanthin. The yield concentration of the obtained crude was 50.5% fucoxanthin. We found strong inhibition against Gram-positive *Staphylococcus aureus* and *Listeria monocytogenes*, and lower inhibition against Gram-negative bacteria and fungi. The probiotic strains progressed between 1.2 and 1.67 log cycles at a concentration of 30 μg/mL. The antioxidant activity ranged between 54.76% and 88.36% at a concentration of 40 μg/mL. The 50% lethal dose of algal fucoxanthin was shown to be more than 2511.88 mg/kg. The production of stirred yoghurt incorporated with 20 mg and 30 mg of fucoxanthin per kilogram of milk was evaluated through chemical, microbiological, and sensory analyses during storage for 21 days and compared with control samples. The maximum growth for probiotics (*Bifidobacterium bifidum* and *Lacticaseibacillus casei*) was found on day 14, but more viability counts were detected in the treatment with 30 mg/kg. All treatments were free from mould and yeast counts up to 7 days, and the small numbers of mould, yeast, and psychrotrophic counts appeared first in control samples. Also, the highest dry matter content was observed for treatments with 30 mg/kg. Moreover, the protein, fat, and ash content of all treatments increased with a progressive cold storage period. Greater reductions in the pH were found in treatments than in the control, and were consistent with the development of acidity. During storage, the amount of crude fucoxanthin had no significant impact on the flavour, colour, or appearance scores.

Significance:

- Fucoxanthin is a type of carotenoid that offers many benefits to human health.
- The fucoxanthin of edible *Dilophys fasciola* had a strong antimicrobial effect against Gram-positive bacteria, Gram-negative bacteria, and fungi.
- Stirred yoghurt fortified with crude fucoxanthin had good overall acceptability and the percentage of crude fucoxanthin had no noticeable effects on the flavour, colour, or appearance. Fucoxanthin, therefore, has potential benefit as a food additive.

Introduction

A molecule produced by an organism is simply a natural product that can be classified as either a primary or secondary metabolite. Secondary metabolites are not directly responsible for the growth and reproduction of the organism and are produced from metabolic pathways that play an important role in drug discovery.[1] Due to their unconventional growth requirements, algae have generated a great deal of curiosity. Seaweed (macroalgae) is a general term for a type of plant found in the sea. Seaweeds today fall into three main classes: Chlorophyta, Phaeophyta, and Rhodophyta. Fucoxanthin constitutes an important natural pigment derived from brown seaweed and is one of the major marine xanthophylls which are produced in large quantities only after exposure of the algae to a specific environment.[2] Fucoxanthin is one of the four major carotenoids that are found along with violaxanthin, lutein, and neoxanthin. It accounts for more than 10% of the total carotenoids. Carotene is mostly found in the chloroplasts of brown seaweed but can sometimes be isolated in relatively larger amounts in diatoms.[3]

The bioactive potential of fucoxanthin and the benefits from its use as a powerful nutrient have been consistently reported.[4] The biological activity of the compound has been investigated, in connection with its combined and standalone use, in the treatment of specific conditions such as cancer and obesity.[4] The remarkable biological activities are attributed to fucoxanthin's distinct structure, which differs from that of most common carotenoids such as carotene and astaxanthin.[5] The powerful biological activities of fucoxanthin include antioxidant, anti-inflammatory, anti-cancer, anti-obesity, anti-diabetic, anti-angiogenic, anti-malarial effects, and its protective power on the liver, blood vessels of the brain, bones, skin, and eyes as well.[6] The formulations of brown seaweed have been developed by health supplement companies as raw substances in the form of gels, capsules, and patches.[7] The emphasis could be on maximising commercial sources of fucoxanthin raw materials from established reservoirs. In view of these properties, fucoxanthin has a probable application in numerous industrial areas such as food, cosmetics, and pharmaceutical areas.[8] As a result, using algae to improve fermented products like yoghurt or cheese represents an excellent opportunity to launch a new category of highly nutritional and healthy food products that combine a high number of lactic acid bacteria with rich quantitative and qualitative profiles of natural bioactive metabolites.[9]

Yoghurt is a generally accepted fermented dairy product that provides a sufficient level of proteins, carbohydrates, calcium, and B vitamins. Yoghurt (stirred yoghurt) is produced by two lactic acid strains: *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus*, which are responsible for the fermentation.[10] The addition of different types of probiotic strains (*Bifidobacterium* and lactobacilli) during yoghurt production can be attractive and beneficial to consumers, as some probiotic strains are able to increase the therapeutic value of yoghurt.[11] So,

prebiotic substances or other nutritional materials can enhance the growth of probiotic strains to reach the human colon in a sufficient amount to provide a beneficial effect.

Given this context, in the present work, we aimed to extract the fucoxanthin from brown algae common to Egypt (*Dilophys fasciola*) and to evaluate the antioxidant and antimicrobial activities of the crude fucoxanthin. A second aim was to add crude fucoxanthin to stirred yoghurt as a supplement and then evaluate its impact on probiotic growth, the chemical and sensory properties of the stirred yoghurt, and the ability to prolong the shelf life of its storage to 21 days.

Materials and methods

Materials

Collection of marine algal material

The seaweed was collected from the Marsa Matrouh governorate during April 2020. Algae were identified as Egyptian isolates *Dilophys fasciola.* The sample was thoroughly washed to remove sand, debris, and epiphytes and then washed several times with both tap and distilled water. Finally, samples were spread in a dark room at room temperature for drying.

Microbial strains and culture media

The antimicrobial assay of crude fucoxanthin was investigated against Gram-positive bacteria (*Staphylococcus aureus* ATCC 6538 and *Listeria monocytogenes* ATCC 5980), Gram-negative bacteria (*Escherichia coli* ATCC 8739*, Salmonella typhimirum* ATCC 14028*, Pseudomonas aeruginosa* ATCC 27853), and fungi (*Aspergillus niger* ATCC 10404 and *Aspergillus flavus* ATCC 9643). The bacteria were grown in tryptone soy broth at 35 °C for 24 h; whereas the fungi were grown using potato dextrose broth at 30 °C for 72 h. All tested strains were serially diluted to inoculate 0.5 McFarland (approximate cell density of 10^7 CFU/mL).

The starter cultures of yoghurt and probiotic strains – notably *Streptococcus thermophilus* CH-1*, Lactobacillus bulgaricus* Lb-12 DRI-VAC, *Lactobacillus acidophilus* CH-2, *Lactobacillus gasserii* B-14168, *Lacticaseibacillus paracasei* NRRL B-4564, *Lacticaseibacillus casei* B-1922, *Lacticaseibacillus rhamnosus* NRRL B-442, *Lactiplantibacillus plantarum* B-4496, *Bifidobacterium bifidum* NRRL B-41410 and *Bifidobacterium lactis* BB12 – were activated individually. *Streptococci* were incubated in M17 broth at 37 °C for 24 h aerobically, *Lactobacilli* sp. in MRS broth at 37 °C for 48 h anaerobically, and *Bifidobacterium* sp. in Reinforced Clostridial Agar (Oxoid) with 0.03 g/L aniline blue[12,13] These strains were activated before being used to study the influence of crude fucoxanthin extract on their viability in order to select some of them for the production of functional stirred yoghurt.

Methods

Sample preparation

Initially, the dried seaweed was ground by mixer without generating heat to convert it to a powder and then stored at -18 °C in the dark for further experiments.

Extraction of fucoxanthin

Extraction and purification of fucoxanthin were done according to the method by Haugen et al.[14] with little modification. Dried and ground brown seaweed was extracted by cold acetone-methanol (7:3 v/v) in a 1-L flask.[15] The algae with solvent were homogenised on ice for 10–15 min, and then the mixtures were filtered through a filter paper. The extraction steps were repeated at least three times. The solvent extracts were collected and left at room temperature, under nitrogen, and in the dark until the extract became colourless. The extract was dried at 30–35 °C on a rotary evaporator, and the residue was dissolved in methanol. The methanol of residue was partitioned in a separation funnel between *n*-hexane and 90% (v/v) aqueous methanol three times until the hexane phase had been eliminated. Fucoxanthin from the aqueous phase was moved to diethyl ether. The diethyl ether phase was evaporated to dryness on a rotary evaporator.

Spectrophotometric assay of pigment

Fucoxanthin concentration was determined by spectrophotometer according to Wang et al.[16] Chlorophyll *a* and *b* and carotenoids were assayed according to Yang et al.[17]

Antioxidant activity of crude fucoxanthin from *Dilophys fasciola*

In-vitro DPPH free radical-scavenging assay

The ability of crude fucoxanthin extracted from *Dilophys fasciola* to scavenge DPPH was examined. The determination of free radicals was done according to Ye et al.[18] Different concentrations of samples were added to 3 mL of ethanolic DPPH solution (0.1 mM). After incubation for 30 min in the dark, the absorbance was measured at 517 nm. The percentage of scavenged DPPH was calculated using Equation 1:

$$\%\text{Scavenged} = [1 - (A_{sample} - A_{blank}/A_{control})] \times 100 \qquad \text{Equation 1}$$

In-vitro total antioxidant capacity assay

A volume of 1 mL of crude fucoxanthin from *Dilophys fasciola* and standard ascorbic acid (10–40 μg/mL) were mixed with 3 mL of reagent solution (0.6 M sulfuric acid, 28 mM sodium phosphate, and 4 mM ammonium molybdate). Tubes were incubated at 95 °C for 90 min. After cooling, the absorbance of each sample was measured at 695 nm.[19]

In-vitro hydrogen peroxide scavenging assay

The antioxidant effect of crude fucoxanthin on hydrogen peroxide was determined according to Gulcin et al.[20] Phosphate buffered saline (0.1 M, pH 7.4) was used to prepare hydrogen peroxide solution (10 mM). Volumes of 1 mL of different crude fucoxanthin concentrations (10–40 μg/ mL) were mixed with 2 mL of hydrogen peroxide solution. The mixtures were incubated for 10 min at 37 °C, after which the absorbance was measured by a UV spectrophotometer at 230 nm against a blank (without hydrogen peroxide). The scavenging percentage of hydrogen peroxide was calculated using Equation 2:

$$\%\text{Scavenged} = ((A_0\text{-}A_1) / A_0) \times 100, \qquad \text{Equation 2}$$

Where A is the absorbance of the tested samples and A_0 is the absorbance of the control.

Antimicrobial activity of crude fucoxanthin from *Dilophys fasciola*

The antimicrobial activities of crude fucoxanthin were determined by conventional agar diffusion according to El-Sayed and El-Sayed[21]. The Mueller–Hinton agar medium (oxide) was poured into sterile petri plates and allowed to set. After that, 20 μL of different microbial suspensions (0.5 McFarland) were spread on the agar surface with a sterile swab. The wells were punched into the agar media (6 mm). The crude fucoxanthin was diluted by dimethyl sulfoxide solution (DMSO) to the concentrations of 10, 15, 20, 25 and 30 μg/mL, and the wells were filled with 100 μL of different crude fucoxanthin concentrations. The antimicrobial activity was identified after incubating the plates at 37 °C for 24 h for bacterial strains and at 30 °C for 72 h for fungal strains. For the positive control, gentamycin (100 ug/mL) was used for bacteria while clotrimazole (100 μg/mL) was used for fungi. The appearance of a clear microbial free inhibition zone around each well was then measured iZetres.

The influence of crude fucoxanthin on probiotics and starter culture viability

The effect of dye concentrations on the viability of lactic acid strains was evaluated according to the method of El-Sayed and El-Sayed[22]. Different concentrations of fucoxanthin from 15 μg/mL to 30 μg/mL were added to 10 mL of sterilised media broth (M17 for streptococci, MRS for lactobacilli and Reinforced Clostridial Agar with 0.03 g/L aniline blue for *Bifidobacterium*) individually, likewise, using media broth only as the controls. Each test tube was inoculated with ~ x10^6 CFU of a different test strain and was incubated at 37 °C for 24 h. The counts of the strains were evaluated using the pour plate method with selective

agar media and they were then incubated at 37 °C for 48 h. The counts were expressed as log CFU/mL. All the media used in this study were provided by Oxoid Ltd., an agency in England.

Determination of LD_{50} value by graphical method

The LD_{50} value of the extract was determined using the graphical method according to Miller and Tainter[23]. To determine the LD_{50} value, Wister albino rats (24 rats) of the same mass were divided into four groups (n=6). Each group was housed separately. Animal groups were orally administered the extract at different doses (1000, 1500, 2000, and 2500 mg/kg). The number of dead animals was recorded after 24 h. The percentages of death were corrected using the formulae: correction formula for 0% dead group = 100(0.25/n), and correction formula for 100% dead group = 100(n0.25/n), where 'n' represents the number of animals in the group. After correction, the percentages were converted into probits and the values were plotted against log dose. The LD_{50} value was determined by finding the dose that was intersected by probit 5.[24]

Production of stirred yoghurt fortified with crude fucoxanthin

Fresh cow's milk was heated at 80±1 °C for 15 min, then cooled and adjusted to 42 °C according to El-Sayed et al.[25] The milk was inoculated with starting cultures (*L. bulgaricus* and *S. thermophilus*) with ratio 1:1%, (w/w) and probiotic strains of *B. bifidum* and *L. casei* with ratio 1:1% (w/w). The inoculated milk was then divided into three portions; the first was saved as the control, the second was complemented with 20 mg crude fucoxanthin per kilogram of milk, and the third was complemented with 30 mg crude fucoxanthin per kilogram of milk. All portions were incubated at 42 °C for 2–4 h until coagulation. After incubation the result yoghurt treatments were stirred with a glass rod and separately transferred into plastic cups (50 mL). The produced stirred yoghurt treatments were stored at 5±1 °C for 21 days and analysed during the storage period at 7-day intervals.

Microbiological evaluation of stirred yoghurt

The pour plate method was used to determine the microbiological activities of stirred yoghurt using decimal dilutions by sterile saline solution (NaCl, 0.9% w/v). The counts of *B. bifidum* were evaluated using Reinforced Clostridial Agar with 0.03 g/L aniline; the plates were incubated at 37 °C for 72 h anaerobically.[12] The count of *L. casei* was enumerated using MRS agar medium supplemented with 0.15% bile salts and 0.05% cellobiose as a carbon source[26] and the plates were incubated at 37 °C for 48 h. *S. thermophilus* was enumerated using M17 agar and the plates were incubated at 37 °C for 48 h aerobically.[27] The count of *L. bulgaricus* was enumerated using MRS agar with 10% sorbitol and the plates were incubated at 37 ºC for 48 h anaerobically.[12] The plate count agar was used to count psychrotrophic bacteria and was incubated at 5 °C for 7 days.[28] The counts of mould and yeast were detected using Chloramphenicol Rose Bengal medium[29] and the plates were incubated at 25 ºC for 4 days aerobically. All the microbiological populations were counted by log CFU/mL values.

Physicochemical analysis of stirred yoghurt

Stirred yoghurt samples were analysed for dry matter, titratable acidity, fat and protein as previously described.[30] The pH measurements of stirred yoghurt were carried out with a digital pH meter equipped with a combined electrode (Hanna Instruments, Italy).

Sensory evaluation of stirred yoghurt

The sensory evaluation of different stirred yoghurt treatments was undertaken by 15 members in the Dairy Department, National Research Centre. The yoghurts were evaluated according to colour and appearance, body and texture, flavour, and overall acceptability.[31] Stirred yoghurt samples were evaluated at fresh state (Day 0) and during storage at refrigerator temperature until 21 days. The assessment degree scale ranged from 1 to 5 with 5 being the most acceptable degree.

Statistical analysis

Descriptive statistics and significant differences were tested using Statistical Analysis System, SAS Institute Inc., Cary, NC. Differences were considered statistically significant when p-values were equal to or less than 0.05.

Ethical clearance

Ethical approval for this study was granted by the Medical Research Ethics Committee of the National Research Centre of Egypt (reference 8521212021).

Results and discussion

Crude fucoxanthin

The yield of crude fucoxanthin from *Dilophys fasciola* was 2.1±0.01 g/100 g dry weight. The crude extract of fucoxanthin has other pigments such as total chlorophyll (36.3%) and carotenoid (8.71%) as shown in Table 1. Yeni et al.[32] found that the yield of fucoxanthin from *S. polycystum, S. granuliferum* and *Sirophysalis trinodis* was 1.24±0.01 g/100 g, 1.43 ±0.01 g/100 g, and 1.48±0.01 g/100 g, respectively. Pigments from brown algae extracted by DMSO and acetone were chlorophyll *a* (7.1%), fucoxanthin (56.5%), and chlorophyll *c* (65.5%) in *Laminaria saccharina* while *Sargassum muticum* pigments were chlorophyll *a* (10.7%) and fucoxanthin (78.4%).[33]

Table 1: Crude fucoxanthin and other pigments from *Dilophys fasciola*

Crude fucoxanthin	**%**
Yield of crude fucoxanthin	2.1
Concentration of fucoxanthin	50.5
Other pigments in crude fucoxanthin	
Total chlorophyll ($a+b$)	36.3
Carotenoid	8.71

Antioxidant activity of crude fucoxanthin from *Dilophys fasciola*

The antioxidant properties of crude fucoxanthin were determined using three different methods: $DPPH^{\cdot}$, H_2O_2 and total antioxidant activity assays.

DPPH scavenging activity

Free radical scavenging of the tested extract was measured by $DPPH^{\cdot}$ assay and the results are shown in Table 2. The results show significant differences in antioxidant activity. The highest activity of crude fucoxanthin was 54.76% at 40 μg/mL followed by 42.13% at 30 μg/mL compared to synthetic antioxidant vitamin C.

H_2O_2 scavenging activity

The H_2O_2 method was used to determine the antioxidant activity of crude fucoxanthin. The results are shown in Table 2. Crude fucoxanthin reduced the level of pro-oxidants against hydrogen peroxide. The antioxidant effect of crude fucoxanthin against hydrogen peroxide ranged from 29.48% to 88.36% compared to controls. This antioxidant activity of crude fucoxanthin against hydrogen peroxide is higher than that of the control.

Total antioxidant activity

The total antioxidant activity of crude fucoxanthin followed the reduction of phosphomolybdic acid (Table 2). The results obtained show that the antioxidant activity increased relative to the concentration of crude fucoxanthin. Brown algae have high antioxidant activity. In-vitro antioxidant chemical methods confirm that the crude extracts, fractions, and pure components of brown seaweed are comparatively similar to synthetic antioxidants.[34] The carotenoid and fucoxanthin present in brown algae possess a high antioxidant activity and they are widely used in applications in various nutraceutical and pharmaceutical arenas as commercially important bioactive compounds.[35,36]

Table 2: Antioxidant activity of crude fucoxanthin from *Dilophys fasciola* against DPPH, H_2O_2 and phosphomolybdic acid

	Inhibition %		
Concentration of fucoxanthin (μg)	**DPPH**	**H_2O_2**	**Phosphomolybdic acid**
10	25.26^{f}	29.48^{f}	38.80^{f}
20	29.21^{e}	42.58^{d}	51.10^{e}
30	42.13^{d}	70.71^{b}	54.88^{d}
40	54.76^{c}	88.36^{a}	65.05^{b}
Control (vitamin C, μg)			
10	16.55^{g}	10.04^{h}	52.92^{de}
20	29.94^{e}	20.54^{g}	60.01^{c}
30	59.87^{b}	37.54^{e}	68.67^{a}
40	69.10^{a}	47.63^{c}	70.21^{a}
LSD	2.40	1.31	2.33

Data are expressed as a mean of three replicates. Means with the same lowercase letter superscripts indicate insignificant difference between samples (p>0.05).

Antimicrobial activity of crude fucoxanthin from Dilophys fasciola

Table 3 shows the results of the well diffusion test of crude fucoxanthin against some bacteria and fungi. Crude fucoxanthin had antimicrobial activity in all of the examined strains starting with a concentration of 15 μg/mL; at this concentration, the clear zone ranged between 3 mm and 8 mm and was enhanced with increased concentration. At a concentration of 20 μg/mL, the clear zone ranged between 8 mm and 15 mm. These zones increased to record 10–17 mm at a concentration of 25 μg/mL and recorded 17–27 mm at a concentration of 30 μL/mL. Compared with the positive control, where gentamycin was used as the antibacterial agent, the zones of inhibition ranged from 14 mm to 19 mm against all the bacterial strains. Where clotrimazole was used as the antifungal agent, the zones of inhibition were 19 mm and 20 mm against *A. niger* and *A. flavus*, respectively. On the other hand, there were not any detected clear zones against tested bacterial strains at a 10 μL/mL concentration of crude fucoxanthin. Gram-positive *S. aureus* was the most sensitive strain to crude fucoxanthin at all concentrations, followed by *L. monocytogens*. However, our results show little difference between tested strains at the same concentration. The results indicate that crude fucoxanthin had good antimicrobial activity when compared with the positive control, and this makes fucoxanthin suitable for use as a preservative in foods to maintain safety during food storage (as stated later when applying the extract to stirred yoghurt). Our results are in agreement with the findings of Karpiński and Adamczak[37] who found that fucoxanthin had a stronger influence on Gram-positive than Gram-negative bacteria. Mean zones of growth inhibition for Gram-positive bacteria ranged between 9.0 mm and 12.2 mm, while in the case of Gram-negative bacteria, they ranged from 7.2 mm to 10.2 mm.

The influence of crude fucoxanthin on probiotics and starter culture viability

Primarily, it was important to study the effects of crude fucoxanthin on the viability of probiotics and starter cultures in vitro before integration into stirred yoghurt. The viability of different probiotics and other starter cultures was significantly enhanced in the presence of crude fucoxanthin concentrations, as shown in Figure 1. At the fucoxanthin concentration of 10 μg/mL, the viable bacterial counts were detected in the same log cycles as for the control (inoculated media without fucoxanthin), where the values ranged from 7.15 to 7.90 log CFU/mL. These bacterial counts were improved at a concentration of crude fucoxanthin of 20 μg/mL, where the counts were increased in range from 0.31 to 0.97 log cycles.

Further improvements in bacterial counts were found at the concentrations 25 μg/mL and 30 μg/mL, where the improvement was recorded in the ranges from 1.00 to 1.42 and from 1.2 to 1.67 log cycles at the concentrations 25 μg/mL and 30 μg/mL, respectively. Additionally, the *Bifidobacterium* strains and *L. casei* were the species with relatively more counts in the presence of different crude fucoxanthin concentrations. Liu et al.[38] confirmed the effect of fucoxanthin on probiotics and indicated that fucoxanthin encourages the growth of intestinal microbes at low concentrations. Conversely, the high concentration of fucoxanthin promoted the growth of *Lactobacillus* and inhibited the deployment of *E. coli*. Finally, microalgae and other derivatives have been shown to stimulate the growth of lactic acid bacteria in fermented milk, as reported by many researchers.[39-41]

LD_{50} value of algal fucoxanthin

The LD_{50} value of algal fucoxanthin in rats was calculated as 2511.88 mg/kg (Table 4). Acute lethal toxicity tests showed that algal fucoxanthin is very safe for humans. Our LD_{50} results were in good agreement with those of Iio et al.[42] who found that the 50% lethal dose of microalgal fucoxanthin oil was more than 2000 mg/kg in experimental rats. The authors did not observe any mortalities after a single oral dose of 2000 mg/kg.

Table 3: Antimicrobial activity of crude fucoxanthin from *Dilophys fasciola*

Test strains	10 μg/mL	15 μg/mL	20 μg/mL	25 μg/mL	30 μg/mL	Gentamycin 100 μg/mL	Clotrimazole 100 μg/mL
	Diameter of inhibition zone (mm)						
Staphylococcus aureus	ND	8Da	15^{Ca}	17^{Ba}	27^{Aa}	15^{Cc}	NA
Listeria monocytogenes	ND	7Da	13^{Cb}	17^{Ba}	24^{Ab}	17^{ab}	NA
Escherichia coli	ND	4^{Cb}	7^{Cd}	10^{Bc}	17^{Ae}	17^{Ab}	NA
Salmonella typhimirum	ND	3^{Cc}	11^{Cc}	15^{Bb}	19^{Ad}	19^{Aa}	NA
Pseudomonas aeruginosa	ND	5^{Db}	12^{Cb}	16^{Ba}	21^{Ac}	14^{Bd}	NA
Aspergillus niger	ND	5^{Db}	10^{Cc}	15^{Bb}	20^{Ac}	n/a	19^{Aa}
Aspergillus flavus	ND	5^{Db}	8^{Cd}	14^{Bb}	19^{Ad}	n/a	21^{Aa}

Data are expressed as a mean of three replicates. Means with the same uppercase letter superscripts indicate insignificant difference between rows (p>0.05), and means with the same lowercase letter superscripts indicate insignificant difference between columns (p>0.05).

ND, not detectable inhibition zone; NA, not applicable

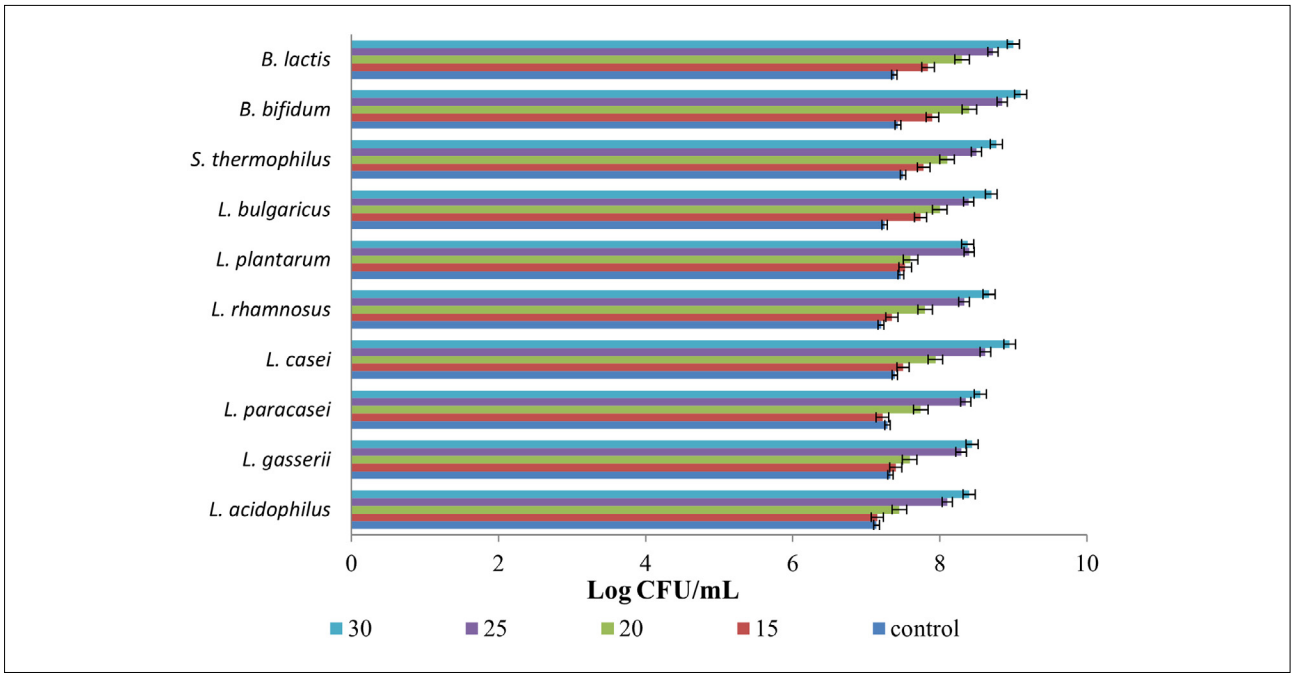


Figure 1: The influence of crude fucoxanthin (μg/mL medium) on probiotics and starter culture viability. Data are expressed as a mean of three replicates ± SD.

Microbiological evaluation of stirred yoghurt

Probiotic strain counts in stirred yoghurt treatments

The effect of crude fucoxanthin on the activity of the probiotics (*B. bifidum* and *L. casei*) in stirred yoghurt during storage is shown in Figure 2. It is clear from the figure that the counts of the two probiotic strains significantly increased over time and the maximum growth was found at 14 days of storage in all treatments. In the case of *B. bifidum*, a relatively higher viable count (9.30 log CFU/mL) was observed in T2 (stirred yoghurt with 30 mg crude fucoxanthin per kilogram milk), followed by T1 (stirred yoghurt with 20 mg crude fucoxanthin per kilogram milk) for which the count was 8.80 log CFU/mL. The lowest count (7.94 log CFU/mL) was observed in the case of the control at 14 days. The same trend was observed in the case of *L. casei*, in which T2 had the most viable count (9.98 log CFU/mL), followed by T1 (9.22 log CFU/mL), and the control showed the lowest count (8.35 log CFU/mL) after 14 days. A small drop in the viable counts of probiotic strains was found on day 21, likely related to the development of more acidity during the storage period. Notably, the counts of all probiotic strains in the stirred yoghurt treatments were higher than the recommended limit of 7.00 log CFU/mL. These results were confirmed by Liu et al.[38], Hosseini et al.[39], Beheshtipour et al.[40] and Abu-Ghannam and Shannon[43] who indicated different lactic acid and probiotic strains progress in the presence of microalgae and their derivatives.

Table 4: Determination of LD_{50} value of crude fucoxanthin

Rat group	Dose (mg/kg)	Log dose	Death	Corrected death %	Probit
1	1000	3.00	0	4.16	3.25
2	1500	3.17	1	16	4.01
3	2000	3.30	2	33	4.56
4	2500	3.39	3	50	5.00

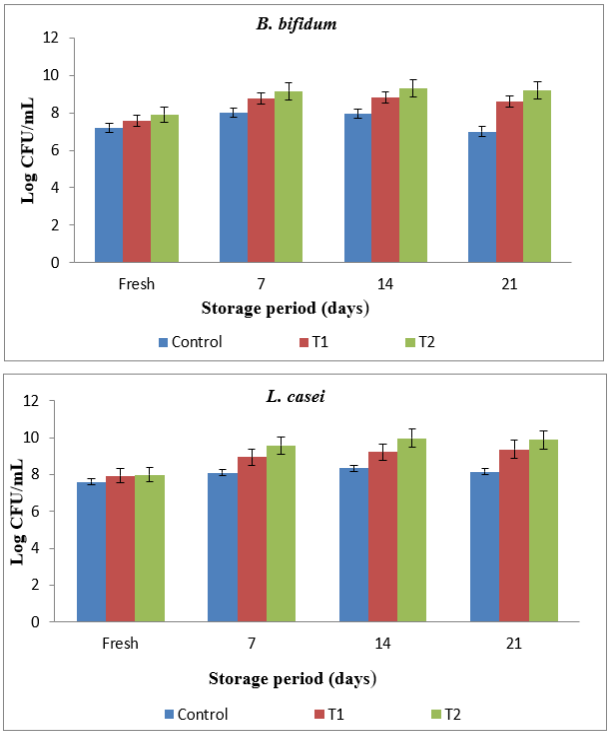


Figure 2: Counts of probiotic strains in stirred yoghurt treatments during storage. Control: plain yoghurt; T1: yoghurt containing 20 mg crude fucoxanthin / kg milk; T2: yoghurt containing 30 mg crude fucoxanthin / kg milk. Data are expressed as a mean of three replicates ± SD.

Starter culture counts in stirred yoghurt treatments

The counts of *L. bulgaricus* and *S. thermophilus* (yoghurt starter cultures) also showed the same trend, that is, the bacterial count was higher in the case of T2 (9.5 and 9.40 log CFU/mL), followed by T1 (9.00

and 8.89 log CFU/mL), on day 14 (Figure 3). The moderately lower count was detected in the stirred yoghurt controls (7.81 and 8.50 log CFU/mL, respectively). But after 21 days, the counts of starter cultures relatively decreased due to the development of acidity.[26] Mok et al.[44] found that milk products can be used as a food matrix for fucoxanthin application and that protein content in milk is an important factor for fucoxanthin stability.

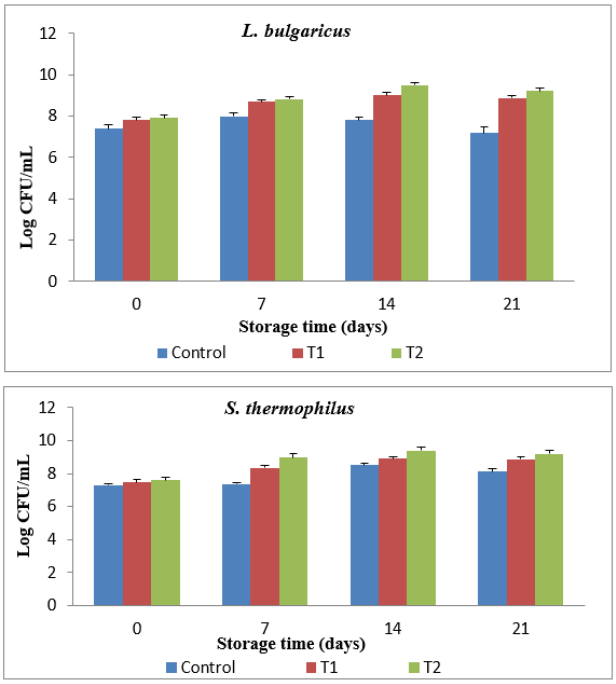


Figure 3: Counts of starter cultures in stirred yoghurt treatments during storage period. Control: plain yoghurt; T1: yoghurt containing 20 mg crude fucoxanthin / kg milk; T2: yoghurt containing 30 mg crude fucoxanthin / kg milk. Data are expressed as a mean of three replicates ± SD.

Detection of other microbial counts in treatments

The shelf life of the stirred yoghurt treatments was assessed through detection of moulds and yeasts and psychrotrophic bacteria. Table 5 indicates that all treatments were free from mould and yeast up to 7 days storage at 4 °C. On day 14, a small number of moulds and yeasts was detected in the control (1.20 log CFU/mL) and T1 (1.00 log CFU/mL) only. Additionally, the low count of moulds and yeasts was observed for T2 (0.89 log CFU/mL) on day 21, and the count was increased in the control to 2.18 log CFU/mL and in T1 to 1.23 log CFU/mL on day 21. Likewise, in the case of psychrotrophic bacterial counts, the count was detected firstly in the control in fresh samples for which a count of 2.00 log CFU/mL was recorded. Afterwards, the psychrotrophic bacteria were detected for T1 and T2 on day 7, where counts of 1.60 and 1.18 log CFU/mL for T1 and T2, respectively, were recorded. These counts significantly increased with the storage period in all treatments, but higher counts were recorded for the control (4.10 log CFU/mL) at the end of storage. Therefore, the lower counts of moulds, yeasts, and psychrotrophic bacteria in treatments than in controls could indicate a preserving activity of crude fucoxanthin as established by the antimicrobial activity mentioned before. The obtained result is in agreement with the findings of other studies that detected the antimicrobial activity of fucoxanthin.[9,37,45]

Physicochemical analysis of stirred yoghurt

Table 6 shows that the dry matter content of stirred yoghurt was significantly ($p<0.05$) affected by crude fucoxanthin fortification and progression of the cold storage period. The increase in the dry matter content was more pronounced in the stirred yoghurt fortified with both levels of crude fucoxanthin (T1 and T2) than in the control samples (without the addition of crude fucoxanthin). After 21 days of storage, the highest dry matter content was found in stirred yoghurt fortified with 30 mg crude fucoxanthin (15.79%) while the lowest was in the control (15.09%).

Table 5: Counts of mould, yeast and psychrotrophic bacteria counts in stirred yoghurt treatments during the storage period

Storage (days)	Mould and yeast counts			Psychrotrophic bacterial counts		
	Control	T1	T2	Control	T1	T2
Fresh	ND	N.D	ND	2.00C	ND	ND
7	ND	N.D	ND	3.70^{Ba}	1.60^{Bb}	1.18^{Bc}
14	1.20^{B}	1.00^{B}	ND	4.22^{Aa}	2.19^{Ab}	2.00^{Ac}
21	2.18^{A}	1.23^{A}	0.89^{A}	4.10^{Aa}	2.00^{Ab}	1.94^{Ab}

Data are expressed as a mean of three replicates. Means with the same uppercase letter superscripts indicate insignificant difference between columns (p>0.05), and means with the same lowercase letter superscripts indicate insignificant difference between rows (p>0.05).

Control: plain yoghurt; T1: yoghurt containing 20 mg crude fucoxanthin / kg milk; T2: yoghurt containing 30 mg crude fucoxanthin / kg milk.

Table 6: Chemical properties of stirred yoghurt during cold storage for 21 days

Parameter	Storage (days)	Treatments		
		Control	T1	T2
Dry matter (%)	Fresh	14.50^{Bc}	14.89^{Bb}	15.51^{Ca}
	7	14.66^{Bc}	15.00^{Bb}	15.63^{Ba}
	14	14.91^{Ac}	15.49^{Bb}	15.68^{Ba}
	21	15.09^{Ac}	15.58^{Ab}	15.79^{Aa}
Protein (%)	Fresh	4.38^{Da}	4.32^{Db}	4.28^{Dc}
	7	4.44^{Ca}	4.39^{Cb}	4.33^{Cc}
	14	4.51^{Ba}	4.45^{Bb}	4.39^{Bc}
	21	4.58^{Aa}	4.50^{Ab}	4.46^{Ac}
Fat (%)	Fresh	3.8^{Ba}	3.7^{Ba}	3.6^{Bb}
	7	3.8^{Ba}	3.8^{Aa}	3.6^{Bb}
	14	4.00^{Aa}	3.8^{Ab}	3.7^{Ab}
	21	4.00^{Aa}	3.9^{Aa}	3.8^{Ab}
Ash (%)	Fresh	1.14^{Ba}	1.11^{Ba}	1.08^{Bb}
	7	1.15^{Ba}	1.12^{Ba}	1.10^{Bb}
	14	1.19^{Aa}	1.17^{Aa}	1.14^{Ab}
	21	1.22^{Aa}	1.19^{Aa}	1.17^{Aa}

Data are expressed as a mean of three replicates. Means with the same uppercase letter superscripts indicate insignificant difference (p>0.05) between columns, and means with the same lowercase letter superscripts indicate insignificant difference between rows (p>0.05).

Control: plain yoghurt; T1: yoghurt containing 20 mg crude fucoxanthin / kg milk; T2: yoghurt containing 30 mg crude fucoxanthin / kg milk

The protein, fat, and ash contents of stirred yoghurt from all treatments including control samples increased significantly ($p<0.05$) with the period of cold storage, reaching maximum values after 21 days of storage. These increases can be attributed principally to the increase in the dry matter or water losses during storage. These changes in the chemical analysis of stirred yoghurt during storage are in agreement with those reported by El-Shibiny et al.[46] However, the protein, fat, and

ash contents of stirred yoghurt were significantly lower in the fortified stirred yoghurt compared to control samples, and this decrease was observed in proportion with an increase in crude fucoxanthin from 20 to 30 mg per kilogram milk.

Changes in the titratable acidity and pH values of stirred yoghurt

The titratable acidity values of stirred yoghurt increased gradually, whereas the pH measurements decreased significantly with progressive cold storage in control and stirred yoghurt treatments compared to other treatments (Figure 4). In both fortified yoghurts, T1 and T2, the development of acidity was slightly faster than in control yoghurt samples during fermentation and storage. Moreover, it was noticed that the progress of acidity increased with an increased fortification with crude fucoxanthin from 20 to 30 mg per kilogram milk; the control stirred yoghurt had the lowest titratable acidity value compared to the other treatments. Mise et al.[47] indicated in their research that inclusion of fucoxanthin with food that contains probiotic cultures may increase the acidity in this food, which may explain the present results.

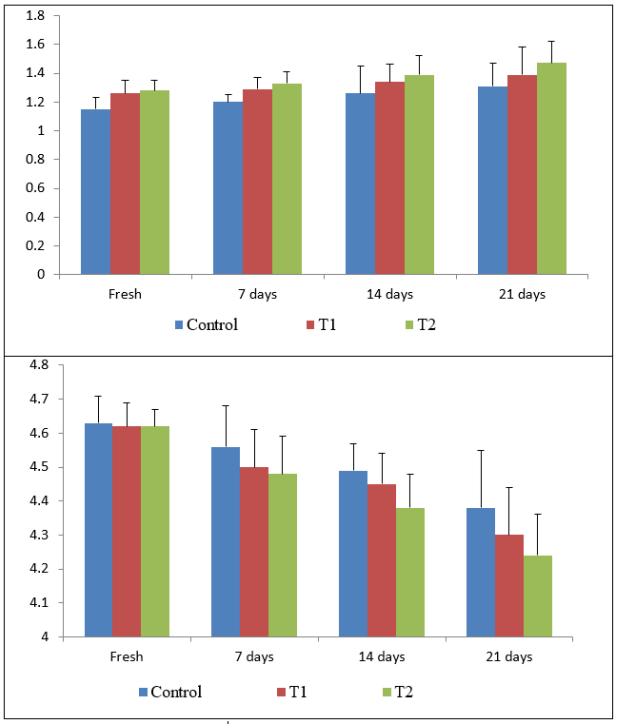


Figure 4: Changes in the titratable acidity and pH values of stirred yoghurt fortified with crude fucoxanthin during 21 days of cold storage. Control: plain yoghurt; T1: yoghurt containing 20 mg crude fucoxanthin / kg milk; T2, yoghurt containing 30 mg crude fucoxanthin / kg milk. Data are expressed as a mean of three replicates ± SD.

Sensory evaluation of stirred yoghurt

Figure 5 displays the overall acceptability scores of the control stirred yoghurt and stirred yoghurt fortified with crude fucoxanthin during 21 days of storage at refrigerator temperature. The overall acceptability of stirred yoghurt fortified with crude fucoxanthin at the concentrations of 20 and 30 mg per kilogram milk (T1 and T2, respectively) was approximately nearer to the control stirred yoghurt when fresh and after 7 days of cold storage. After 14 days of storage, the overall acceptability of stirred yoghurt in all treatments, including control treatments, slightly decreased and this can be attributed to the syneresis effect or the development of acidity in the stirred yoghurt treatments.[48-50] The lowest total overall acceptability scores were given to the control stirred yoghurts after storage for 21 days. No chalkiness was detected in stirred yoghurts fortified with different ratios of crude fucoxanthin. Moreover, both storage time and percentage of crude fucoxanthin had no noticeable effects on the flavour as well as on the colour and appearance scores. The same trend was found by Prabhasankar et al.[51] who reported the successful addition of fucoxanthin into semolina pasta and found no significant or discernible organoleptic differences between the control and the fortified pasta. The body and texture of all stirred yoghurt treatments were affected by the storage time, and these effects were amplified with advanced cold storage period, mainly due to syneresis and water loss during storage.

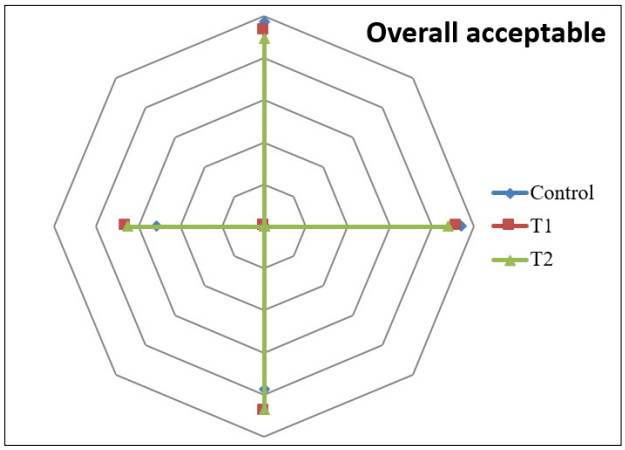


Figure 5: The overall acceptability of yoghurt fortified with crude fucoxanthin during 21 days of cold storage. Control: plain yoghurt; T1: yoghurt containing 20 mg crude fucoxanthin / kg milk; T2: yoghurt containing 30 mg crude fucoxanthin / kg milk. Data are expressed as a mean of three replicates ± SD.

Conclusion

The fucoxanthin of edible *Dilophys fasciola* has a strong antimicrobial effect against Gram-positive bacteria, Gram-negative bacteria, and fungi. Furthermore, in the presence of 30 g/mL crude fucoxanthin, the probiotic strains progressed from 1.2 to 1.67 log cycles. Algal fucoxanthin is a safer pigment and a strong antioxidant functional food. The stirred yoghurt containing crude fucoxanthin that was produced was free from mould and yeast for up to 21 days. The overall acceptability of stirred yoghurt fortified with crude fucoxanthin at concentrations of 20 and 30 mg/kg was very near to that of the control stirred yoghurt and the percentage of crude fucoxanthin had no noticeable effects on the flavour, colour, or appearance scores during storage.

Acknowledgements

We thank the National Research Centre of Egypt (Research point E 120703) for financial support.

Competing interests

We have no competing interests to declare.

Authors' contributions

All authors were responsible for the isolation of fucoxanthin from the edible *Dilophys fasciola*; stirred yoghurt preparation; methodology, data collection and analysis; characterisation; and writing, reviewing and editing. All authors approved the final manuscript.

References

1. Fraenkel GS. The raison d'etre of secondary plant substances: These odd chemicals arose as a means of protecting plants from insects and now guide insects to food. Science. 1959;129:1466–1470. https://doi.org/10.1126/science.129.3361.1466

2. Eonseon J, Polle JEW, Lee HK, Hyun SM, Chang M. Xanthophylls in microalgae: From biosynthesis to biotechnological mass production and application. J Microbiol Biotechnol. 2003;13(2):165–174.

3. Peng J, Yuan J, Wang J. Fucoxanthin, a marine carotenoid present in brown seaweeds and diatoms: Metabolism and bioactivities relevant to human health. Mar Drugs. 2011;9:1806–1828. https://doi.org/10.3390/md9101806

4. Maeda H, Tsukui T, Sashima T, Hosokawa M, Miyashita K. Seaweed carotenoid, fucoxanthin, as a multi-functional nutrient. Asia Pac J Clin Nutr. 2008;71:196–199.

5. Sachindra NM, Sato E, Maeda H, Hosokawa M, Niwano Y, Kohno M, et al. Radical scavenging and singlet oxygen quenching activity of marine carotenoid fucoxanthin and its metabolites. J Agric Food Sci. 2007;55:8516–8522. https://doi.org/10.1021/jf071848a

6. Mikami K, Hosokawa M. Biosynthetic pathway and health benefits of fucoxanthin, an algae-specific xanthophyll in brown seaweeds. Int J Mol Sci. 2013;14:13763–13781. https://doi.org/10.3390/ijms140713763

7. Li Y, Liu Q. Compositions containing fucoxanthin extract. Google patents 2013 [webpage on the Internet]. c2014 [cited 2014 Feb 15]. Available from: http://www.google.com/patents/US8454970

8. Lourenço-Lopes C, Garcia-Oliveira P, Carpena M, Fraga-Corral M, Jimenez-Lopez C, Pereira AG, et al. Scientific approaches on extraction, purification and stability for the commercialization of fucoxanthin recovered from brown algae. Foods. 2020;9(8):1113. https://doi.org/10.3390/foods9081113

9. Silva A, Silva SA, Carpena M, Garcia-Oliveira P, Gullón P, Barroso MF, et al. Macroalgae as a source of valuable antimicrobial compounds: Extraction and applications. Antibiotics. 2020;9(10):642. https://doi.org/10.3390/antibiotics9100642

10. Hekmat S, Reid G. Sensory properties of probiotic yogurt is comparable to standard yogurt. Nutr Res. 2006;26(4):163–166. https://doi.org/10.1016/j.nutres.2006.04.004

11. Salvatierra M, Molina A, Gamboa MDM, Arias ML. Evaluación del efecto de cultivos probióticos presentes en yogurt sobre *Staphylococcus aureus* y la producción de termonucleasa [Evaluation of the effect of probiotic cultures present in yogurt on *Staphylococcus aureus* and thermonuclease production]. Arch Latinoam Nutr. 2004;54(3):298–302. Spanish.

12. Saccaro DM, Hirota CY, Tamime AY, de Oliveira N. Evaluation of different selective media for enumeration of probiotic micro-organisms in combination with yogurt starter cultures in fermented milk. Afr J Microbiol Res. 2011;5(23):3901–3906. https://doi.org/10.5897/AJMR11.598

13. Fayed B, Abood A, El-Sayed HS, Hashem AM, Mehanna NSH. A synbiotic multiparticulate microcapsule for enhancing inulin intestinal release and Bifidobacterium gastro-intestinal survivability. Carbohydr Polym. 2018;93:137–143. https://doi.org/10.1016/j.carbpol.2018.03.068

14. Haugan JA, Akermann T, Liaaen-Jensen S. Isolation of fucoxanthin and piridinin. In: Packer L, editor. Carotenoids part A: Chemistry, separation, quantitation and antioxidants. Methods in enzymology. Volume 213. Cambridge, MA: Academic Press; 1992. p. 231–245. https://doi.org/10.1016/0076-6879(92)13124-G

15. Haugan JA, Liaaen-Jensen S. Improved isolation procedure for fucoxanthin. Phytochem. 1989;28(10):2797–2798. https://doi.org/10.1016/S0031-9422(00)98091-9

16. Wang LJ, Fan Y, Parsons R, Hu GR, Zhang PY, Li F. A rapid method for the determination of fucoxanthin in diatom. Mar Drugs. 2018;16(1):33. https://doi.org/10.3390/md16010033

17. Yang CM, Chang KW, Yin MH, Huang HM. Methods for the determination of the chlorophylls and their derivatives. Taiwania. 1998;43(2):116–122.

18. Ye H, Zhou C, Sun Y, Zhang X, Liu J, Hu Q, et al. Antioxidant activities in vitro of ethanol extract from brown seaweed *Sargassum pallidum*. Eur Food Res Technol. 2009;230(1):101–109. https://doi.org/10.1007/s00217-009-1147-4

19. Prieto P, Pineda M, Aguilar M. Spectrophotometric quantitation of antioxidant capacity through the formation of a phosphomolybdenum complex: Specific application to the determination of vitamin E. Anal Biochem. 1999;269:337–341. https://doi.org/10.1006/abio.1999.4019

20. Gülçin İ, Küfrevioğlu Ö, Oktay M, Büyükokuroğlu ME. Antioxidant, antimicrobial, antiulcer and analgesic activities of nettle (*Urtica dioica* L.). J Ethnopharmacol. 2004;90:205–215. https://doi.org/10.1016/j.jep.2003.09.028

21. El-Sayed HS, El-Sayed SM. A modern trend to preserve white soft cheese using nano-emulsified solutions containing cumin essential oil. Environ Nanotechnol Monit Manag. 2021;16:100499. https://doi.org/10.1016/j.enmm.2021.100499

22. El-Sayed SM, El-Sayed HS. Production of UF-soft cheese using probiotic bacteria and *Aloe vera* pulp as a good source of nutrients. Ann Agric Sci. 2020;65(1):13–20. https://doi.org/10.1016/j.aoas.2020.05.002

23. Miller LC, Tainter ML. Estimation of LD_{50} and its error by means of log-probit graph paper. Proc Soc Exp Biol Med. 1944;57:261. https://doi.org/10.3181/00379727-57-14776

24. Raj J, Chandra M, Dogra TD, Pahuja M, Raina A. Determination of median lethal dose of combination of endosulfan and cypermethrin in Wistar rat. Toxicol Int. 2013;20(1):1–5. https://doi.org/10.4103/0971-6580.111531

25. El-Sayed SM, El-Sayed HS, Elgamily HM, Youssef AM. Preparation and evaluation of yogurt fortified with probiotics jelly candy enriched with grape seeds extract nanoemulsion. J Food Process Preserv. 2022;46(7), e16713. https://doi.org/10.1111/jfpp.16713

26. El-Shafei K, Elshaghabee F, El-Sayed H, Kassem J. Assessment the viability properties of *Lactobacillus casei* strain using labneh as a carrier. Acta Scient Polon Technol Alimen. 2018;17:267–276. https://doi.org/10.17306/J.AFS.0583

27. International Dairy Federation (IDF). Dairy starter cultures of lactic acid bacteria LAB standard of identity. IDF standard no. 149A. Brussels: IDF; 1997.

28. Marshall RT. Standard methods for the examination of dairy products, 16th ed. American Public Health Association, Washington, DC. (1992).

29. Baggerman WI. A modified Rose Bengal medium for the enumeration of yeasts and moulds from foods. Eur J Appl Microbiol Biotechnol. 1981;12(4):242–247. https://doi.org/10.1007/BF00499496

30. AOAC. Official methods of analysis. 25th ed. Arlington, VA: Association of the Official Analytical Chemists; 2002.

31. Fayed B, El-Sayed HS, Abood A, Hashem AM, Mehanna NS. The application of multi-particulate microcapsule containing probiotic bacteria and inulin nanoparticles in enhancing the probiotic survivability in yoghurt. Biocatal Agric Biotechnol. 2015;22:101391. https://doi.org/10.1016/j.bcab.2019.101391

32. Yeni S, Sabdono N, Afiati N, Haeruddin H. Fucoxanthin identification and purification of brown algae commonly found in Lombok Island, Indonesia. Biodiversitas. 2021;22(3):1527–1534. https://doi.org/10.13057/biodiv/d220358

33. Seely GR, Duncan MJ, Vidaver WE. Preparative and analytical extraction of pigments from brown algae with dimethyl sulfoxide. Mar Biol. 1972;12(2):184–188. https://doi.org/10.1007/BF00350754

34. Balboa E, Conde E, Moure A, Falqué E, Domínguez H. In vitro antioxidant properties of crude extracts and compounds from brown algae. Food Chem. 2013;138(2–3):1764–1785. https://doi.org/10.1016/j.foodchem.2012.11.026

35. Agrahari AK, Panda SK, Ashutosh M, Padhan AR, Mohd K. Phytochemical screening of *Curculigo orchioides* Gaertn. root tubers. J Chem Pharm Res. 2010;2(2):107–111. https://jopr.com/vol2-1ss2-2010/JOCPR

36. Sudharsan S, Seedevi P, Ramasamy P, Subhapradha N, Vairamani S, Shanmugam A. Heavy metal accumulation in seaweeds and sea grasses along southeast coast of India. J Chem Pharm Res. 2012;4(9):4240–4244.

37. Karpiński TM, Adamczak A. Fucoxanthin – An antibacterial carotenoid. Antioxidants. 2019;8(8):239. https://doi.org/10.3390/antiox8080239

38. Liu Z, Sun X, Sun X, Wang S, Xu Y. Fucoxanthin isolated from *Undaria pinnatifida* can interact with *Escherichia coli* and lactobacilli in the intestine and inhibit the growth of pathogenic bacteria. J Ocean Univ. 2019;18(4):926–932. https://doi.org/10.1007/s11802-019-4019-y

39. Hosseini MS, Shahbazizadeh S, Khosravi-Darani K, Reza Mozafari M. *Spirulina paltensis*: Food and function. Curr Nutr Food Sci. 2013;9(3):189–193. https://doi.org/10.2174/1573401311309030003

40. Beheshtipour H, Mortazavian AM, Haratian P, Darani KK. Effects of *Chlorella vulgaris* and *Arthrospira platensis* addition on viability of probiotic bacteria in yogurt and its biochemical properties. Eur Food Res Technol. 2012;235(4):719–728. https://doi.org/10.1007/s00217-012-1798-4

41. Gyenis B, Szigeti JF, Ásványi-Molnár N, Varga L. Use of dried microalgal biomasses to stimulate acid production and growth of *Lactobacillus plantarum* and *Enterococcus faecium* in milk. Acta Agraria Kaposváriensis. 2005;2:53–59.

42. Iio K, Okada Y, Ishikura M. Single and 13-week oral toxicity study of fucoxanthin oil from microalgae in rats. Shokuhin Eiseigaku zasshi. Journal of the Food Hygienic Society of Japan. 2011;52(3):183–189. https://doi.org/10.3358/shokueishi.52.183

43. Abu-Ghannam N, Shannon E. Seaweed carotenoid, fucoxanthin, as functional food. In: Gupta VK, Treichel H, Shapaval V, Antonio de Oliveira L, Tuohy MG. Microbial functional foods and nutraceuticals. Hoboken, NJ: John Wiley & Sons; 2018. p. 39–64. https://doi.org/10.1002/9781119048961.ch3

44. Mok IK, Yoon JR, Pan CH, Kim SM. Development, quantification, method validation, and stability study of a novel fucoxanthin-fortified milk. J Agric Food Chem. 2016;64(31):6196–6202. https://doi.org/10.1021/acs.jafc.6b02206

45. Heo SJ, Jeon YJ. Protective effect of fucoxanthin isolated from *Sargassum siliquastrum* on UV-B induced cell damage. J Photochem Photobiol B Biol. 2009;95(2):101–107. https://doi.org/10.1016/j.jphotobiol.2008.11.011

46. El-Shibiny S, El-Gawad MA, Assem FM, El-Sayed SM. The use of nano-sized eggshell powder for calcium fortification of cow's and buffalo's milk yogurts. Acta Sci Pol Technol Aliment. 2018;17(1):37–49. https://doi.org/10.17306/J.AFS.2018.0541

47. Mise T, Ueda M, Yasumoto T. Production of fucoxanthin rich powder from *Cladosiphon okamuranus*. Adv J Food Sci Technol. 2011;3:73–76.

48. El-Sayed HS, El-Sayed SM, Youssef AM. Novel approach for biosynthesizing of zinc oxide nanoparticles using *Lactobacillus gasseri* and their influence on microbiological, chemical, sensory properties of integrated yogurt. Food Chem. 2021;365:130513. https://doi.org/10.1016/j.foodchem.2021.130513

49. El-Naggar ME, Hussein J, El-Sayed SM, Youssef AM, El Bana M, Latif YA, et al. Protective effect of the functional yogurt based on *Malva parviflora* leaves extract nanoemulsion on acetic acid-induced ulcerative colitis in rats. J Mater Res Technol. 2020;9(6):14500–14508. https://doi.org/10.1016/j.jmrt.2020.10.047

50. El-Sayed HS, El-Sayed SM, Youssef AM. Designated functional microcapsules loaded with green synthesis selenium nanorods and probiotics for enhancing stirred yogurt. Sci Rep. 2022;12(1):1–15. https://doi.org/10.1038/s41598-022-18781-w

51. Prabhasankar P, Ganesan P, Bhaskar N, Hirose A, Stephen N, Gowda LR. Edible Japanese seaweed, wakame (*Undaria pinnatifida*) as an ingredient in pasta: Chemical, functional and structural evaluation. Food Chem. 2009;115(2):501–508. https://doi.org/10.1016/j.foodchem.2008.12.047

GUEST LEADER

What's in a face? Introducing the special section on Face Science.

AUTHORS:
Colin Tredoux[1,2]
Alicia Nortje[1]

AFFILIATIONS:
[1]Department of Psychology, University of Cape Town, Cape Town, South Africa
[2]Cognition, Languages, Language, Ergonomics Laboratory, University of Toulouse – Jean Jaurès, Toulouse, France

HOW TO CITE:
Tredoux C, Nortje A. What's in a face? Introducing the special section on Face Science. S Afr J Sci. 2023;119(3/4), Art. #15663. https://doi.org/10.17159/sajs.2023/15663

FUNDING:
Oppenheimer Memorial Trust

The faces of animals are central to their functioning in lived environments. For mammals, faces are usually the location of the sense organs, and thus provide the primary way of living in *umwelten*[1] (umwelt is the world as experienced by a particular organism). Mammalian faces are also an important location of non-verbal communication and expression, and convey a wide range of emotional and social information; for humans and many other primates, faces are also key for personal identity. The special significance of the face is likely the reason that the brains of primates have multiple face-selective cortical areas, the most well-known region in humans being the fusiform face area.[2]

For these, and many other reasons, faces are of considerable interest to scientists across a range of disciplines. Often the interest is at the level of basic scientific work, for instance describing the brain regions that are involved in the complex act of cognitively processing faces[3], or representing human faces with statistical models to compare populations of *Homo sapiens* across evolutionary time periods[4]. The interest in human faces is also driven by practical applications, such as automatic recognition of faces by computers[5], forensic portraiture of living or deceased people for identification[6], and rapid identification of potential illness or physical disorder in low-cost settings[7], among many others. Within policing and other law enforcement services, there exist several groups of practitioners who are especially involved in applied work that involves the human face; in the South African Police Service, at least three such clusters of applications exist: the Face Recognition Unit assists witnesses in creating composite portrait images of faces of suspects; a Face Reconstruction Group that reconstructs faces post-mortem to enable identification of deceased people; and then in standard detective work, police officers conduct identification parades in order to test and adduce evidence of identification in ongoing criminal investigations. A fourth application is matching photographs of faces to individuals, as performed by immigration officials at ports of entry into South Africa; although a different type of law enforcement agency, their work and abilities at matching photographs of faces in passports to faces of people physically co-present has long been important.

We noted the wide-ranging transdisciplinary interest in the human face some years ago during some research in collaboration with the South African Police Service[8], and instituted an annual Face Science Symposium (now in its eleventh year) to bring together researchers and practitioners from these diverse academic disciplines. This interdisciplinary symposium provides a platform for researchers and practitioners to present their work, to share their knowledge, and to engage in discussions and collaborations.

The special section on Face Science in this issue of the *South African Journal of Science* is a collection of articles that appeared in one form or another at one of these symposia. The articles present cutting-edge research and new perspectives on the study of faces within South Africa and provide insights into the promise of interdisciplinarity in the scientific study of faces.

In the first of the articles, Felix Atuhaire and Tinashe Mutsvangwa of the University of Cape Town, and Bernhard Egger of the Friedrich-Alexander-Universität, Erlangen-Nürnberg, describe research on using 3D face modelling to assess foetal alcohol syndrome (FAS). FAS is a condition caused by maternal alcohol consumption during pregnancy, and is especially prevalent in South Africa, where it has been estimated as being between 59.3 to 91.0 cases per 1000 people.[9] Facial dysmorphology of FAS is a key factor in early diagnosis. Current methods for automated analysis of the FAS facial phenotype use 3D facial image data from expensive surface scanning devices. The research reported by Atuhaire and colleagues used a 3D face model learned from a database of registered 3D face scans to reconstruct 3D face surfaces from single frontal 2D images. An important consideration driving their research was to find a solution that is low cost, and that can be used in resource-challenged contexts. The authors show the potential of the proposed framework to reliably estimate 3D landmark positions for components of the face associated with the FAS facial phenotype, using input images obtained from relatively low-resolution cameras. The study concludes by emphasising that future work should focus on improving accuracy and adapting the approach to predict face data of individuals with FAS. The article thus applies biomedical engineering and computer science skills to an important practical problem in southern Africa.

A second article, by Kyra Scott, Colin Tredoux and Alicia Nortje, all three associated with the University of Cape Town, applied computing methods to face images to generate synthetic faces from descriptions given by eyewitnesses, and to compare these to face composites made by witnesses with extant composite software. The reliability of facial identification evidence obtained from eyewitnesses, such as person descriptions and facial composites, is often questioned, and Scott and colleagues were interested in whether statistical models of faces could be used to produce better quality face composites than are presently possible (for a review of research on face composites, see the article by Vredeveldt and colleagues[10]). To test this hypothesis, a study with 167 participants compared the accuracy and precision of identifying a target face using person descriptions, facial composites, and computer-generated synthetic faces produced from person descriptions. The former two conditions used methods that are already used extensively in police forces around the world and were thus a useful basis for comparing the new idea of generating synthetic faces from descriptions. Results showed that person descriptions had higher accuracy but lower precision in narrowing down the suspect pool than the other two methods. The synthetic faces generated from descriptions did not fare any better than faces generated with extant composite systems. Their study highlighted the importance of person descriptions in accurately identifying unknown perpetrators, which although surprising – as a simple verbal description seems unlikely to capture more information about a human face than a complete visual representation – is consistent with earlier research. The study also introduces a distinction between concepts of identification precision and accuracy when assessing the utility of facial identification information.

Milton Gering, Tayla Johnson and Colin Tredoux, all from the University of Cape Town, considered the effects that stress has on face recognition in a simulated criminal context. The extant literature is unclear about this relation, and as Gering and colleagues point out, do not use experimental designs that honour the long-recognised[11] non-linear nature of the relation between arousal and performance. They present the results of two experiments examining the impact of stress on eyewitness performance in line-up face recognition tasks. In Experiment 1 they replicated previous studies in the area, finding a null result, while in Experiment 2 they found a non-linear relationship between stress and performance, with the stress group experiencing moderate levels of stress showing better recognition accuracy than low or high stress groups. These kinds of studies are difficult to conduct, given the obvious ethical constraints when inducing simulated stress, but Gering and colleagues found an ingenious way to induce different, and high levels of realistic stress in a naturalistic way. Their results support the proposition that the relationship between stress and face recognition is likely non-linear, and contend that the failure to find this, or even really address it in earlier studies, suggests the need for a different approach in future experiments on the topic.

The final article included in the special section is by Colin Tredoux, Ahmed Megreya, Alicia Nortje and Kate Kempen, and grew out of a collaboration between the University of Cape Town and Qatar University. Their study examines the own group bias (OGB) in face recognition, which is a tendency to better recognise faces of one's own group than another group; typically, 'group' is defined as a race or ethnicity, which is how it was defined in this study, but in other research has been extended to age and sex. The OGB has serious implications in criminal investigations and can lead to incorrect identifications, mistaken convictions, and imprisonment, which has happened several hundred times in the USA.[12] The experiment reported by Tredoux et al. aimed to determine whether the OGB occurs during encoding or retrieval of faces from memory. Participants, both black and white South Africans, encoded faces of both same and other races, and immediately tried to match the faces to members of photograph arrays. After a further delay, they tried to identify the faces from long-term memory. The results showed a crossover OGB in the delayed matching task, but an asymmetrical OGB at retrieval. Tredoux and colleagues then reasoned that in order to distinguish encoding from retrieval processes one should consider only stimuli successfully matched in the first phase of the experiment. Further investigation of recognition performance using these stimuli showed a non-significant OGB at retrieval but the authors concede that they may not have had enough statistical power to detect an OGB. Their study highlights the rapidly changing nature of the OGB across encoding and retrieval and uses a novel approach to distinguish effects at encoding and retrieval, but further work is needed to provide clear evidence of the origin of OGB.

Conclusion

We believe that the articles in this special section of the *South African Journal of Science* have demonstrated that there is active interest and research on the human face in a variety of disciplines in South Africa. Although we have showcased work here that is founded in the disciplines of Biomedical Engineering, Computer Science, and Psychology, there is much more work to be found in South Africa from disciplines such as Anatomy, Fine and Forensic Art, Physical and Social Anthropology, Policing, Semiotics, and Sociology. The special section also shows the promise of interdisciplinarity for better understanding of basic processes and for cogent applications of knowledge to practice. As the present articles all emerged in some form from the annual Face Science Symposium that we have organised since 2012, it is also evidence that explicitly interdisciplinary initiatives of such kind can bear fruit and bring interdisciplinary work into being.

References

1. Von Uexküll J. A foray into the worlds of animals and humans with a theory of meaning. Minneapolis, MN: University of Minnesota Press; [1934] 2010.
2. Kanwisher N, Yovel G. The fusiform face area: A cortical region specialized for the perception of faces. Proc Royal Soc B. 2006;361(1476):2109–2128. https://doi.org/10.1098/rstb.2006.1934
3. Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. TiCS. 2000;4(6):223–233. https://doi.org/10.1016/S1364-6613(00)01482-0
4. Lieberman DE, McBratney BM, Krovitz G. The evolution and development of cranial form in *Homo sapiens*. Proc Natl Acad Sci USA. 2002;99(3):1134–1139. https://doi.org/10.1073/pnas.022440799
5. Kumar A, Kaur A, Kumar M. Face detection techniques: A review. Artif Intell Rev. 2019;52:927–948. https://doi.org/10.1007/s10462-018-9650-2
6. Tredoux CG, Frowd C, Vredeveldt A, Scott K. Construction of facial composites from eyewitness memory. In: Shapiro, L, Rea PM, editors. Biomedical visualisation. Volume 13. New York: Cham Springer International Publishing; 2022. p. 149–190. https://doi.org/10.1007/978-3-031-13021-2_8
7. Sforza C, Ferrario VF. Soft-tissue facial anthropometry in three dimensions: From anatomical landmarks to digital morphology in research, clinics and forensic anthropology. J Anthropol Sci. 2006;84:97–124.
8. Vredeveldt A, Tredoux CG, Nortje A, Kempen K, Puljević C, Labuschagne GN. A field evaluation of the Eye-Closure Interview with witnesses of serious crimes. Law Hum Behav. 2015;39(2):189–197. https://doi.org/10.1037/lhb0000113
9. Olivier L, Viljoen DL, Curfs LM. Fetal alcohol spectrum disorders: Prevalence rates in South Africa: The new millennium. S Afr Med J. 2016;106(sup1):103–106. https://doi.org/10.7196/SAMJ.2016.v106i6.11009
10. Vredeveldt A, Tredoux CG. Composite communication: How dissemination of facial composites in the media affects police investigations. Memory Mind Media. 2022;1, e10. https://doi.org/10.1017/mem.2022.4
11. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol 1908;18(5):459–482. https://doi.org/10.1002/cne.920180503
12. Innocence Project. How eyewitness misidentification can send innocent people to prison [webpage on the Internet]. c2020 [cited 2023 Feb 03]. Available from: https://innocenceproject.org/how-eyewitness-misidentification-can-send-innocent-people-to-prison/

AUTHORS:
Felix Atuhaire[1]
Bernhard Egger[2,3]
Tinashe Mutsvangwa[4]

AFFILIATIONS:
[1]Department of Biomedical Sciences and Engineering, Mbarara University of Science and Technology, Mbarara, Uganda
[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[3]Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[4]Division of Biomedical Engineering, Department of Human Biology, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Tinashe Mutsvangwa

EMAIL:
tinashe.mutsvangwa@uct.ac.za

DATES:
Received: 24 Aug. 2021
Revised: 22 Feb. 2022
Accepted: 23 Nov. 2022
Published: 29 Mar. 2023

HOW TO CITE:
Atuhaire F, Egger B, Mutsvangwa T. Evaluating 3D human face reconstruction from a frontal 2D image, focusing on facial regions associated with foetal alcohol syndrome. S Afr J Sci. 2023;119(3/4), Art. #12064. https://doi.org/10.17159/sajs.2023/12064

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Michael Inggs

KEYWORDS:
3D morphable models, model fitting, face reconstruction, fetal alcohol syndrome, 3D face analysis

FUNDING:
European Commission; South African Department of Science and Innovation; South African National Research Foundation (grant no. 98788); Swiss National Science Foundation (P400P2_191110)

Evaluating 3D human face reconstruction from a frontal 2D image, focusing on facial regions associated with foetal alcohol syndrome

Foetal alcohol syndrome (FAS) is a preventable condition caused by maternal alcohol consumption during pregnancy. The FAS facial phenotype is an important factor for diagnosis, alongside central nervous system impairments and growth abnormalities. Current methods for analysing the FAS facial phenotype rely on 3D facial image data, obtained from costly and complex surface scanning devices. An alternative is to use 2D images, which are easy to acquire with a digital camera or smart phone. However, 2D images lack the geometric accuracy required for accurate facial shape analysis. Our research offers a solution through the reconstruction of 3D human faces from single or multiple 2D images. We have developed a framework for evaluating 3D human face reconstruction from a single-input 2D image using a 3D face model for potential use in FAS assessment. We first built a generative morphable model of the face from a database of registered 3D face scans with diverse skin tones. Then we applied this model to reconstruct 3D face surfaces from single frontal images using a model-driven sampling algorithm. The accuracy of the predicted 3D face shapes was evaluated in terms of surface reconstruction error and the accuracy of FAS-relevant landmark locations and distances. Results show an average root mean square error of 2.62 mm. Our framework has the potential to estimate 3D landmark positions for parts of the face associated with the FAS facial phenotype. Future work aims to improve on the accuracy and adapt the approach for use in clinical settings.

Significance:

Our study presents a framework for constructing and evaluating a 3D face model from 2D face scans and evaluating the accuracy of 3D face shape predictions from single images. The results indicate low generalisation error and comparability to other studies. The reconstructions also provide insight into specific regions of the face relevant to FAS diagnosis. The proposed approach presents a potential cost-effective and easily accessible imaging tool for FAS screening, yet its clinical application needs further research.

Introduction

Early detection of foetal alcohol syndrome (FAS) allows for early intervention, mitigates the onset of secondary disorders such as mental breakdown or improper sexual behaviours, and leads to significantly better clinical outcomes.[1] The diagnosis of FAS is based on the evidence of central nervous system abnormalities, evidence of growth abnormalities, and a characteristic pattern of facial anomalies, specifically short palpebral fissure length, smooth philtrum, flat upper lip, and flat midface.[2,3] The FAS facial phenotype has been emphasised clinically for diagnosis.[4-7] However, clinical evaluation requires the expertise of trained dysmorphologists. This requirement limits efforts for large-scale screening in suspected high prevalence regions, such as South Africa, which has a prevalence rate estimated to be between 93 and 128 per 1000 live births[8], and a shortage of highly trained clinical personnel. Alternative methods for assessing the FAS facial phenotype are possible but require careful acquisition of face data. Face data collection methods include direct anthropometry using handheld rulers and callipers. Indirect anthropometry, on the other hand, is possible through the acquisition of face data through 2D photogrammetry, 3D stereophotogrammetry, and 3D surface imaging scanners.[2,9,10] Direct anthropometry introduces inaccuracies due to the indentation of some features during contact measurements with physical instruments. For this reason, more efforts have been put into indirect anthropometry, which has the added benefit of near-instantaneous patient data acquisition. Furthermore, with indirect approaches, measurements on the images can be repeated in the absence of subjects. Indirect evaluation on 3D image data is typically more accurate than on 2D images.[11] However, acquiring 3D face images using 3D surface scanners tends to be costly and precludes large-scale deployment in low-resource settings.

Reconstruction of the 3D human face from a single 2D image is a popular topic of research, with applications in face recognition, face tracking, face animation, and medical analysis of faces.[12] However, to date, there has not been any report on the quantitative suitability of 3D from 2D face reconstruction for FAS-related facial phenotype characterisation. In this study, our aim was to evaluate the geometric accuracy of a 3D human face reconstruction from a single 2D facial image, using a 3D morphable model of the face.[13] We focused on 3D reconstruction of the complete face to enable surface-based approaches, and to allow us to evaluate landmark and distance-based measurements. We tested if such a reconstruction algorithm could be suitable for automated analysis of facial features related to FAS.

Related work

Three-dimensional morphable models (3DMMs) are high-resolution generative models containing shape and texture variations from sample populations.[13-17] Typically, 3DMMs are built from a set of 3D face scans after establishing anatomical dense correspondences across the face data set. Establishing correspondences ensures that similar

features across a set of 3D face scans match each other (e.g. the tip of the nose or the eye corners) – we call this process 'registration'.

Several methods for building 3DMMs from a set of 3D face scans have been presented over the years.[12] In pioneering work, Blanz and Vetter[13] built a 3DMM from a set of face scans after computing dense correspondences with an optical flow-based registration technique. The shape and texture variations in a collection of face scans were then modelled using principal component analysis (PCA), resulting in a low dimensional representation. The learned face models were used to estimate a 3D face surface from a single 2D face image. Early 3DMMs were built using just hundreds of face scans. However, a recent study by Booth et al.[16] constructed a 3DMM known as the Large-Scale Facial Model using 9663 3D facial scans. Booth et al.[16] used the non-rigid iterative closest point (NICP) algorithm[18] for registration of the template face surface to each target face scan in the data set, aided by generalised Procrustes analysis (GPA) for similarity alignment of the registered face scans. They then used PCA[19] for statistical analysis of the registered face scans. The 3DMMs have already been successfully applied in various application areas including face tracking, face recognition, face segmentation, and face reconstruction.[12] However, additional research focusing on human face variations would still be required before the morphable model could be used for medical purposes.[16]

Blanz and Vetter's[13] work was seminal, but their approximate 3D face meshes were only qualitatively evaluated. Romdhani and Vetter[20] took a different approach, extracting multiple features from a single image. The extracted features were then used to estimate a 3D face surface by minimising a cost function. In 2009, 3DMM, the Basel face model, was made available for research purposes and enabled the community to grow faster.[21] Aldrian and Smith[22] developed the first publicly available inverse graphics algorithm based on a 3DMM. Schönborn et al.[23] employed a sampling-based approach to fit a Gaussian process morphable model to a single 2D image. The face shape reconstruction accuracy as measured by a root mean squared average was 3.79 mm. Recently, a first benchmark was established for 3D reconstruction from 2D images.[24] This benchmark is, however, strongly biased towards light skin tones, which is a narrow subset of the world's population and might not be representative for general clinical application. The state-of-the-art method on this benchmark is a deep learning based method for 3DMM reconstruction, with an average reconstruction error of 1.38 mm.[25]

While reconstruction algorithms are reported in the literature, there is limited research evaluating the accuracy of these algorithms, which has implications for the algorithm performance on medical-related applications. Additionally, to the best of our knowledge, model-based 3D face reconstruction from 2D image approach has not been evaluated with a focus on FAS applications, perhaps because 3D ground truth data may not be available. A robust single image-based reconstruction approach could offer a cost-effective alternative to 3D surface capturing systems.

Methods

Data description

We based our experiments on the BU-3DFE face database, which is a publicly available data set of high-quality 3D scans, acquired using the 3dMD face system.[26] It consists of face scans of 98 subjects of different ethnicities (56 female and 42 male subjects aged between 18 and 70 years). We used only the facial scan with a neutral expression for each identity (see Figure 1 for an example of the images). The data were used with ethical approval from both the University of Cape Town and the State University of New York. To maximise the number of faces for training, we performed a leave-one-out cross-validation scheme for our experiments. From each face scan, we derived the 3D ground truth face shape, established correspondence to our model template, and rendered a frontal 2D image for our 2D to 3D reconstruction task. To reach maximal accuracy, we used 12 manual landmarks to initialise the 2D to 3D reconstruction process: right outer and inner canthi, glabella, left inner and outer canthi, right and left alares, pronasale, subnasale, right and left cheilions. We did not rely strictly on these landmarks as the fitting framework used has been shown to work with automatic landmark detection. This gave us a set of 2D images with known ground truth 3D shapes for learning and evaluating our model and reconstruction framework.

Figure 1: Some examples of the 2D images in the test set generated from the BU-3DFE face data set.

Rigid alignment of face scans

The goal of rigid alignment is to bring all the face scans into a common coordinate system without deformations. Given a set of pre-processed 3D face scans (pre-processing involves trimming the face scans to remove the unwanted regions such as the hair and neck regions) and a set of facial landmarks for each face scan, the facial landmarks were used to calculate a least-squares alignment that brought landmarks corresponding across scans as close together as possible (Procrustes alignment). The training face scans were mapped, using rigid transformations, to the mean of the Basel face model[27], which represents a common reference face surface. The results of these alignments are a collection of rigidly aligned 3D face scans.

Registration of face scans

After rigid alignment, we used a deformable model to establish dense correspondences between a reference face surface and each target face scan in the training data. By dense correspondences, we mean finding the mappings between similar features across the data set. The goal of registration is to re-parameterise the face scans to have the same number of vertices and triangulations across face scans in the training set with the key feature that each vertex corresponds to the same point on each face. The reference face surface is fitted to each target face scan, using a Gaussian process fitting approach[27], to obtain dense face surface deformations, which best match a target face scan to a common reference face surface. The time for registering the reference to each target face was 5–8 minutes computed on an Intel(R) Core (TM) i5-8350U CPU @ 1.7 GHz. This registration approach builds on a Gaussian process defined by mean and covariance functions to model smooth deformations of the template shape.[28,29] During registration, we searched the optimal set of parameters of our Gaussian process model to match the 3D scan at hand. The results of applying the fitting approach are registered 3D face scans. As we wanted to build a 3DMM, at this stage, we also extracted the colour per vertex from the closest vertex on the face scan – this enabled us to not only build a shape but also a texture model.

Building face models

With dense correspondences established among the training data, we removed translations and rotations on the data to retain shape deformation. To perform these non-shape-related transformations on the training data, we applied the Procrustes analysis approach.[30-32] After removing alignments, the principal modes of variation were extracted from training data using PCA[28,29] to build 3D morphable face models (3DMMs). The 3DMMs consisted of the face means and the principal components as modes of variation. The 3DMMs are expressed as linear combinations of shape and texture vectors in the face subspace. The time required to build a surface model from each registered face scan was 10–20 minutes computed on an Intel(R) Core (TM) i5-8350U CPU @ 1.7 GHz. An example of registration results as well as a resulting 3DMM built from all 98 scans are illustrated in Figure 2.

3D from 2D face reconstruction

The key application of a 3DMM that we were interested in was the estimation of 3D face surfaces and 3D landmark positions for FAS measurements from 2D images. In the reconstruction setting, the 3DMM

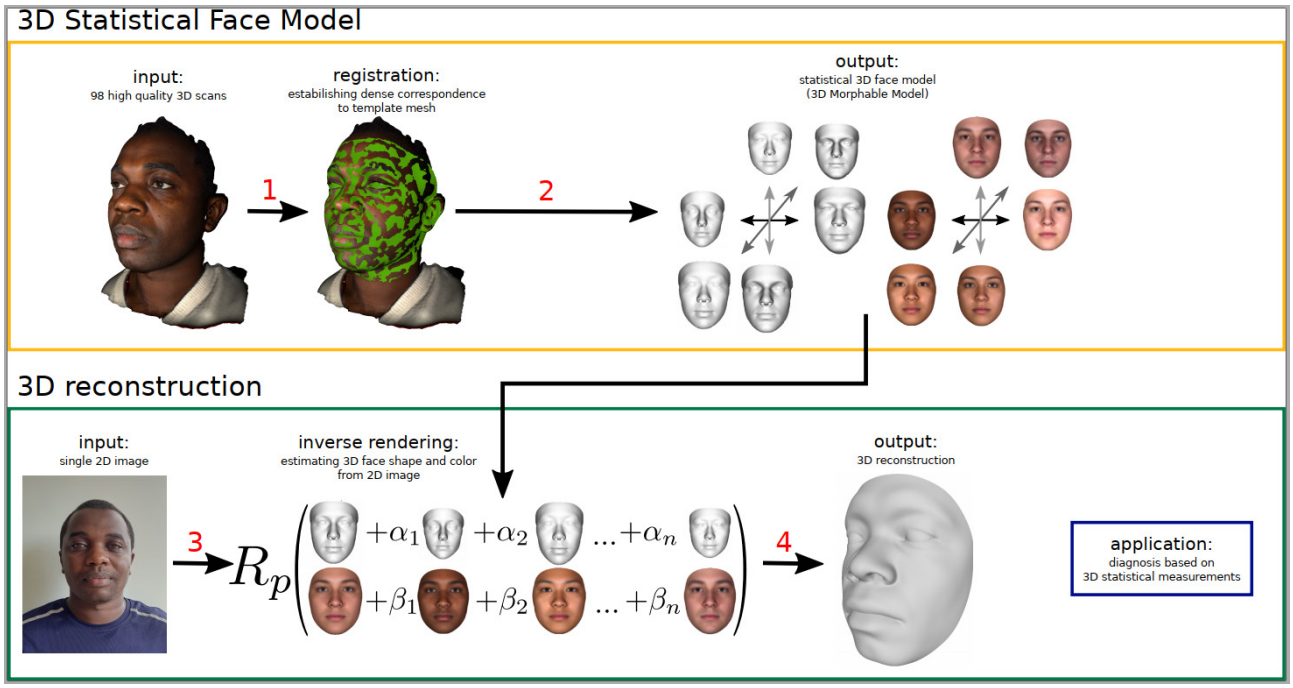


Figure 2: Statistical face modelling and model fitting pipelines. The yellow box on top shows the model building steps and the green box at the bottom illustrates the 3D from 2D estimation. In Step 1, 3D face scans are registered based on the reference scan, and in Step 2, the registered 3D scans are used to build the face model. Step 3 illustrates the 3D from 2D reconstruction process and Step 4 presents the 3D reconstruction result based on the single input image. In the reconstruction step, *R* is the rendering function, *p* represents rendering parameters, α_n are shape parameters, and β_n are colour parameters. Note that the images used are for illustrative purposes only.

acts as a prior of 3D face shape, and we searched for the most likely reconstruction given only 2D images. This is potentially useful because 2D images, in contrast to 3D scans, are easy to acquire using either a mobile phone or a portable camera. One of the goals of this study was to reconstruct a neutral 3D face with shape and texture information from a single frontal 2D image and evaluate how close that reconstruction was to the known ground truth.

Given single 2D images, we estimated 3D face reconstructions by fitting the morphable model. The reconstruction time measured on an Intel(R) Core (TM) i5-8350U CPU @ 1.7 GHz was 58 minutes. We applied an approach proposed by Schönborn et al.[23] to fit a 3DMM to a single 2D image. The fitting algorithm recovers a full posterior model of the face by simultaneously optimising facial shape and texture as well as illumination and camera parameters for a test face image. We used a spherical harmonics illumination model which can recover a broad range of natural illumination conditions in combination with a pinhole camera model. Illumination estimation is a critical step and optimised early and regularly in the sampling process. The fitting algorithm tries to reconstruct the 2D image, producing a rendering from the 3D model that matches the 2D image as closely as possible. The results for fitting a morphable model to a single 2D image are 3D face shape and texture reconstructions, as illustrated in the pipeline in Figure 2.

Experiments and results

Evaluating the face shape model

Before applying the 3DMM in our downstream reconstruction of a 3D face surface from a single 2D image, it was necessary to evaluate the quality of the built face model in terms of generalisation, specificity, and compactness. The details of the model evaluation metrics are discussed by Styner et al.[33]

Shape model generalisation: This refers to the ability of the shape model to accurately represent an instance for which it was not trained. The leave-one-out approach[33] was used to evaluate the generalisation ability of the face shape model. For each iteration, a shape model was constructed from a set of training face surfaces, leaving out one face shape instance. With all the training data in correspondence, the left-out face instance was projected into the shape model space to generate a face estimate. To evaluate the geometric accuracy of the estimated face, the distance between the face estimate and the original left-out face instance was calculated. The average vertex-to-vertex root mean squared (RMS) distance between the left-out face instance and the estimated face instance was computed. The procedure was repeated until all the face instances in the training set were used and each time the evaluation metric was calculated. The model generalisation ability results are presented in Figure 3, which demonstrates the generalisation error represented as RMS distance (*y* axis), plotted against shape principal components (*x* axis). After 5 principal components, the generalisation accuracy was close to 1.5 mm, and with 50 principal components, we reached an accuracy of approximately 0.5 mm.

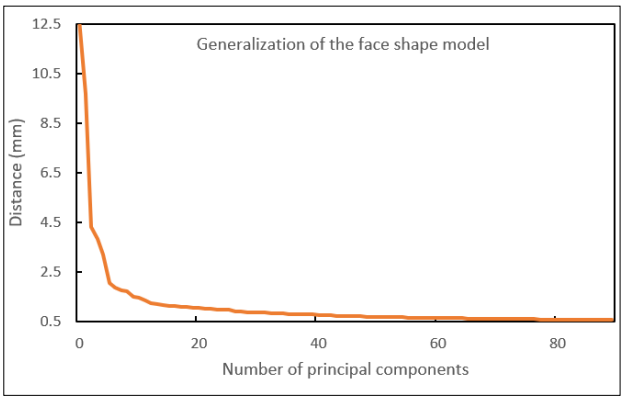


Figure 3: Generalisation: root mean squared (RMS) distances as a function of the number of principal components in the reconstruction.

Shape model specificity: This is defined as the ability of the shape model to randomly generate valid synthetic shape instances that are similar to real shape instances present in the training data set.[33] To evaluate

model specificity, a set of 90 shape instances was randomly generated from a distribution of the 3D morphable face model. The RMS distance between the randomly generated shape instances and the closest face surfaces in the training set was calculated as a specificity estimate. Lower RMS deviations are desirable because they indicate that the synthesised shape instances are close to the real shape instances in the training set. Figure 4 shows the specificity results. The results are in common ranges for specificity.[15] Note that it is typical that the specificity decreases (distance increases) with greater model complexity (number of principal components).

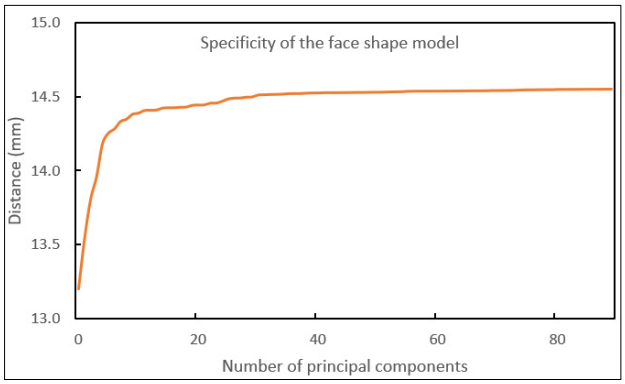


Figure 4: Specificity: root mean squared (RMS) distances as a function of the number of principal components.

Shape model compactness: This indicates the percentage of variability accounted for by increasing numbers of principal components. Fewer principal components capture variability in shape information more efficiently. To validate the model compactness, the cumulative variance accounted for by the shape model was plotted as a function of the number of principal components of the shape model (illustrated in Figure 5). The line reflecting cumulative variance flattens as the number of shape principal components increases. Using only the first 20 shape principal components, the shape model accounts for more than 90% of shape variation in the training data set. This implies that the shape model is compact as it describes the training data set using a small number of principal components.

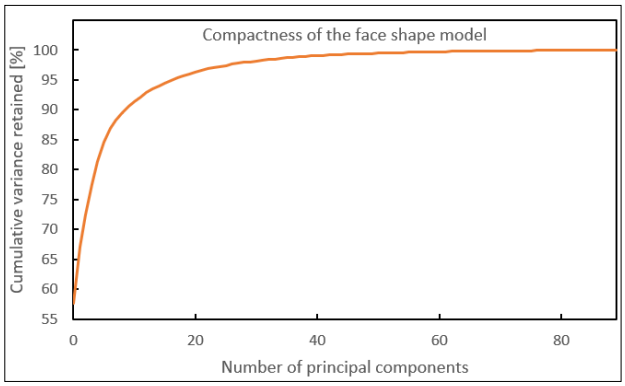


Figure 5: Compactness: cumulative variance against the number of principal components.

Evaluating 3D from 2D reconstruction results

To evaluate the geometric accuracy of face shape reconstructions from a single 2D image, the predicted 3D face surfaces were compared to the ground truth 3D face scans. We performed the reconstruction for each of the 98 2D images separately and build a separate 3DMM, removing that identity from the training data (leave-one-out cross-validation).

To measure the reconstruction error, we first rigidly aligned each predicted face mesh with the associated ground truth face shape. Following surface alignment, we computed the difference between the aligned face surfaces using the RMS distance metric.[34] The RMS metric gives the surface-to-surface assessment value for each pair of surface comparisons. To visualise the reconstruction error distribution on the face surface, we additionally generated the surface colour maps from the comparisons of the predicted face surfaces and the associated ground truth face scans.

The overall average RMS error between the pairs of the predicted 3D face surfaces and the ground truth 3D face shapes is 2.62 mm with a deviation of 1.41 mm, with errors ranging from 1.00 mm to 6.75 mm. Furthermore, the visual shape comparisons between the predicted face surfaces and the corresponding ground truth face shapes are represented using colour gradients as illustrated in column (e) of Figure 6.

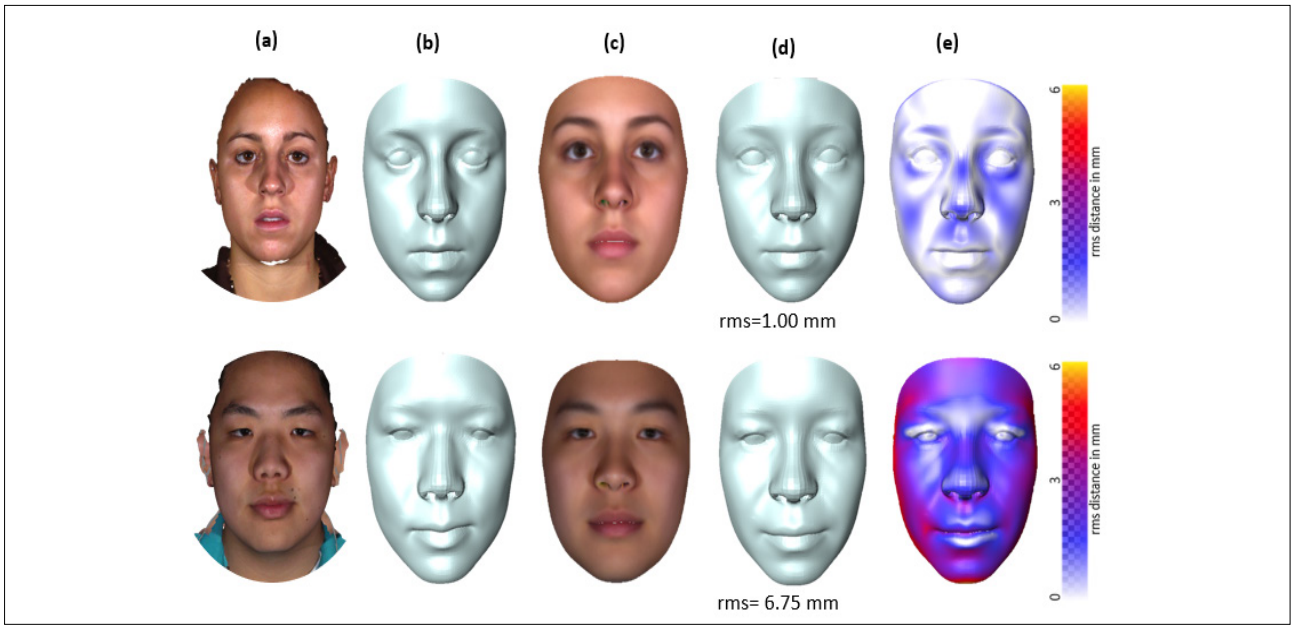


Figure 6: Visualisations of the global best and worst face predictions. The first column (a) indicates the target 2D face images (input). The second column (b) represents the ground truth face shapes. The third column (c) illustrates the reconstructed face textures (output). The fourth column (d) shows the predicted face shapes (output). The fifth column (e) illustrates the face surface colourmap comparisons. The best reconstruction is in the top row, and the worst in the bottom.

Figure 6 shows the identities with the best and worst RMS values for the predicted face shapes, and their corresponding ground truth face shapes, including the face surface colour map comparison. We observe that the largest reconstruction errors are found in regions of the face that are not relevant when screening for facial phenotypes in FAS. However, we find that the philtrum, which is one of the discriminators for FAS facial analysis, shows larger errors in faces considered to be outliers.

Face surface analysis across skin tones

Foetal alcohol syndrome affects people of all ethnicities. Previous methods did not explore darker-skinned individuals well enough and structured light systems have acquisition issues when it comes to imaging darker skin tones.[35] Furthermore, previous models have a strong bias towards lighter skin tones.[36] We investigated what happens when darker and lighter skin tones are mixed. We can visualise the distribution of skin tones across our data set in Figure 7 and observe a heavy tail in the low intensity range. We also investigated the relationship between skin tone and the reconstruction error per mesh, and observe that we reach comparable reconstruction accuracy for the heavy tail of low intensity skin tones even though they are underrepresented in the training data (Figure 8). Figure 9 shows faces with lowest and highest reconstruction error results across skin tones. The poorest face reconstructions are also indicated in Figure 8 with orange circles, while the best face reconstructions are illustrated in the same figure with green circles. We find that regions of the face that are not related to the FAS facial phenotype are most affected. We also present the average reconstruction error over the whole data set in Figure 10.

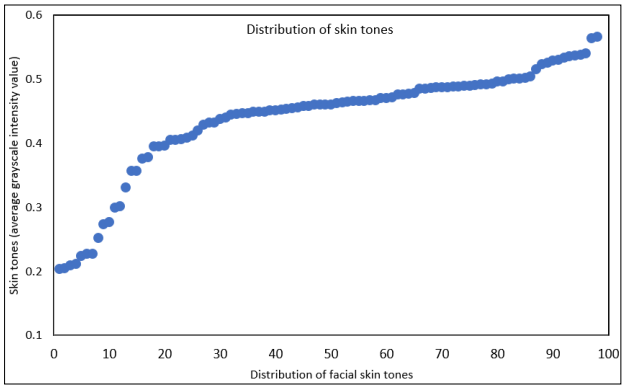


Figure 7: Distribution of skin tones represented in our data set ordered based on average greyscale intensities (scale 0–1, under the assumption that the illumination in the data set is constant) from low to high.

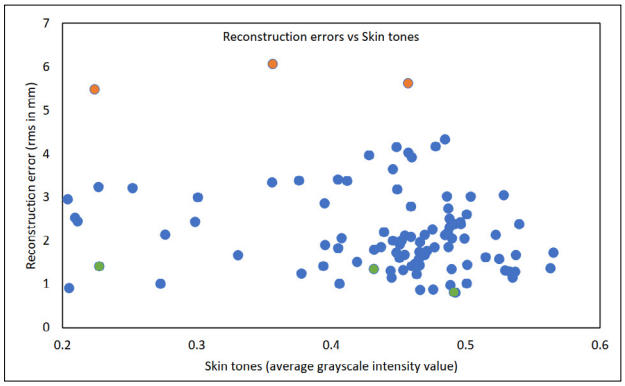


Figure 8: Relation between skin tone (measured as average greyscale intensity as in Figure 7) and the root mean squared (RMS) reconstruction error per mesh. The correlation between skin tone and reconstruction quality in this plot is -0.16 compared to the Basel Face Model (BFM)[27] error which was -0.37. The orange circles represent predictions with high errors across the skin tone while green circles show predictions with low errors across the skin tones. The highlighted examples are shown in Figure 9.

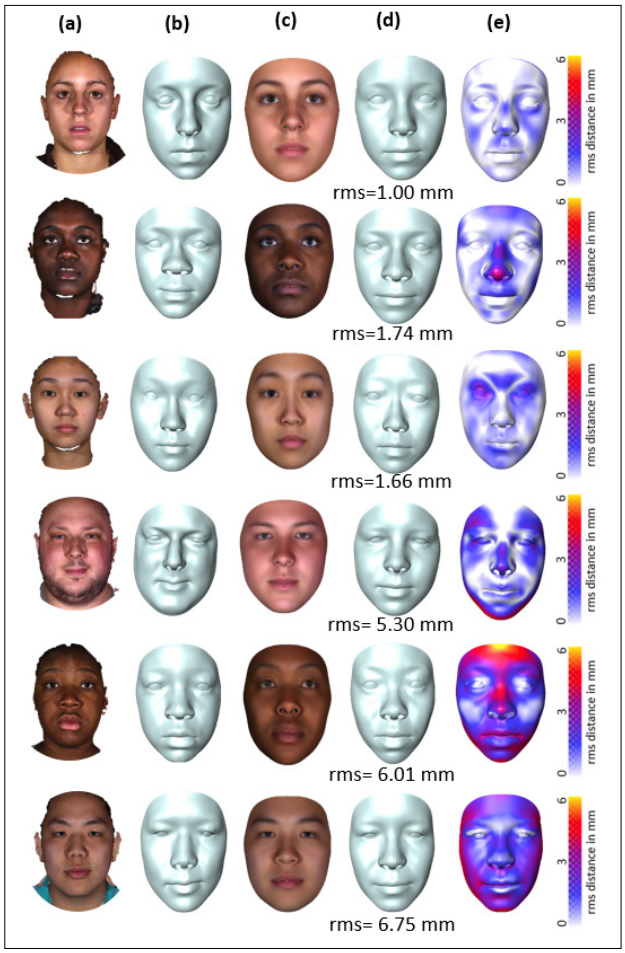


Figure 9: Examples of lowest and highest face predictions across skin tones. The first column (a) shows the target 2D face images (input). The second column (b) illustrates the ground truth face scans. The third column (c) represents the textured face reconstructions (output). The fourth column (d) indicates the predicted face shapes (output) with the root mean squared (RMS) value for different skin tones. The fifth column (e) presents the face surface colourmap comparisons.

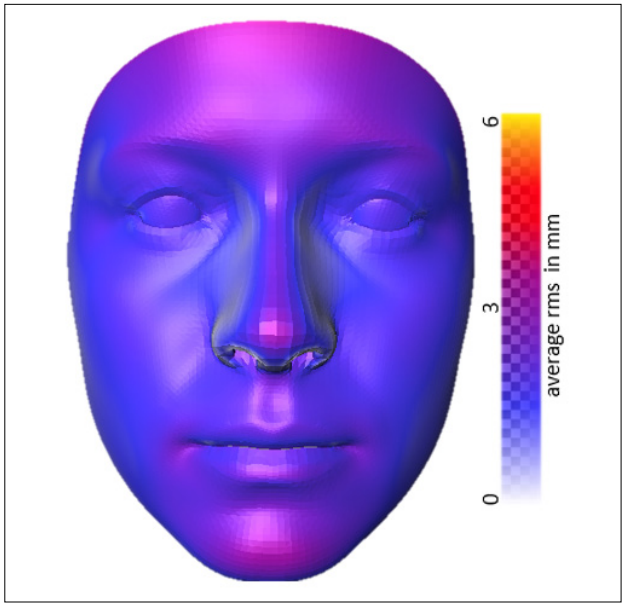


Figure 10: Visualising the average reconstruction error of the whole face data set.

3D surface distance measurements

During FAS facial phenotype assessments, distances between facial features on the face surface can be measured using either a physical instrument or a computer-assisted tool. These surface measurements are used to confirm the diagnosis of facial syndrome. A study by Douglas et al.[37] extracted facial features and performed measurements on the following face distances related to FAS: palpebral fissure length, inner canthal distance, outer canthal distance and interpupillary distance. However, these measurements were conducted on 2D stereo-photogrammetry images projected in 3D space. We extracted landmark points from our face reconstructions and can derive such measurements directly from our 3D reconstruction without manual interaction. We present reconstruction accuracy of those landmark points as well as distance reconstruction accuracy (measured in 3D) based on our 3D reconstructions compared to the ground truth 3D shapes.

Landmark estimation. Landmarks are essential when taking measurements on a face surface. We identified and selected a subset of 14 landmark points which are related to FAS facial phenotype assessments. These landmarks are described in a study by Mutsvangwa and Douglas[38]. The results of the landmark estimation errors were computed and are illustrated in Table 1. The landmark error was computed by measuring position distance from reference surface to the reconstructed surface. As shown in Table 1, 11 of 14 landmark errors were lower than 3.5 mm. The large standard deviations are mainly a result of the few outliers observed in Figure 8.

Table 1: Facial landmark estimation errors

Landmark ID	Landmark name	Mean RMS error (mm)	Standard deviation (mm)
0	Right outer canthus	2.98	1.78
1	Right inner canthus	2.68	1.52
2	Glabella	3.36	1.87
3	Left inner canthus	2.57	1.51
4	Left outer canthus	3.01	1.74
5	Ring alare	2.74	1.25
6	Pronasale	4.21	2.34
7	Left alare	2.70	1.31
8	Subnasale	3.41	1.81
9	Right crista philtre	3.25	1.79
10	Labiale superius	3.38	1.95
11	Left crista philtre	3.19	1.79
12	Right cheilion	3.52	1.74
13	Left cheilion	3.45	1.73

RMS, root mean square

Facial feature distances. The distance measurements characteristic to FAS facial features include the palpebral fissure length, outer canthal distance, and inner canthal distance, as illustrated in Figure 11b. The landmarks required for the distance measurements are described in Table 1 and illustrated in Figure 11a. The corresponding distances on the reconstructed 3D face surface and the 3D ground truth face scan were compared and the difference calculated. Table 2 shows the results of average absolute distance errors and their standard deviations for the palpebral fissure length, outer canthal distance, and inner canthal distance facial feature distances.

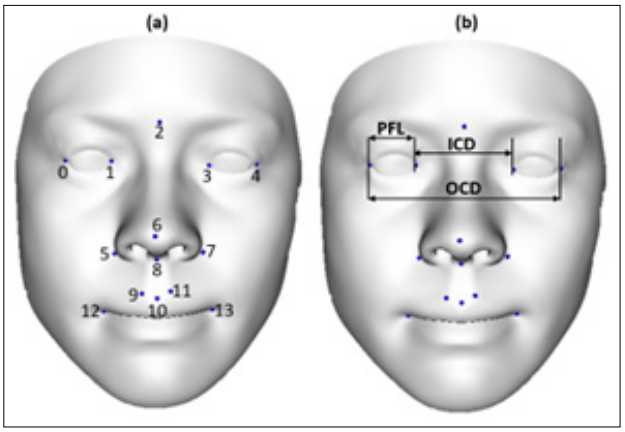


Figure 11: Landmarks (a) and inter-landmark distances (b) used in the study.

Table 2: A comparison of inter-landmark facial distances that would be considered for foetal alcohol syndrome facial analysis. Note that these estimations are based on normal adult faces.

Inter-landmark distance	Mean RMS error (mm)	Standard deviation (mm)
Palpebral fissure length	1.25	0.93
Outer canthal distance	3.56	3.08
Inner canthal distance	2.23	1.97

RMS, root mean square

Discussion

We constructed a 3D face model from 2D face scans and evaluated the accuracy of 3D face shape predictions from single images. The constructed morphable model of the face was evaluated for generalisation, specificity, and compactness parameters. The lowest generalisation error was 0.5 mm which suggests that the face shape model described the unseen face shapes well when given data outside the training set. The generalisation results of the face shape model compare well with other results found in the literature.[15,16] The specificity results of the face shape model are in the range of 13.2 mm to 14.5 mm, which is in the common ranges for specificity.[15] The compactness results of our face shape model indicate that more than 90% of the variability in the training set is retained with just 20 principal components and this compares well with other results in the literature. For example, Booth et al.[16] report that the first 40 principal components retained more than 90% of variability in their training set. Overall, our morphable model construction and evaluation seem successful.

The numerical average reconstruction error between the reconstructed face shapes and the corresponding ground truth face shapes in our data set was 2.62 mm. These findings are comparable to other results in the literature. For example, Zollhofer et al.[39] compared reconstructed face surfaces obtained from 3D face scans via a Kinect sensor to ground truth face scans, reporting an average deviation of about 2 mm. Additionally, Feng et al.[40] reported a root mean square error of 2.83 mm from surface comparisons between the predicted 3D face meshes and the corresponding ground truth 3D face scans.

For FAS facial phenotype assessment, we are interested in specific regions of the face such as the eyes, the midface, the upper lip, and the philtrum. These regions provide cues to clinicians when examining the FAS facial phenotype. The whole face surface reconstruction was examined using the colourmap surface comparisons shown in Figures 6, 9, and 10. From visual observation, the midface, the philtrum, and the regions around the eyes show lower levels of surface variation, as represented by the surface error colourmaps, whereas the upper lip areas show slight surface differences. The lower levels of surface variation around the eyes,

the philtrum, and the midface could be attributed to the ease of identifying landmarks in those regions. From Figures 9 and 10, we observe that regions of the face which are not relevant to FAS facial phenotype analysis are most affected. A study by Hammond et al.[41] suggested that visual inspections of the 3D surfaces using heat maps can delineate and discriminate facial features. In Figure 8, we find that the reconstruction quality in our data set is not affected by skin tones. On top of the heatmap representation, we also evaluated the landmarks and distances previously explored for the facial phenotype in FAS, as shown in Tables 1 and 2, respectively. We show accuracies in a minimum range of 2.57 mm for landmarking errors and 1.25 mm for distance errors. Similar results are reported in the literature. Regarding landmark localisation error, a study by Sukno et al.[42] reported an average error per landmark of below 3.4 mm. For inter-landmark distances, Douglas et al.[37] reported an average difference, between the manual and automated approaches, within 1 mm for palpebral fissure length, but with greater variations for outer canthal distance and inner canthal distance.

The highest face surface reconstruction errors belong to a relatively small set of 3D scans. Furthermore, the surface differences could imply that, during the model fitting phase of the reconstruction process, our statistical model did not completely capture all geometric cues in the 2D image of the face. We define a geometric cue as the information contained in a 2D image of the face, such as shading or contours.

Limitations and future research

Although we used a data set of scans of normal adult controls, with no known FAS indications, we assume that the framework is invariant to the data when built and applied to a population of interest. Ideally, training and test data sets would be collected from FAS and non-FAS control populations, with similar demographics. It is a challenge to access 3D data of individuals with FAS; however, the acquired face database (BU-3DFE) is very diverse. Future work could focus on reducing the reconstruction errors to acceptable clinical standards by collecting and analysing larger data sets, including more training data, especially from underrepresented populations. This would broaden the applicability of the morphable models of the face. To improve on the surface reconstruction performance, future developments could consider using multi-view 2D images of the face to provide more geometric cues during the model fitting of the face.

Conclusion

In this study, we aimed to evaluate whether an inverse graphics-based 3D from 2D reconstruction algorithm is suitable for acquiring 3D face data for FAS facial shape analysis. The reconstruction task was accomplished by fitting a 3DMM to a 2D image to recover a 3D face representation. Additionally, 3DMMs were built from a collection of 3D face scans with shape and texture information. We provided an evaluation performance of face reconstruction for future applications to FAS diagnosis. The resulting accuracies are promising for these future applications, even across different ethnicities.

Acknowledgements

This research was supported by the African Biomedical Engineering Mobility (ABEM) project, under the Intra-Africa Academic Mobility Scheme of the European Commission's Education, Audio Visual and Culture Executive Agency, and by the South African Research Chairs Initiative of the Department of Science and Innovation and the National Research Foundation of South Africa (grant no. 98788). B.E. is supported by a Postdoc Mobility Grant from the Swiss National Science Foundation (P400P2_191110).

Competing interests

We have no competing interests to declare.

Authors' contributions

All authors participated in the project conceptualisation. F.A. prepared the data, performed, and evaluated the experiments, and drafted the initial manuscript. B.E. performed and evaluated experiments and revised the manuscript. T.M. formulated the research goals and objectives, supervised the project, and provided financial support. All authors contributed towards the writing of the final manuscript.

References

1. Streissguth AP, Bookstein FL, Barr HM, Sampson PD, O'Malley K, Young JK. Risk factors for adverse life outcomes in fetal alcohol syndrome and fetal alcohol effects. J Dev Behav Pediatr. 2004;25(4):228–238. https://doi.org/10.1097/00004703-200408000-00002
2. Astley SJ, Clarren SK. A fetal alcohol syndrome screening tool. Alcohol Clin Exp Res. 1995;19(6):1565–1571. https://doi.org/10.1111/j.1530-0277.1995.tb01025.x
3. Hoyme HE, May PA, Kalberg WO, Kodituwakku P, Gossage JP, Trujillo PM, et al. A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: Clarification of the 1996 Institute of Medicine criteria. Pediatrics. 2005;115(1):39–47. https://doi.org/10.1542/peds.2004-0259
4. Clarren SK, Smith DW. The fetal alcohol syndrome. N Engl J Med. 1978;298(19):1063–1067. https://doi.org/10.1056/NEJM197805112981906
5. Astley SJ, Clarren SK. A case definition and photographic screening tool for the facial phenotype of fetal alcohol syndrome. J Pediatr. 1996;129(1):33–41. https://doi.org/10.1016/S0022-3476(96)70187-7
6. Astley SJ, Clarren SK. Diagnosing the full spectrum of fetal alcohol-exposed individuals: Introducing the 4-digit diagnostic code. Alcohol Alcohol. 2000;35(4):400–410. https://doi.org/10.1093/alcalc/35.4.400
7. Moore ES, Ward RE, Jamison PL, Morris CA, Bader PI, Hall BD. The subtle facial signs of prenatal exposure to alcohol: An anthropometric approach. J Pediatr. 2001;139(2):215–219. https://doi.org/10.1067/mpd.2001.115313
8. May PA, De Vries MM, Marais A-S, Kalberg WO, Adnams CM, Hasken JM, et al. The continuum of fetal alcohol spectrum disorders in four rural communities in South Africa: Prevalence and characteristics. Drug Alcohol Depend. 2016;159:207–218. https://doi.org/10.1016/j.drugalcdep.2015.12.023
9. Honrado CP, Larrabee WF Jr. Update in three-dimensional imaging in facial plastic surgery. Curr Opin Otolaryngol Head Neck Surg. 2004;12(4):327–331. https://doi.org/10.1097/01.moo.0000130578.12441.99
10. Lekakis G, Claes P, Hamilton GS, Hellings PW. Three-dimensional surface imaging and the continuous evolution of preoperative and postoperative assessment in rhinoplasty. Facial Plast Surg. 2016;32:88–94. https://doi.org/10.1055/s-0035-1570122
11. Blanz V, Vetter T. Face recognition based on fitting a 3D morphable model. IEEE Trans Pattern Anal Mach Intell. 2003;25(9):1063–1074. https://doi.org/10.1109/TPAMI.2003.1227983
12. Egger B, Smith WA, Tewari A, Wuhrer S, Zollhoefer M, Beeler T, et al. 3D morphable face models – past, present, and future. ACM Trans Graph. 2020;39(5):1–38. https://doi.org/10.1145/3395208
13. Blanz V, Vetter T. Morphable model for the synthesis of 3D faces. In: Proceedings of the ACM SIGGRAPH Conference on Computer Graphics. New York: ACM Press/Addison-Wesley Publishing Co.; 1999. p. 187–194. https://doi.org/10.1145/311535.311556
14. Booth J, Antonakos E, Ploumpis S, Trigeorgis G, Panagakis Y, Zafeiriou S, editors. 3D face morphable models "in-the-wild". In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017. p. 48–57. https://doi.org/10.1109/CVPR.2017.580
15. Dai H, Pears N, Smith WA, Duncan C, editors. A 3D morphable model of craniofacial shape and texture variation. In: Proceedings of the IEEE International Conference on Computer Vision. IEEE; 2017. p. 3085–3093. https://doi.org/10.1109/ICCV.2017.335
16. Booth J, Roussos A, Ponniah A, Dunaway D, Zafeiriou S. Large scale 3D morphable models. Int J Comput Vision. 2018;126(2–4):233–254. https://doi.org/10.1007/s11263-017-1009-7
17. Weiwei Y, editor. Fast and robust image registration for 3D face modeling. In: 2009 Asia-Pacific Conference on Information Processing; 2009 July 18–19; Shenzhen, China. IEEE; 2009. p. 23–26. https://doi.org/10.1109/APCIP.2009.142

18. Amberg B, Romdhani S, Vetter T, editors. Optimal step nonrigid ICP algorithms for surface registration. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition; 2007 June 17–22; Minneapolis, MN, USA. IEEE; 2007. p. 1–8. https://doi.org/10.1109/CVPR.2007.383165

19. Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065), Art. #20150202. https://doi.org/10.1098/rsta.2015.0202

20. Romdhani S, Vetter T. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); 2005 June 20–25; San Diego, CA, USA. IEEE; 2005. p. 986–993.

21. Paysan P, Knothe R, Amberg B, Romdhani S, Vetter T. A 3D face model for pose and illumination invariant face recognition. In: 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. 2009;70(80):296–301. https://doi.org/10.1109/AVSS.2009.58

22. Aldrian O, Smith WA. Inverse rendering of faces with a 3D morphable model. IEEE Trans Pattern Anal Mach Intell. 2013;35(5):1080–1093. https://doi.org/10.1109/TPAMI.2012.206

23. Schönborn S, Egger B, Morel-Forster A, Vetter T. Markov chain Monte Carlo for automated face image analysis. Int J Comput Vision. 2016;123(2):160–183. https://doi.org/10.1007/s11263-016-0967-5

24. Sanyal S, Bolkart T, Feng H, Black MJ, editors. Learning to regress 3D face shape and expression from an image without 3D supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019 June 15–20; Long Beach, CA, USA. IEEE; 2019. p. 7763–7772. https://doi.org/10.1109/CVPR.2019.00795

25. Feng Y, Feng H, Black MJ, Bolkart T. Learning an animatable detailed 3D face model from in-the-wild images. ACM Trans Graph. 2021;40(4), Art. #88. https://doi.org/10.1145/3450626.3459936

26. Yin L, Wei X, Sun Y, Wang J, Rosato MJ, editors. A 3D facial expression database for facial behavior research. In: 7th International Conference on Automatic Face and Gesture Recognition (FGR06); 2006 April 10–12; Southampton, UK. IEEE; 2006. p. 211–216.

27. Gerig T, Morel-Forster A, Blumer C, Egger B, Luthi M, Schönborn S, et al., editors. Morphable face models – an open framework. In: 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018); 2018 May 15–19; Xi'an, China. IEEE; 2018. p. 75–82. https://doi.org/10.1109/FG.2018.00021

28. Lüthi M, Forster A, Gerig T, Vetter T. Shape modeling using gaussian process morphable models. In: Zheng G, Li S, Székely G, editors. Statistical shape and deformation analysis: London: Academic Press; 2017. p. 165–191. https://doi.org/10.1016/B978-0-12-810493-4.00008-0

29. Lüthi M, Gerig T, Jud C, Vetter T. Gaussian process morphable models. IEEE Trans Pattern Anal Mach Intell. 2017;40(8):1860–1873. https://doi.org/10.1109/TPAMI.2017.2739743

30. Gower JC. Generalized Procrustes analysis. Psychometrika. 1975;40(1):33–51. https://doi.org/10.1007/BF02291478

31. Goodall C. Procrustes methods in the statistical-analysis of shape. J Roy Stat Soc B Met. 1991;53(2):285–339. https://doi.org/10.1111/j.2517-6161.1991.tb01825.x

32. Komalasari D, Widyanto MR, Basaruddin T, Liliana DY, editors. Shape analysis using generalized Procrustes analysis on active appearance model for facial expression recognition. In: 2017 International Conference on Electrical Engineering and Computer Science (ICECOS); 2017 August 22–23; Palembang, Indonesia. IEEE; 2017. p. 154–159. https://doi.org/10.1109/ICECOS.2017.8167123

33. Styner MA, Rajamani KT, Nolte L-P, Zsemlye G, Székely G, Taylor CJ, et al. Evaluation of 3D correspondence methods for model building. In: Taylor C, Noble JA, editors. Information Processing in Medical Imaging (IPMI 2003). Lecture Notes in Computer Science vol 2732. Berlin: Springer; 2003. p. 63–75. https://doi.org/10.1007/978-3-540-45087-0_6

34. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geosci Model Dev. 2014;7(3):1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

35. Voisin S, Page DL, Foufou S, Truchetet F, Abidi MA. Color influence on accuracy of 3D scanners based on structured light. In: Meriaudeau F, Niel KS, editors. Proceedings volume 6070 Machine Vision Applications in Industrial Inspection XIV. SPIE; 2006, #607009. https://doi.org/10.1117/12.643448

36. Grother P, Ngan M, Hanaoka K. Face recognition vendor test (FRVT). Part 2: Identification. Gaithersburg, MD: US Department of Commerce, National Institute of Standards and Technology; 2019. https://doi.org/10.6028/NIST.IR.8271

37. Douglas TS, Martinez F, Meintjes EM, Vaughan CL, Viljoen DL. Eye feature extraction for diagnosing the facial phenotype associated with fetal alcohol syndrome. Med Biol Eng Comput. 2003;41(1):101–106. https://doi.org/10.1007/BF02343545

38. Mutsvangwa T, Douglas TS. Morphometric analysis of facial landmark data to characterize the facial phenotype associated with fetal alcohol syndrome. J Anat. 2007;210(2):209–220. https://doi.org/10.1111/j.1469-7580.2006.00683.x

39. Zollhofer M, Martinek M, Greiner G, Stamminger M, Sussmuth J. Automatic reconstruction of personalized avatars from 3D face scans. Comput Animation Virtual Worlds. 2011;22(2–3):195–202. https://doi.org/10.1002/cav.405

40. Feng Z-H, Huber P, Kittler J, Hancock P, Wu X-J, Zhao Q, et al., editors. Evaluation of dense 3D reconstruction from 2D face images in the wild. In: 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018); 2018 May 15–19; Xi'an, China. IEEE; 2018. p. 780–786. https://doi.org/10.1109/FG.2018.00123

41. Hammond P, Hutton TJ, Allanson JE, Buxton B, Campbell LE, Clayton-Smith J, et al. Discriminating power of localized three-dimensional facial morphology. Am J Hum Genet. 2005;77(6):999–1010. https://doi.org/10.1086/498396

42. Sukno FM, Waddington JL, Whelan PF. 3-D facial landmark localization with asymmetry patterns and shape regression from incomplete local features. IEEE Trans Cybern. 2014;45(9):1717–1730. https://doi.org/10.1109/TCYB.2014.2359056

AUTHORS:
Kyra Scott[1]
Colin Tredoux[1]
Alicia Nortje[1]

AFFILIATION:
[1]Eyewitness and ACSENT Laboratories, Department of Psychology, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Kyra Scott

EMAIL:
kyra.em.scott@gmail.com

DATES:
Received: 23 Aug. 2021
Revised: 19 Nov. 2022
Accepted: 05 Dec. 2022
Published: 29 Mar. 2023

HOW TO CITE:
Scott K, Tredoux C, Nortje A. Evaluating the utility of facial identification information: Accuracy versus precision. S Afr J Sci. 2023;119(3/4), Art. #12067. https://doi.org/10.17159/sajs.2023/12067

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Leslie Swartz

KEYWORDS:
facial composites, facial descriptions, identification, eyewitness, face recognition

FUNDING:
National Research Foundation of South Africa; University of Cape Town

Evaluating the utility of facial identification information: Accuracy versus precision

Facial identification evidence obtained from eyewitnesses, such as person descriptions and facial composites, plays a fundamental role in criminal investigations and is regularly regarded as valuable evidence for apprehending and prosecuting perpetrators. However, the reliability of such facial identification information is often queried. Person descriptions are frequently reported in the research literature as being vague and generalisable, whilst facial composites often exhibit a poor likeness to an intended target face. This raises questions regarding the accuracy of eyewitness facial identification information and its ability to facilitate efficient searches for unknown perpetrators of crimes. More specifically, it questions whether individuals, blind to the appearance of a perpetrator of a crime (i.e. the public), can correctly identify the intended target face conveyed by facial identification information recalled from eyewitness memory, and which of the two traditional facial identification formats would be better relied upon by law enforcement to enable such searches. To investigate this, in the current study ($N=167$) we employed two metrics – identification accuracy and identification precision – to assess the utility of different formats of eyewitness facial identification information in enabling participants to correctly identify an unknown target face across three different formats: facial descriptions, facial composites and computer-generated description-based synthetic faces. A statistically significant main effect for the format of facial identification information on identification accuracy ($p<0.001$) was found, with a higher target identification accuracy yielded by facial descriptions in comparison to composites and description-based synthetic faces. However, the reverse relationship was established for identification precision, where composites and description-based synthetic faces enabled significantly greater precision in the narrowing down of a suspect pool than did facial descriptions, but did not necessarily result in the retainment of the intended target face ($p<0.001$).

Significance:

- This study highlights the relative importance of person descriptions in being as effective as, if not better than, facial composites in allowing for accurate identifications when solely relying upon eyewitness facial identification information to facilitate the search for unknown perpetrators.
- We introduce the metric of identification precision to evaluate the utility of facial identification information obtained by eyewitnesses.
- The study provides a novel approach to directly model facial composites based on a person description using traditional fourth-generation composite systems, thus producing a computer-generated description-based synthetic face that resembles a target face observed by an eyewitness.

Introduction

Eyewitness identification evidence (e.g. person descriptions and facial composites) refers to any information provided by an eyewitness, from memory, regarding the physical appearance of the perpetrator. Such evidence frequently plays a pivotal role in the criminal justice system[1,2], often impacting the outcome of criminal cases, especially in the absence of physical incriminating evidence, such as DNA or fingerprints, or still and video imagery, such as CCTV footage, enabling identification of the offender[3]. When there is a lack of incriminating evidence and suitable suspects, police officials regularly depend upon eyewitness accounts and descriptive evidence to facilitate wider searches amongst the public in order to apprehend unidentified perpetrators.[4,5] The general sentiment towards eyewitness identification evidence held by society and law enforcement is that it is a reliable source of evidence to correctly establish the identity of a perpetrator.[1,6-8] However, misidentifications and convictions of innocent people often occur when such convictions are primarily dependent upon this form of evidence.[9,10] Furthermore, researchers have repeatedly raised concerns surrounding the risks of solely relying upon eyewitness identification evidence to aid in the search and accurate identification of offenders, as it has been determined to be easily susceptible to suggestibility, bias and error.[11-14] Moreover, facial identification information recalled by witnesses may be particularly unreliable because of a frequently reported lack of specificity of facial descriptors[15,16] and a poor similarity of facial composites to intended target faces[8]. Nonetheless, evidence obtained from eyewitnesses is essential to criminal investigations, specifically in the absence of other incriminating evidence and in time-sensitive cases, where retrieving camera imagery may take too long.[17]

Law enforcement officers routinely collect eyewitness person descriptions of offenders[5], and indeed are expected to do so by South African law (see *R v Shekelele*[18]). Typically the first evidence gathered during a criminal investigation, person descriptions are verbal or written recollections of a perpetrator's appearance given by eyewitnesses.[19] Various interview procedures, which attempt to achieve maximum reliability and comprehensive recall of details[20], are employed by police officers to elicit person descriptions from eyewitness memory[11]. These include the use of person-feature checklists[21], the Cognitive Interview[22,23] and its variations such as the Holistic-Cognitive Interview[24] for facial-composite construction, the Person Description Interview[25], Self-Administered Interview[26] and free-recall descriptions[11]. Despite recalled descriptors achieving accuracy rates of over 80%[15,27-29], person descriptions are still

frequently perceived as vague and incomplete in terms of portraying an offender's appearance[1,8,15,16]. This is because witnesses tend to provide information regarding basic physical attributes (e.g. gender, age, height, build and clothing) instead of more distinct and identifiable attributes that are vital for person identification, such as inner-facial features (e.g. eye shape and colour, and relative size and shape of the nose, lips and mouth).[29] Moreover, eyewitness reporting of inner-facial features is typically less than 40% accurate.[29] Hence, facial composites are predominately relied upon by law enforcement to locate unidentified offenders.[5]

Facial composites are attempted visual depictions of a perpetrator's face, constructed by an eyewitness's recall with the aid of a software operator or sketch artist who guides the process.[30] Composites facilitate the search for suspects by providing an image of the perpetrator that it is hoped members of the public familiar with the perpetrator will recognise, so they can then report the individual in question.[5] Originally, composites were drawn by artists working alongside eyewitnesses to create a sketch of the offender.[6] This later evolved to the use of 'featural' systems (mechanical systems and computerised 'feature' systems), allowing police with less artistic expertise to facilitate the construction process for facial composites.[17] These systems enable witnesses to build a likeness of the perpetrator's face using a selection of set facial features.[6] However, the composites produced by such 'featural' systems tended to bear a poor similarity to the suspect[31], and so led to the introduction of 'holistic' fourth-generation composite systems, such as EvoFit[32], EFIT-V[33] and ID[34], which utilise holistic-configural facial processing[35]. Currently, the most widely used facial composite systems, fourth-generation composite systems, outperform previous sketch and 'featural' systems in enabling accurate identifications.[31] Nonetheless, facial composites still attain low correct identification naming rates, with laboratory studies documenting naming rates of less than 50% under forensically valid conditions.[36] Similarly, police utilising a fourth-generation composite system reported an identification naming rate of approximately 60% calculated on suspect arrests.[37]

The reliance on both person descriptions and face composites to search for unidentified offenders raises the question of which of these methods law enforcement should be emphasising for better identification accuracy. Several studies have sought to directly compare and determine the effectiveness of person descriptions and facial composites, and the consensus is that facial descriptions significantly outperform facial composites, resulting in fewer false identifications of wanted offenders.[38-41]

In the current study we sought to determine the effective utility of facial identification information obtained from eyewitnesses by assessing its effectiveness in facilitating efficient and accurate searches for unknown perpetrators of crimes. More specifically we considered whether individuals, blind to the appearance of a perpetrator of a crime (i.e. the public), could correctly identify the intended target face conveyed by the facial identification information recalled from eyewitness memory. This was attempted by evaluating the performance of three different formats of eyewitness facial identification information in enabling participants blinded to the appearance of simulated offenders to identify accurately and unambiguously the said persons. To do this, two performance metrics, identification accuracy and identification precision, were employed to measure differences in the effectiveness of different formats of eyewitness facial identification information to enable a correct identification. The three different formats of eyewitness facial identification information were (1) eyewitness facial descriptions, (2) traditional facial composites and (3) novel computer-generated description-based synthetic faces produced by fourth-generation facial composite software.

Given that facial descriptions have been shown to facilitate better recognition accuracy than facial composites[38-41], computer-generated description-based synthetic faces were used in the present study as a visual form of eyewitness facial descriptions. This produced a target face observed by an eyewitness whilst minimising human involvement in facial composite development, in an attempt to control and limit human-induced inaccuracies and potential eyewitness memory issues during composite construction, such as further memory decay, operator bias[35] and pattern-specific interference[21]. Description-based synthetic faces were also generated with the aid of fourth-generation facial composite software to standardise the creation medium for the visual facial identification information across the two different visual formats. Furthermore, by incorporating these computer-generated description-based synthetic faces, it becomes possible to explore whether facial descriptions do capture important information that facial composites typically overlook. The current study hypothesises that (1) facial descriptions will exhibit significantly higher target identification accuracy than both facial composites and computer-generated description-based synthetic faces, and (2) traditional facial composites will have significantly poorer target identification accuracy than computer-generated description-based synthetic faces.

Method

Design and setting

This study was conducted in two stages: (1) facial identification information was obtained from mock witnesses for use as the primary experimental material, and (2) the utility of the respective facial identification information was evaluated in terms of identification accuracy and identification precision. To obtain facial identification information, participants acting as eyewitnesses were exposed to two simulated offenders and asked to (1) describe the face and (2) construct a traditional facial composite based on (a) memory or (b) in-view observation. These descriptions and composites were then evaluated via an online experiment hosted on the survey platform Qualtrics. New participants, independent of stage one, were required to complete two identification tasks to assess the relative identification accuracy and identification precision of eyewitness facial descriptions and face composites.

Factors

A mixed repeated-measures design, consisting of $3 \times 2 \times 2$ randomised experimental cells, was implemented. The three factors were: (1) 'Format of Facial Identification Information' (within-subjects factor; description versus computer-generated description-based synthetic face versus facial composite), (2) 'Mode of Recall' (between-subjects factor; from memory versus in-view) and (3) 'Target' (between-subjects factor; Target A versus Target B). For this study, two simulated offenders' faces (target faces) were introduced to control for possible target bias (e.g. through differential facial distinctiveness).

Dependent variables

The utility of different formats of facial identification information was measured through two dependent variables: 'Identification Accuracy' and 'Identification Precision'. 'Identification Accuracy' was operationalised as correct identification of an intended target face from an array. 'Identification Precision' measured how efficient identification information was in narrowing down a pool of potential suspects from a starting face-matrix set of 24 faces.

Participants

The initial sample consisted of 169 volunteers recruited electronically with opportunistic sampling. It was established that no participants knew or recognised either of the simulated offenders. Two participants were excluded as they had not resided in southern Africa for at least the past 5 years (all target faces were South African). This criterion was imposed due to the sensitivity of facial identification, which can be negatively impacted by own group bias.[22] Thus, the final sample size was 167, with the majority (59.28%) being female participants. Mean age was 38.65 years (standard deviation (SD) = 15.55). Approximately 68% of the sample self-identified as white, 14% as coloured, 11% as Indian, 5% as black and 1% as Asian.

Materials

Face stimuli

The simulated offenders were two coloured South African men in their early 20s. The two identification tasks were constructed to assess identification accuracy and identification precision. To do so, separate 6×4 photographic face matrices were constructed for each simulated offender. These face matrices acted as the suspect pools, each

consisting of the target faces and 23 standardised filler faces that bore a resemblance to the simulated offenders. A modal facial description was produced for each simulated offender based on the most frequently mentioned facial attributes given by 17 individuals in delayed-matching, free-recall facial descriptions. These individuals did not participate in any other part of the study. Subsequently, 23 filler faces were selected from a database of coloured South African men's faces for each simulated offender's suspect pool, based on the modal facial descriptions. Facial photographs were presented in colour, with faces bearing neutral expressions. To ensure unbiased arrangements of photographs in the face matrices, placement of photographs was randomised.

Facial identification information

Facial identification information was gathered independently of the online evaluation tasks. The process used to gather the facial identification information is depicted in Figure 1. A total of 16 undergraduate psychology students were recruited, via convenience sampling, to act as mock witnesses and produce the required eyewitness facial identification information of the two simulated offenders. Individuals were recruited under the pretense of a 'tarot-card reading' study to mitigate potential priming and subject-expectancy effects.

To gather facial identification information from memory, the mock witnesses were randomly exposed to a live, in-person, 10-minute encoding to one of the two simulated offenders, who acted as a tarot-card reader. Exposure to the two simulated offenders during the encoding was counter-balanced between participants. Following an 8-minute distractor task, witnesses were required to give free-recall facial descriptions and complete a facial checklist, which was a modified version of the Aberdeen University Face Rating Schedule (FRS)[21], for the two simulated offenders. This consisted of 41 ratings of facial attributes on a five-point Likert scale. Subsequently, facial descriptions of the simulated offender (tarot-card reader) were elicited from witness memory, followed by another elicited facial description of the other simulated offender whilst they sat in view. A 15-minute distractor task followed the elicitation of facial descriptions before witnesses constructed facial composites for both simulated offenders. First, witnesses produced the facial composite of the tarot-card reader from memory, and then they produced an in-view facial composite of the other simulated offender. This process produced a total of 32 free-recall facial descriptions (16 from memory and 16 in view) and 32 facial composites (16 from memory and 16 in view). Refer to Figure 2 for examples of the facial identification information collected.

Composite software

A fourth-generation composite system, ID[34], was used to produce all visual facial identification information (i.e. facial composites and computer-generated description-based synthetic faces). ID is a contemporary eigenface composite construction software similar to EvoFit[32] and EFIT-V[33], which are currently utilised by police personnel[17]. The system promotes holistic-configural facial processing by presenting witnesses with an array of 12 synthetic faces and allowing for the repeated selection, morphing and blending of multiple faces together. This iterative process utilises underlying eigenfaces and evolutionary genetic algorithms, such as Population Incremental Learning and M-Choice, to yield new 'generations' of faces bearing a likeness to previous selected faces. Overall variation between generated faces in the arrays continuously lessens upon each iteration as formerly selected faces are combined to create new 'generations' until a synthetic face is produced that resembles the intended target face. Witnesses are also capable of altering individual facial features and featural spacing.

Synthetic description-to-face generation

As for the traditional facial composites, ID was also used as the primary system to generate the description-based synthetic faces. This enabled standardisation across visual facial identification information formats in the creation medium so that the utility of eyewitness facial identification evidence could be better assessed. Computer-generated description-based synthetic faces were produced by initially training a front-end multivariate regression model to model the underlying eigenface coefficients required by ID to construct facial composites. To create this model, an independent online survey was circulated to a total of 72 individuals that required them to rate ID-generated facial composites using the modified FRS (checklist rating schedule). Sixty composites were rated in total, by between five and seven raters. An average rating was computed for each composite on each of the 41 dimensions in the FRS, thus averaging out differences between raters on all dimensions. The composite faces were created using an appearance model that had 138 eigenfaces (basis dimensions) underlying it (see Tredoux et al.[34] for a description of how ID works). In other words, each of the 60 synthetic or composite faces had a known set of coefficients that created the composite face from the underlying eigenfaces. The perceptual ratings on the FRS were then entered into a multivariate regression model, in which each of the underlying eigenfaces in the ID system was modelled by ratings on the perceived (FRS) face dimensions. In other words, a predictive model was built that allowed us to generate coefficient values from ratings, and the coefficient values could in

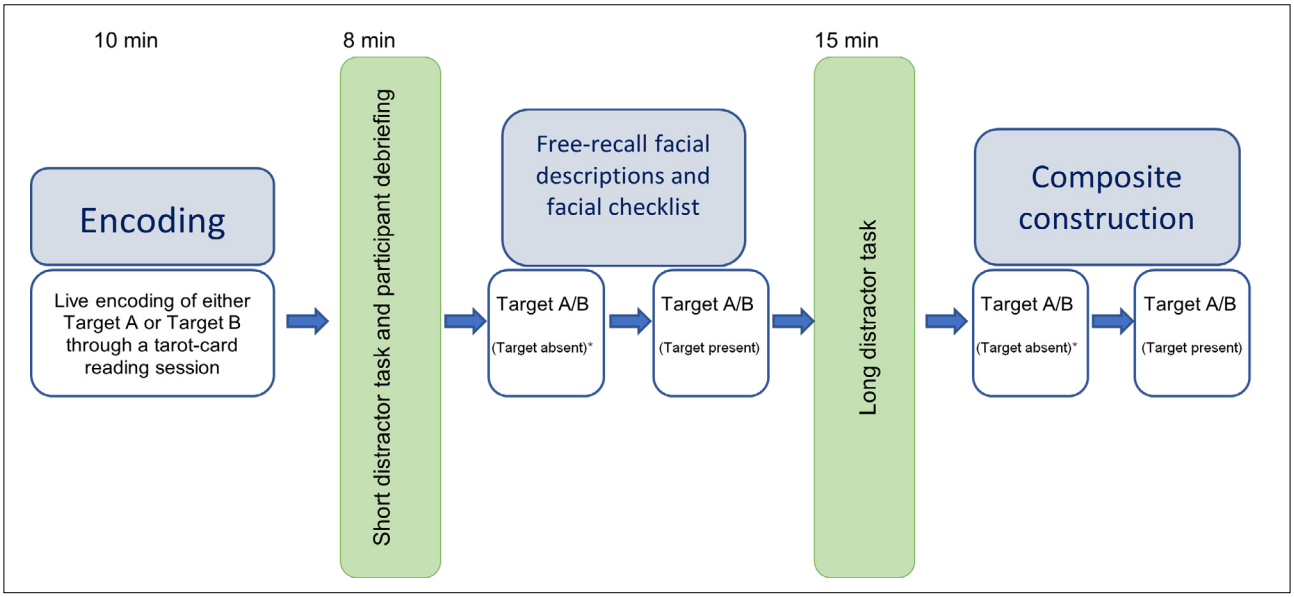


Figure 1: Procedure for gathering facial identification information. Target encoding was counterbalanced between individuals. Target-absent condition was used for the simulated offender encoded during the tarot-card reading session. The target-present condition was used as an in-view control of strong encoding.

turn be used to generate synthetic faces. FRS facial descriptions obtained for the two simulated offenders were then entered into the multivariate regression model and the relevant description-based synthetic faces were produced in ID, based on the model's output. This created a total of 32 computer-generated description-based synthetic faces (16 from memory and 16 in view). Refer to Figure 2 for an example of a computer-generated description-based synthetic face.

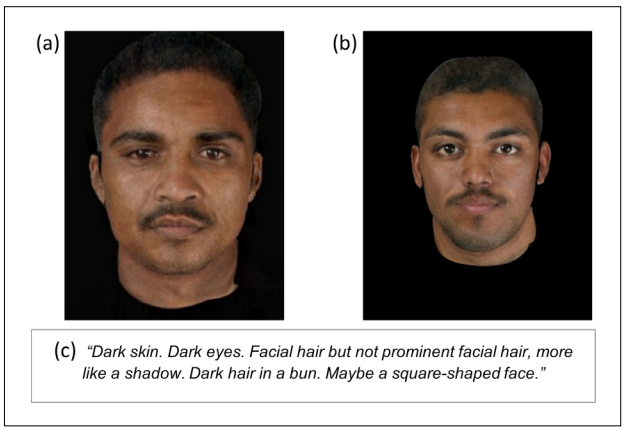


Figure 2: Samples of gathered facial identification information: (a) computer-generated synthetic face produced using an in-view description; (b) in-view facial composite constructed in ID; and (c) in-view facial description.

Procedure

To assess the utility of the three sources of facial identification information, a ranking task and a set-reduction task were conducted online to test identification accuracy and identification precision, respectively. Demographic information was gathered virtually from participants before they commenced the online tasks. This was done to measure the potential impact of own race bias[42] on facial identification results.

Participants were randomly assigned to one of four groups based upon 'Target' and 'Mode of Recall'. Hence, for both tasks they were exposed to only one of the two target faces and facial identification information produced either in view or from memory. However, all participants were presented with at least one instance of each source of eyewitness facial identification information (facial description, facial composite and computer-generated description-based synthetic face). Order of exposure to the different formats of facial identification information was randomised to control for order effects. Additionally, facial identification information was arbitrarily allocated to participants for each respective format.

Jointly presented with the 6 × 4 face matrix and the relevant facial identification information, participants were required first to complete a set-reduction task and then a ranking task for each format of facial identification information. The set-reduction task aimed to evaluate identification precision by asking participants to eliminate all faces from the face matrix that were not deemed to bear a resemblance to the provided facial identification information, resulting in a reduced subset of faces. The ranking task assessed identification accuracy by requiring participants to rank the 24 faces in the face matrix from least to most likely to resemble the given facial identification information.

Between the two identification tasks for each format of facial identification information, participants engaged in a short 3-minute distractor task, mitigating potential interference effects. See Figure 3 for an overview of the procedure.

Data analyses

All statistical analyses were completed in SPSS[434], with the alpha value set to 0.05 as per convention[44]. Analyses commenced with descriptive summaries of each variable. Further examination of group differences for 'Identification Accuracy' and 'Identification Precision' across formats of facial identification information and mode of recall (from memory versus in view) were carried out via a series of mixed analyses of variance (ANOVA). Analysis of the data pertaining to 'Identification Accuracy' found strong negative skewing of the standardised residual distributions across most experimental cells. To correct for this, a '$\ln(25 - X_i)$' transformation was applied to the data. Further assessment of the transformed data also revealed significant outliers[45] ($n=16$), which were subsequently omitted. Outliers were identified as transformed standardised residuals greater than 2.30 or less than −2.30.[46,47] The Greenhouse–Geisser estimate ($\varepsilon=0.79$) was used in the interpretation of the ANOVA results relating to 'Identification Accuracy' to correct degrees of freedom, as Mauchly's test of sphericity was found to be violated ($\chi^2(2)=46.05$, $p<0.001$). Visual assessment of the standardised residual distributions for 'Identification Precision' also found the data to be strongly skewed and thus an appropriate two-step transformation[47] was applied to correct for this. The Greenhouse–Geisser estimate ($\varepsilon=0.85$) was also used in the interpretation of the ANOVA results for 'Identification Precision', as Mauchly's test of sphericity indicated violation of the assumption of sphericity ($\chi^2(2)=32.25$, $p<0.001$).

Coding of dependent variables

'Identification Accuracy' was scored as a value between 1 and 24, determined by the positional rank a participant placed on the target face of the simulated offender during the ranking task. Higher scores are indicative of a greater accuracy than lower scores, with a score of 24 reflecting an accuracy of 100%. 'Identification Precision' was coded as the size of the remaining suspect pool after the set-reduction task took

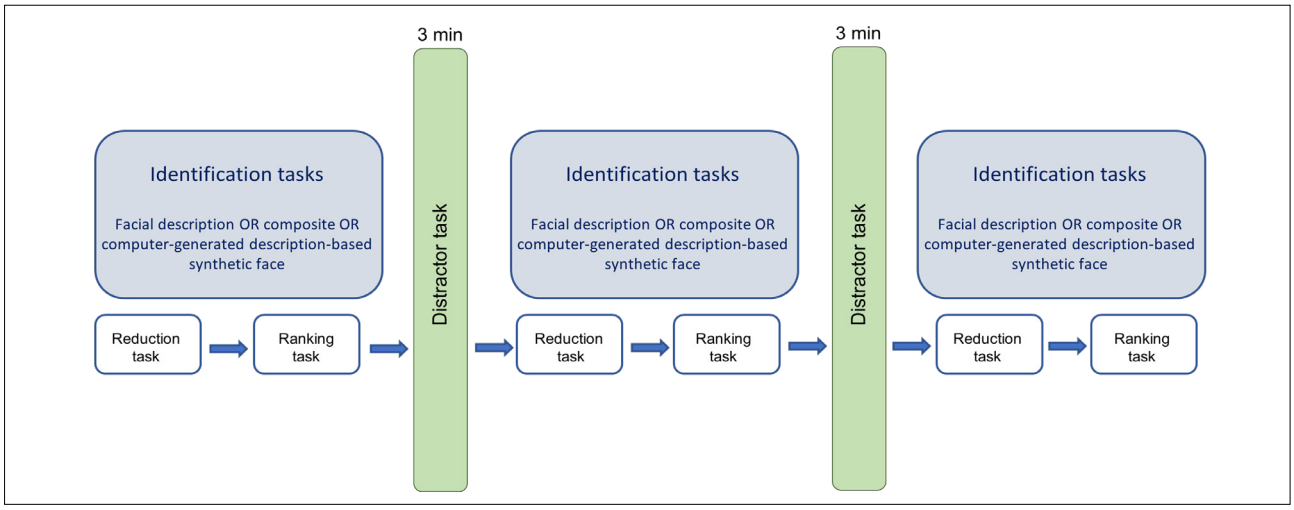


Figure 3: Procedure for testing identification accuracy and identification precision of facial identification information. The order of presentation of the type of facial identification information was counterbalanced.

place to narrow down the face matrix. This variable was measured as a value ranging between 0 and 24, where smaller set sizes related to higher levels of identification precision than larger set sizes.

Results

Identification accuracy

Descriptive statistics of identification accuracy across the different experimental conditions are reported in Table 1. On average, facial descriptions were able to facilitate the highest identification accuracy (mean = 16.74 (SD=6.87)). Facial composites (mean = 12.38 (SD=6.77)) and computer-generated description-based synthetic faces (mean = 11.05 (SD=5.99)) achieved lower identification accuracy scores. This resulted in facial descriptions performing on average, 18.17% and 23.71% better in accurately identifying a simulated offender than facial composites and computer-generated description-based synthetic faces, respectively. This claim in performance difference is further supported by the statistical analysis that follows.

Table 1: Descriptive statistics for identification accuracy (N=167)

	Target A		Target B		Total
	Memory (n=37)	In view (n=39)	Memory (n=51)	In view (n=40)	
Facial description	16.89 (5.88)	16.41 (5.69)	16.71 (8.05)	16.98 (7.36)	16.74 (6.87)
Facial composite	12.95 (5.65)	15.00 (6.47)	10.00 (7.21)	12.33 (6.60)	12.38 (6.77)
Computer-generated description-based synthetic face	12.97 (6.37)	12.38 (5.05)	9.08 (5.36)	10.50 (6.57)	11.05 (5.99)

Note: Identification accuracy was scored from 1 to 24, where higher scores indicate better identification accuracy.

Values shown are mean (standard deviation).

A summary of the results from the 3 × 2 × 2 mixed-designs ANOVA are reported in Table 2. A significant main effect of the type of facial identification information format on identification accuracy was indicated: $F(1.57, 231.40) = 56.79$, $p<0.001$, $\eta_p^2 = 0.28$. Post-hoc analysis, adopting a Bonferroni adjustment as a correction for multiple significance testing, found facial descriptions (mean = 1.95 (standard error (SE)=0.07), 95% confidence interval (CI) [1.81, 2.08]) allow for significantly greater identification of an intended target face than facial composites (mean = 2.50 (SE=0.05), 95% CI [2.42, 2.59]) or computer-generated description-based synthetic faces (mean = 2.66 (SE=0.03), 95% CI [2.60, 2.73]), which tended to result in poorer identification accuracy. However, no significant difference in identification accuracy was revealed between facial composites and computer-generated description-based synthetic faces.

Interpretation of interactions found a significant interaction between 'Mode of Recall' and 'Format of Facial Identification Information', with facial composites produced from memory yielding significantly poorer identification accuracy than facial composites generated in view of a simulated offender: $F(1, 147)=8.99$, $p=0.003$. Facial descriptions and computer-generated description-based synthetic faces performed similarly in relation to identification accuracy, regardless of 'Mode of Recall'. See Figure 4 for an illustration of the results. 'Format of Facial Identification Information' and 'Target' showed a relatively small two-way interaction effect: $F(1.57, 231.40)=7.06$, $p=0.003$. Further analyses of this two-way interaction, via simple main effects and pairwise comparisons, revealed a significantly lower identification accuracy for all facial composites of Target B over all facial composites of Target A: mean = 0.30 (SE = 0.09), $p=0.001$, 95% CI [0.12, 0.48]; $F(1, 147)=11.19$, $p=0.001$. A similar result with respect to poorer identification accuracy on Target B over Target A was also found for the computer-generated description-based synthetic faces: mean = 0.22 (SE=0.06), $p=0.001$, 95 CI [0.10, 0.35]; $F(1, 147)=12.61$, $p=0.001$. No significant difference was indicated between facial descriptions of the two different target faces: mean = 0.19 (SE=0.14), $p=0.17$, 95% CI [–0.08, 0.46]; $F(1, 147)=1.90$, $p=0.170$ (Figure 5).

Identification precision

Computer-generated description-based synthetic faces enabled the greatest narrowing down of the suspect pool by having the fewest remaining faces in the reduced face-matrix set (mean = 7.32 (SD=6.07)), resulting in an overall better identification precision than the other formats of facial identification information. Facial composites were revealed to perform only slightly worse than computer-generated description-based synthetic faces (mean = 7.40 (SD=6.19)). However, facial descriptions, on average, yielded the most retained faces within the face-matrix set (mean = 9.28 (SD=7.47)), leading to the worst identification precision amongst the different formats of facial identification information. Full descriptive statistics are shown in Table 3.

Table 2: Summary of ANOVA tests of between-subject effects and within-subject effects for identification accuracy (N=151)

Source	df	Sum of squares	Mean square	F	p	η_p^2
Between subjects	150.00	25.820				
Target	1.00	0.470	0.47	3.30	0.071	0.02
Mode of Recall	1.00	0.350	0.35	2.49	0.117	0.02
Target × Mode of Recall	1.00	0.000	0.00	0.00	0.958	0.00
Error	147.00	20.770	0.14			
Within subjects	237.68	157.800				
Facial Identification Information Format	1.57	41.980	26.67	56.79	<0.001**	0.28
Facial Identification Information Format × Target	1.57	5.220	3.32	7.06	0.003*	0.05
Facial Identification Information Format × Mode of Recall	1.57	1.770	1.13	2.40	0.106	0.02
Facial Identification Information Format × Target × Mode of Recall	1.57	0.170	0.11	0.23	0.741	0.00
Error	231.40	108.660	0.47			
Total	387.68	134.480				

*p<*0.05; **p<*0.001. All listed p-values are two-tailed; df, degrees of freedom; η_p^2, partial eta squared*

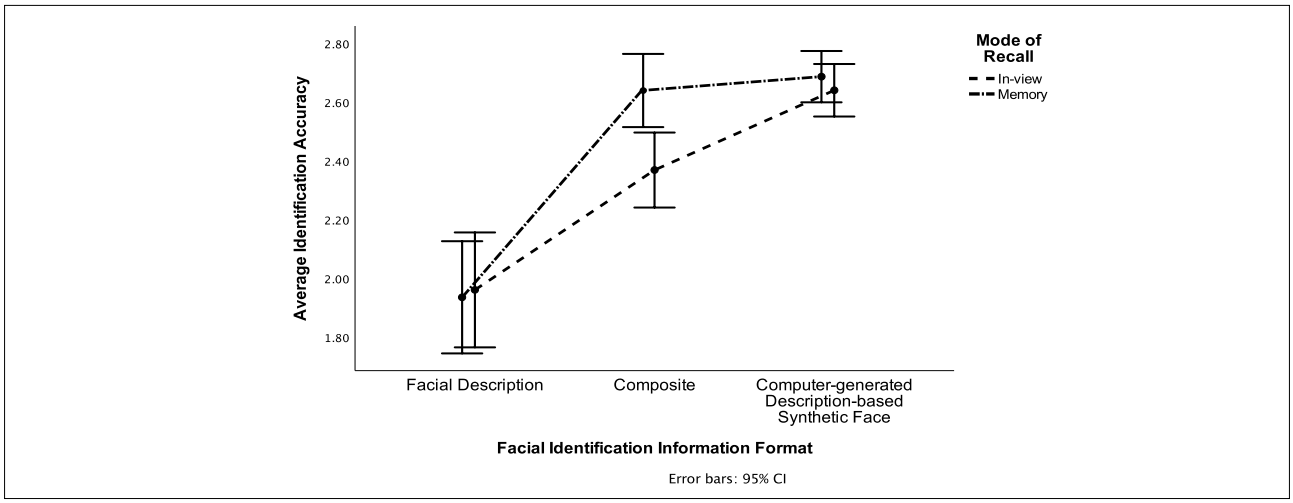


Figure 4: Interaction effect for the type of facial identification information format on mode of recall.

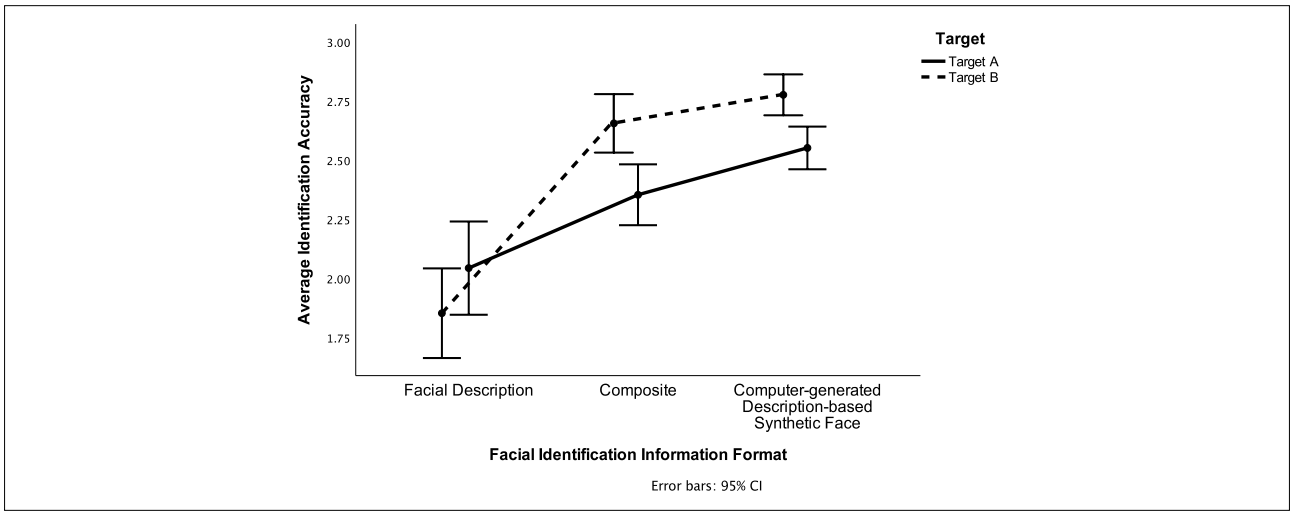


Figure 5: Interaction effect for the type of facial identification information format on target.

Table 3: Descriptive statistics for identification precision ($N=167$)

	Target A		Target B		Total
	Memory ($n=37$)	In view ($n=39$)	Memory ($n=51$)	In view ($n=40$)	
Facial description	9.22 (7.01)	9.79 (7.22)	7.71 (7.13)	10.85 (8.37)	9.28 (7.47)
Facial composite	6.46 (6.19)	7.64 (6.54)	7.39 (6.00)	8.02 (6.22)	7.40 (6.19)
Computer-generated description-based synthetic face	7.43 (6.37)	6.23 (5.64)	7.57 (6.01)	7.98 (6.36)	7.32 (6.07)

Note: Identification accuracy was scored from 1 to 24, where higher scores indicate better identification accuracy.
Values shown are mean (standard deviation).

As no significant main effects or interactions were indicated in relation to 'Target', we collapsed across this factor. A 3 × 2 mixed-designs ANOVA ('Facial Identification Information Format' × 'Mode of Recall') was run on the identification precision outcome. These results are shown in Table 4. A significant main effect of the format of facial identification information in relation to identification precision was revealed: $F(1.70, 427.18)=9.48$, $p<0.001$, $\eta_p^2=0.05$.

Table 4: Summary of ANOVA tests of between-subject effects and within-subject effects for identification precision ($N=167$)

Source	df	Sum of squares	Mean square	*F*	*p*	η_p^2
Between subjects	166.00	12 742.06				
Mode of Recall	1.00	66.09	66.09	0.86	0.355	0.01
Error	165.00	12 675.97	76.82			
Within subjects	283.41	7958.47				
Facial Identification Information Format	1.70	427.18	251.72	9.48	<0.001*	0.05
Facial Identification Information Format × Mode of Recall	1.70	94.83	55.88	2.10	0.132	0.01
Error	280.01	7436.46	26.56			
Total	449.41	20 700.53				

**p<0.05; **p<0.001. All listed p-values are two-tailed; df, degrees of freedom; η_p^2, partial eta squared*

Further post-hoc analysis of this effect, adopting a Sidak adjustment as all assumptions were met, determined that facial descriptions (mean = 9.45 (SE=0.55), 95% CI [8.36, 10.53]) resulted in a significantly worse identification precision than both facial composites (mean = 7.54 (SE=0.47), 95% CI [6.61, 8.46]) and computer-generated description-based synthetic faces (mean = 7.44 (SE=0.46), 95% CI [6.53, 8.34]). Thus, producing a mean difference of 1.91 (SE=0.57, p=0.003, 95 CI [0.54, 3.28]) and 2.01 ((SE=0.58, p=0.002, 95% CI [0.61, 3.41]) between facial descriptions and the other two formats of facial identification information, respectively. No significant difference was established between facial composites and computer-generated description-based synthetic faces in relation to the number of faces eliminated from the face-matrix set. Thus, on average, the use of facial descriptions resulted in the largest groups of suspects after reduction when compared with groups of suspects reduced using composite faces or computer-generated synthetic faces derived from descriptions.

Discussion

The purpose of the present research was to evaluate the utility of eyewitness facial identification information by examining whether individuals, blind to the appearance of a perpetrator of a crime, could accurately and precisely identify the target face conveyed by eyewitness facial identification evidence. This was achieved by independently gathering facial identification information from mock witnesses and then conducting an online experiment, consisting of a sorting task and a set-reduction task, to assess the eyewitness facial identification information in relation to identification accuracy and identification precision, respectively.

In line with previous studies[38-40], and the current hypotheses, the results indicate that facial descriptions outperformed facial composites and computer-generated description-based synthetic faces for identification accuracy. However, facial composites and computer-generated description-based synthetic faces were found to achieve similar rates of identification accuracy, not supporting the original hypothesis that computer-generated description-based synthetic faces would outperform facial composites. In relation to identification precision, facial composites and computer-generated description-based synthetic faces enabled significantly higher levels of precision in narrowing down a suspect pool than facial descriptions did. This inverse relationship between identification accuracy and identification precision could be accredited to numerous factors.

Differing description modalities (i.e. visual versus written) could account for some of the observed identification differences. Facial composites and computer-generated description-based synthetic faces convey facial information visually to enable the search and identification of an unknown target face, which is also a visual stimulus. This allows a straight comparison between the ability of the two visual stimuli to facilitate an identification. Written facial descriptions, on the other hand, convey facial information non-visually, making comparisons between this identification information and the other stimuli less direct.[19] This may account for the higher levels of identification precision achieved for facial composites and computer-generated description-based synthetic faces. Human observers, blind to the true appearance of a target, might well accept a visual facial likeness as an absolute depiction of a target face instead of an approximate, vague impression.[21,40] This may encourage the search for an exact match to the facial composite or computer-generated description-based synthetic face in a set of faces. Moreover, this may consequently produce more precise identifications, as individuals invoke stricter judgement and selection criteria of a chosen target face, resulting in more potential suspects being eliminated from a suspect pool, but not necessarily resulting in better identification accuracy. In comparison, facial descriptions convey non-visual facial information, which requires individuals to internally construct a mental image of the reported target face before searching for it. This could produce more subjective interpretations of a target, potentially leading to lower levels of congruency with a chosen target face as a less stringent judgement and selection criterion is invoked. Ultimately, this might produce less precise identifications as individuals eliminate fewer potential suspects from a suspect pool.

Despite finding that facial descriptions were associated with a poorer filtering ability, they did enable the most accurate identifications in comparison with facial composites and computer-generated description-based synthetic faces. This finding is in line with those of previous studies in which facial descriptions have been shown to outperform facial composites in enabling accurate identification of intended targets. Modality-specific interference and high cognitive loads induced by transferring an internalised, visual representation of an encoded target face into an externalised visual facial composite have been suggested as potential reasons for why facial composites yield poorer identification accuracy than facial descriptions.[21,40] However, we propose some alternative reasons that may account for the differences seen in identification accuracy amongst the differing formats of facial identification information.

Facial descriptions enable less facial information to be conveyed in comparison to visual facial identification information formats, as not all facial features and aspects are necessarily described by an eyewitness. Visual forms of facial identification, such as face composite images, require the inclusion of all facial details to illustrate the face of a perpetrator, regardless of whether the eyewitness is confident regarding all aspects of the face. The necessary inclusion of more information in facial composites and computer-generated description-based synthetic faces creates a higher probability of incorporating erroneous details. This enables greater subjective identification precision (set reduction) in third parties, as a greater number of comparisons can be made against a potential target and the visual facial identification information. However, the increased probability of error-prone facial information increases the chances of poorer identification accuracy rates. Facial descriptions implicitly allow witnesses to omit facial features and details regarding the suspect's face that are uncertain to them. This is likely to lower the probability of error within a facial description, allowing for better identification accuracy. Facial descriptions also allow eyewitnesses to place more emphasis upon facial attributes that they are more confident of, whilst less certain details can remain generalisable through vague or ambiguous word choice, or simple omission. Facial composites and computer-generated description-based synthetic faces, on the other hand, do not allow for emphasis on specific facial features. If better composite systems become available which address this difference, then facial composites could potentially outperform facial descriptions as they already outperform in terms of identification precision, which currently enables better narrowing down of a suspect pool than facial descriptions, although does not always lead to the correct identification of a perpetrator.

The significant difference between identification accuracy targets set for facial composites and for computer-generated description-based synthetic faces may reflect limitations in the ability of the composite software (ID) to adequately capture the visual likeness of the one simulated offender's face. This speculation is justified by the lack of a significant difference being established between facial descriptions of the two simulated offenders, which removes the possibility of facial distinctiveness inflating identification accuracy for the one simulated offender for the facial composites and computer-generated description-based synthetic faces.[48]

Facial composites constructed in view of the simulated offender enabled better identification accuracy than facial composites constructed from memory. This finding was in line with previous research that proposes constructing facial composites from memory enables memory decay and potentially other factors that can negatively impact accurate portrayal of a perpetrator as a result of misleading information leaking into the construction of the facial composite.[49]

Limitations and future research directions

From this study, it is reasonable to assume that facial descriptions will always be lacking in identification precision because of the lower level of facial information conveyed in this format, as it will always be near impossible to verbally describe all aspects of a face. Visual facial identification information, on the other hand, appears to provide a higher level of identification precision because of the amount and depth of

facial information conveyed by a visual likeness of a face, even if much of this information is inaccurate. If facial composites and computer-generated description-based synthetic faces can improve the accuracy of features in the visual medium, they may consistently outperform facial descriptions in identification accuracy. This would further increase the ecological validity of the use of facial composites by the criminal justice system to facilitate the search for unidentified perpetrators.

Given recent advancements in generative models, such as generative adversarial networks[50,51] and diffusion networks[52], as well as conditionalised text-to-image models[53-57], it may be feasible to increase the accuracy of computer-generated description-based synthetic faces. These networks could be utilised to fill in features that the eyewitness is not certain of by constructing the feature based on what is statistically most likely. Furthermore, a composite software that blurs facial features based on the confidence of the eyewitness's memory of a perpetrator's face could reduce the probability of erroneous facial information being introduced. This would lead to increased identification accuracy but at the cost of identification precision. This approach may allow composite operators to scale identification precision against identification accuracy. We are currently conducting research on both aspects.

Conclusion

In this study, the utility of varying formats of eyewitness facial identification information were evaluated in relation to identification accuracy and identification precision. Verbal (or written) facial descriptions from eyewitnesses were found to facilitate the most accurate identification of a conveyed offender from a pool of suspects, by individuals blind to the appearance of the offender. On the other hand, visual formats of eyewitness facial identification information (i.e. facial composites and computer-generated description-based synthetic faces) enabled the most precise narrowing down of a suspect pool, but at the cost of lowering identification accuracy.

Ethical considerations

Ethical clearance to conduct the study was granted by the University of Cape Town's Ethics Review Committee of the Faculty of Humanities (ethical clearance reference no. PSY2019-051). All aspects of the study were ethically approved. All participation was voluntary with informed consent given before engagement with any aspect of the study. Additionally, all participants were informed of their right to withdraw from the study at any stage without justification or penalty. Confidentiality and anonymity of participants were ensured by organising data according to randomly allocated participant numbers as well as encrypting and password-protecting all electronically stored data.

Acknowledgements

We thank the National Research Foundation of South Africa and the University of Cape Town for their financial support.

Competing interests

We have no competing interests to declare.

Authors' contributions

All authors were responsible for the conceptualisation and methodology of the study. K.S. was the primary researcher for the study, whilst C.T. and A.N. provided student supervision and critical feedback throughout the entire process. K.S. conducted the data collection and wrote the original draft of the manuscript. C.T. reviewed the manuscript draft and aided in subsequent writing revisions.

References

1. Brown C, Lloyd-Jones TJ, Robinson M. Eliciting person descriptions from eyewitnesses: A survey of police perceptions of eyewitness performance and reported use of interview techniques. Eur J Cogn Psychol. 2008;20(3):529–560. https://doi.org/10.1080/09541440701728474
2. McQuiston-Surrett D, Topp LD, Malpass RS. Use of facial composite systems in US law enforcement agencies. Psychol Crime Law. 2006;12(5):505–517. http://dx.doi.org/10.1080/10683160500254904
3. Keval HU, Sasse MA. Can we ID from CCTV? Image quality in digital CCTV and face identification performance. In: Agaian SS, Jassim SA, editors. Mobile multimedia/image processing, security, and applications. SPIE; 2008. p. 189–203. https://doi.org/10.1117/12.774212
4. Laughery KR, Duval C, Wogalter MS. Dynamics of facial recall. In: Aspects of face processing. Dordrecht: Springer; 1986. p. 373–387. https://doi.org/10.1007/978-94-009-4420-6_40
5. Kovera MB, Penrod SD, Pappas C, Thill DL. Identification of computer-generated facial composites. J Appl Psychol. 1997;82(2):235–246. http://dx.doi.org/10.1037/0021-9010.82.2.235
6. Davies G, Valentine T. Facial composites: Forensic utility and psychological research. In: Lindsay RCL, Ross DF, Read JD, Toglia MP, editors. Handbook of eyewitness psychology. New York: Routledge; 2007. p. 59–83. https://doi.org/10.4324/9781315805535
7. Gabbert F, Brown C. Interviewing for face identification. In: Valentine T, Davis JP, editors. Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV. Chichester, UK: John Wiley & Sons; 2015. p. 15–41. https://doi.org/10.1002/9781118469538
8. Kebbell MR, Milne R. Police officers' perceptions of eyewitness performance in forensic investigations. J Soc Psychol. 1998;138(3):323–330. http://dx.doi.org/10.1080/00224549809600384
9. Innocence Project. Eyewitness identification reform [webpage on the Internet]. c2016 [2022 Nov 10]. Available from: https://www.innocenceproject.org/eyewitness-identification-reform/
10. Albright TD. Why eyewitnesses fail. Proc Natl Acad Sci USA. 2017;114(30):7758–7764. http://dx.doi.org/10.1073/pnas.1706891114
11. Meissner CA, Sporer SL, Schooler JW. Person descriptions as eyewitness evidence. In: Lindsay RCL, Ross DF, Read JD, Toglia MP, editors. Handbook of eyewitness psychology. New York: Routledge; 2007. p. 3–34. https://doi.org/10.4324/9781315805535
12. Wells GL, Hasel LE. Facial composite production by eyewitnesses. Curr Dir Psychol Sci. 2007;16(1):6–10. http://dx.doi.org/10.1111/j.1467-8721.2007.00465.x
13. Hurley G. The trouble with eyewitness identification testimony in criminal cases [document on the Internet]. c2017 [cited 2021 Oct 23]. Available from: http://ncsc.contentdm.oclc.org/cdm/ref/collection/criminal/id/280
14. Charman SD, Gregory AH, Carlucci M. Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. J Exp Psychol Appl. 2009;15(1):76–90. http://dx.doi.org/10.1037/a0014682
15. Fahsing IA, Ask K, Granhag PA. The man behind the mask: Accuracy and predictors of eyewitness offender descriptions. J Appl Psychol. 2004;89(4):722–729. http://dx.doi.org/10.1037/0021-9010.89.4.722
16. Meissner CA, Sporer SL, Susa KJ. A theoretical review and meta-analysis of the description-identification relationship in memory for faces. Eur J Cogn Psychol. 2008;20(3):414–455. http://dx.doi.org/10.1080/09541440701728581
17. Frowd CD. Forensic facial composites. In: Smith AM, Toglia PT, Lampinen JM, editors. Methods, measures, and theories in eyewitness identification tasks. New York: Routledge; 2021. p. 34–64. https://doi.org/10.4324/9781003138105-5
18. R v Shekelele (1953) 1 SA 636 (T).
19. Sporer SL. Psychological aspects of person descriptions. In: Sporer SL, Malpass RS, Koehnken G, editors. Psychological issues in eyewitness identification. New York: Psychology Press; 1996. p. 53–86. https://doi.org/10.4324/9781315821023
20. Milne B, Bull R. Investigative interviewing: Psychology and practice. Chichester, UK: Wiley; 1999.
21. Shepherd JW, Ellis HD. Face recall – methods and problems. In: Sporer SL, Malpass RS, Koehnken G, editors. Psychological issues in eyewitness identification. New York: Psychology Press; 1996. p. 87–115. https://doi.org/10.4324/9781315821023
22. Mcleod SA. The cognitive interview [webpage on the Internet]. c2019 [cited 2022 Oct 23]. Available from: https://www.simplypsychology.org/cognitive-interview.html

23. Kebbell MR, Milne R, Wagstaff GF. The cognitive interview: A survey of its forensic effectiveness. Psychol Crime Law. 1999;5(1–2):101–115. http://dx.doi.org/10.1080/10683169908414996

24. Frowd CD, Bruce V, Smith AJ, Hancock PJB. Improving the quality of facial composites using a holistic cognitive interview. J Exp Psychol Appl. 2008;14(3):276–287. http://dx.doi.org/10.1037/1076-898X.14.3.276

25. Demarchi S, Py J. A method to enhance person description: A field study. In: Bull R, Valentine T, Williamson T, editors. Handbook of psychology of investigative interviewing: Current developments and future directions. Oxford, UK: Wiley-Blackwell; 2009. p. 241–256. https://doi.org/10.1002/9780470747599

26. Gabbert F, Hope L, Fisher RP. Protecting eyewitness evidence: examining the efficacy of a self-administered interview tool. Law Hum Behav. 2009;33(4):298–307. http://dx.doi.org/10.1007/s10979-008-9146-8

27. Odinot G, Wolters G, Van Koppen PJ. Eyewitness memory of a supermarket robbery: A case study of accuracy and confidence after 3 months. Law Hum Behav. 2009;33(6):506–514. http://dx.doi.org/10.1007/s10979-008-9152-x

28. Woolnough PS, MacLeod MD. Watching the birdie watching you: Eyewitness memory for actions using CCTV recordings of actual crimes. Appl Cogn Psychol. 2001;15(4):395–411. http://dx.doi.org/10.1002/acp.717

29. Van Koppen PJ, Lochun SK. Portraying perpetrators: The validity of offender descriptions by witnesses. Law Hum Behav. 1997;21(6):661–685. http://dx.doi.org/10.1023/a:1024812831576

30. Frowd C, Bruce V, McIntyre A, Hancock P. The relative importance of external and internal features of facial composites. Br J Psychol. 2007;98(1):61–77. http://dx.doi.org/10.1348/000712606x104481

31. Frowd CD, Erickson WB, Lampinen JM, Skelton FC, McIntyre AH, Hancock PJB. A decade of evolving composites: Regression- and meta-analysis. J Forensic Pract. 2015;17(4):319–334. http://dx.doi.org/10.1108/jfp-08-2014-0025

32. Frowd CD, Hancock PJB, Carson D. EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. ACM Trans Appl Percept. 2004;1(1):19–39. http://dx.doi.org/10.1145/1008722.1008725

33. Gibson SJ, Solomon CJ, Maylin MIS, Clark C. New methodology in facial composite construction: From theory to practice. Int J Electron Secur Digit Forensics. 2009;2(2):156–168. http://dx.doi.org/10.1504/ijesdf.2009.024900

34. Tredoux C, Nunez D, Oxtoby O, Prag B. An evaluation of ID : An eigenface based construction system. S Afr Comput J. 2006(37):90–97. https://journals.co.za/doi/10.10520/EJC28017

35. Fodarella C, Kuivaniemi-Smith H, Gawrylowicz J, Frowd CD. Forensic procedures for facial-composite construction. J Forensic Pract. 2015;17(4):259–270. http://dx.doi.org/10.1108/jfp-10-2014-0033

36. Zahradnikova B, Duchovicova S, Schreiber P. Facial composite systems: Review. Artif Intell Rev. 2018;49(1):131–152. http://dx.doi.org/10.1007/s10462-016-9519-1

37. Frowd CD, Pitchford M, Skelton F, Petkovic A, Prosser C, Coates B. Catching even more offenders with EvoFIT facial composites. In: Third International Conference on Emerging Security Technologies; 2012 September 5–7; Lisbon, Portugal. IEEE; 2012. p. 20–26. http://dx.doi.org/10.1109/EST.2012.26

38. Christie DF, Ellis HD. Photofit constructions versus verbal descriptions of faces. J Appl Psychol. 1981;66(3):358–363. http://dx.doi.org/10.1037/0021-9010.66.3.358

39. McQuiston-Surrett D, Topp LD. Externalizing visual images: Examining the accuracy of facial descriptions vs. composites as a function of the own-race bias. Exp Psychol. 2008;55(3):195–202. http://dx.doi.org/10.1027/1618-3169.55.3.195

40. Lech AM, Johnston R. The relative utility of verbal descriptions and facial composites in facial identifications. Int J Bio-Sci Bio-Technol. 2011;3(3):1–16.

41. Ness H. Improving composite images of faces produced by eyewitnesses. Stirling: University of Stirling; 2003. http://hdl.handle.net/1893/12555

42. Meissner CA, Brigham JC. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. Psychol Public Policy Law. 2001;7(1):3–35. http://dx.doi.org/10.1037/1076-8971.7.1.3

43. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp; 2017.

44. Field A. Discovering statistics using IBM SPSS statistics. 5th ed. London: SAGE Publications; 2017.

45. Aguinis H, Gottfredson RK, Joo H. Best-practice recommendations for defining, identifying, and handling outliers. Organ Res Methods. 2013;16(2):270–301. http://dx.doi.org/10.1177/1094428112470848

46. Buzzi-Ferraris G, Manenti F. Outlier detection in large data sets. Comput Chem Eng. 2011;35(2):388–390. http://dx.doi.org/10.1016/j.compchemeng.2010.11.004

47. Templeton GF. A two-step approach for transforming continuous variables to normal: Implications and recommendations for IS research. Commun Assoc Inf Syst. 2011;28(1), Art. #4. http://dx.doi.org/10.17705/1cais.02804

48. Shapiro PN, Penrod S. Meta-analysis of facial identification studies. Psychol Bull. 1986;100(2):139–156. http://dx.doi.org/10.1037/0033-2909.100.2.139

49. Wogalter MS, Bradley Marwitz D. Face composite construction: In-view and from-memory quality and improvement with practice. Ergonomics. 1991;34(4):459–468. http://dx.doi.org/10.1080/00140139108967329

50. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 June 15–20; Long Beach, CA, USA. IEEE; 2019. p. 4396–4405. https://doi.org/10.1109/CVPR.2019.00453

51. Chen X, Qing L, He X, Luo X, Xu Y. FTGAN: A fully-trained generative adversarial network for text to face generation [preprint]. arXiv [cs.CV]. 2019. http://arxiv.org/abs/1904.05729

52. Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis [preprint]. arXiv [cs.LG]. 2021. http://arxiv.org/abs/2105.05233

53. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with CLIP latents [preprint]. arXiv [cs.CV]. 2022. http://arxiv.org/abs/2204.06125

54. Ho J, Chan W, Saharia C, Whang J, Gao R, Gritsenko A, et al. Imagen Video: High definition video generation with diffusion models [preprint]. arXiv [cs.CV]. 2022. http://arxiv.org/abs/2210.02303

55. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative adversarial text to image synthesis. In: Balcan MF, Weinberger KQ, editors. Proceedings of the 33rd International Conference on Machine Learning; 2016 June 19–24; New York, USA. New York: PMLR; 2016. p. 1060–1069.

56. Xu T, Zhang P, Huang Q, Zhang H, Gan Z, Huang X, et al. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018 June 18–23; Salt Lake City, UT, USA. IEEE; 2018. p. 1316–1324.

57. Ding M, Yang Z, Hong W, Zheng W, Zhou C, Yin D, et al. CogView: Mastering text-to-image generation via transformers [preprint]. arXiv [cs.CV]. 2021. http://arxiv.org/abs/2105.13290

Non-linear effects of stress on eyewitness memory

AUTHORS:
Milton Gering[1]
Tayla Johnson[1]
Colin Tredoux[1]

AFFILIATION:
[1]Department of Psychology, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Milton Gering

EMAIL:
miltongering@gmail.com

DATES:
Received: 27 Aug. 2021
Revised: 12 Sep. 2022
Accepted: 18 Oct. 2022
Published: 29 Mar. 2023

HOW TO CITE:
Gering M, Johnson T, Tredoux C. Non-linear effects of stress on eyewitness memory. S Afr J Sci. 2023;119(3/4), Art. #12102. https://doi.org/10.17159/sajs.2023/12102

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Leslie Swartz

KEYWORDS:
stress, eyewitness, memory, line-up, faces

FUNDING:
None

The effect of stress on face identification is not yet clear, with recent experiments finding positive, negative and null results. Here we report the results of two experiments examining the effect of stress on eyewitness performance in line-up face recognition tasks. Both experiments use a stress manipulation and live mock crime paradigm to examine the relationship between stress at encoding and subsequent line-up performance. Experiment 1 replicated an experiment by Sauerland et al. (Behav Sci Law. 2016;34(4):580–594) which induced stress using the Maastricht Acute Stress Test. The replication found the same null result as the original experiment. Experiment 2 aimed to address a limitation of many laboratory experiments which dichotomise stress into low and high groups for comparison. As the Yerkes-Dodson law (1908) suggests that a non-linear relationship exists between stress and performance, it was hypothesised that using a low, medium and high stress manipulation might show clearer results than a dichotomous paradigm. The results of Experiment 2 show a non-linear relationship, with no difference between the low and high stress groups but better performance by the middle stress group. The results suggest that a different approach is required in experiments on stress and face recognition, as the stress–performance relationship is likely non-linear.

Significance:

- Non-linear models are better predictors of face recognition in line-up tasks than are linear models.
- Two group designs provide insufficient resolution to capture the stress–performance relationship.

Non-linear effects of stress on eyewitness memory

People who witness crimes often experience high levels of stress during the event. The effect that this stress might have on memory is important, yet remains unclear. Although an earlier meta-analysis[1] showed a negative effect of stress at encoding, recent studies have found mixed results, and a more recent meta-analysis concludes that extant evidence does not show that stress affects witness memory[2]. Some of the differences may be a result of different methods, particularly for stress induction, used in studies within the literature, but it is likely also that the stress induction paradigms used in the corpus of studies have produced unreliable results. Where the Deffenbacher review found a negative linear effect[1], the more recent review suggests that a non-linear model may offer a better explanation[2]. We report on two studies that show that a dichotomous stress manipulation will not achieve reliable results, especially because the stress–performance relation is explained by the universally cited Yerkes-Dodson law[3] to be non-linear (likely quadratic). To see a non-linear effect, it is necessary to induce a minimum of three levels of stress. In doing so, the studies presented here show a flaw in the existing literature and describe an experimental solution.

Stress and memory

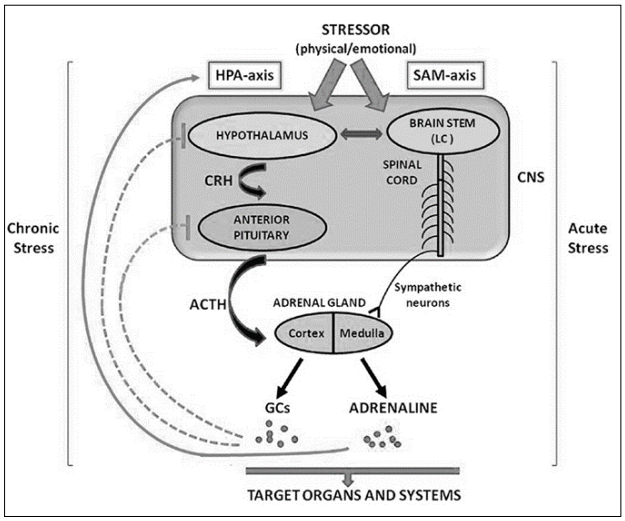
Acute episodes of stress have known physiological consequences, primarily preparing the body for action[4], but simultaneously having effects on memory. These responses occur through the activation of two systems: the fast-acting sympathetic-adrenal-medullary (SAM) axis, or the slower-acting hypothalamic-pituitary-adrenal (HPA) axis (see Figure 1). The SAM axis response increases activity of the sympathetic nervous system, releasing adrenaline, which in turn stimulates the release of norepinephrine in the brain.[5] The norepinephrine activates the amygdala, which interacts with the brain regions involved in encoding memories, including the hippocampus and frontal lobes.[4] These areas are also activated through the HPA axis, their glucocorticoid receptors responding to the release of a different hormone, cortisol.[5] This complex response is experienced by many witnesses of crimes and may affect how they process and encode events. The combined cortisol and noradrenaline response is thought to affect memory, impacting different brain regions involved in perceiving, encoding and storing memories for later recall and retrieval.

Several factors influence the effect of acute stress on memory, including the amount of stress experienced, the degrees of activation of the SAM and HPA axes, the types of memory process, and the time since onset of stress.[6] While both axes activate the amygdala and hypothalamus, SAM activation appears to boost memory consolidation, while HPA axis activation weakens it.[7] These counteracting paths may explain why mild or moderate stress during encoding improves memory, yet intense stress impairs it.[8] This non-linear effect of stress on memory has been known experimentally since 1908[3] and is often referred to as the 'Yerkes-Dodson law', shown here as Figure 2. The quadratic curve shown in the figure may not be as symmetrical as suggested, perhaps having a steep drop off at extreme levels of stress ('catastrophic forgetting').[9] Despite some scholars contesting the Yerkes-Dodson law, or its general application, there has been no empirical work with the type of stress inductions needed to falsify the theory.[2]

A significant limitation in the extant literature on stress and witness memory is that almost all studies use a two-group design (typically 'high' and 'low' stress groups), despite the obvious impossibility of detecting a non-linear effect with two levels of a predictor. This may be why different stress responses are seen, as a dichotomous division of a non-linear variable will show different patterns depending on the intensity of stress induced.

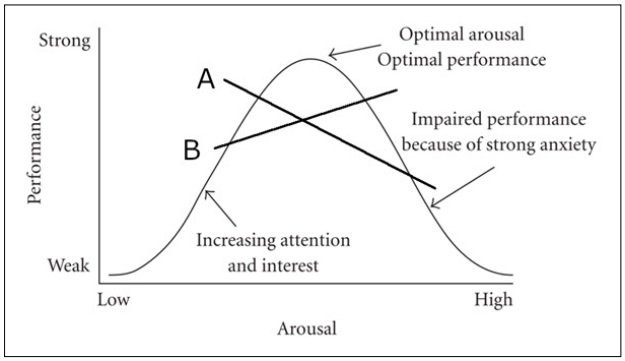
In understanding the effects of stress on witness memory, it is also important to note that, to remember details of an event, one must notice and attend to it. Attentional resources are limited, and not everything in the environment can be attended to. Emotional events, such as fear-inducing crimes, tend to be conspicuous and thus attract attention, particularly to central, life-threatening aspects of the event, leading to poorer encoding of other information.[3] Arousal

will also impair memory in situations where cognitive demands are high, as the brain must process the stress response while simultaneously using other cognitive resources.[10] This may affect how successfully different information is encoded by witnesses into their memory.



Source: Diamond et al.[21] licensed under a Creative Commons licence

Figure 1: Schematic representation of the actions of the hypothalamic-pituitary-adrenal (HPA) and sympathetic-adrenal-medullary (SAM) axes in the regulation of the stress response.



Source: Baritaki et al.[22] licensed under a Creative Commons licence

Figure 2: The Yerkes-Dodson Curve. The lines A and B show how it is possible to find negative and positive (but misleading) effects of stress on performance. The Yerkes-Dodson curve could be simply quadratic, or taper off at the end to look more Gaussian, or be asymmetrical and piecewise and only roughly approximate a quadratic curve.

Another factor to consider is the nature of the memory tasks required of eyewitnesses to crimes. Witnesses to a crime are required to *recall* events, and to *recognise* perpetrators, which are explicit and implicit memory tasks, respectively.[11] Laboratory studies of memory show that stress at encoding benefits recall tasks more than recognition tasks.[12] Similar task-dependent differences have been seen in studies on depression. While depression is generally considered an affective disorder, it also produces changes in cognition.[13] Participants with major depressive disorder perform worse than healthy controls in explicit memory tasks.[14] This suggests that depression may impair a witness's ability to recall events, and probably more than it would impair recognition of perpetrators. As such, it should be measured and included as a control variable. While mood disorders may be a source of confounding influences in the literature, whether the memory task requires implicit or explicit memory appears to be an overarching moderator of the stress–performance relationship.

We report two studies here that investigate the effect of stress on witness memory, simulating a criminal event as part of a live interaction. We aimed to capture the essence of a stressful crime by inducing stress and presenting a critical event on which participants were later tested. Both these studies include elements considered best practice in the stress induction literature, including a meaningful delay between encoding and recognition, the use of physiological and self-report manipulation checks of stress induction, and well-constructed line-ups for testing recognition memory.[15,16] We believe that these methodological elements are important parts of the three-group design that is critical for mapping the non-linear relationship between stress and performance. Although we consider both recall and recognition memory, our analysis is focused mostly on participant recognition of perpetrators from line-ups.

Method

Design

Both studies used an experimental design to investigate the effect of induced stress on eyewitness memory. In Experiment 1, stress was manipulated, following common practice in the literature[15], into low and high levels of stress, and effects tested on perpetrator recognition and event recall. For Experiment 2, stress was manipulated to have three levels: a control group had no intentional stress induction, one experimental group received a moderate stress induction, and a second experimental group received a high stress induction.

Participants

For both experiments, participants were recruited via the Student Research Participation Programme (SRPP) run by the Department of Psychology at the University of Cape Town (UCT). Participants were recruited via an advertisement on a university online portal. For Experiment 1, 180 students were screened for eligibility and 123 of those were invited to take part in the laboratory session. Of those 123 participants, 40 (8 men) took part in the 45-min laboratory session, completing the Maastricht Acute Stress Test (MAST), with 28 participants (6 men) providing a full set of data. As such, there were 40 participants in the stress induction phase and 28 in the eyewitness simulation phase. For Experiment 2, 190 (of 266) participants were invited to take part in the laboratory phase of the experiment after screening. Of those, 89 took part in the experiment. Of these, 14 were previously familiar with the 'perpetrator' and so were excluded from analysis, leaving 75 participants.

Screening instruments

Various instruments were used to screen, check manipulations, or measure covariates. These processes were undertaken through an online survey in the weeks prior to the in-person laboratory sessions, and, as they are not central to our hypotheses, we mention them only briefly, as follows (values in parentheses are Cronbach alpha estimates of internal consistency, or test-rest reliability): The 4-item Primary Care Post-Traumatic Stress Disorder Screen (PC-PTSD; $\alpha=0.89$; $r=0.83$), the State-Trait Anxiety Inventory Form Y1 and Y2 (STAI-Trait and STAI-State; $\alpha=0.89$; $r=0.69$–0.89), Beck Depression Inventory-II (BDI-II; $\alpha=0.90$; $r=0.93$), the Beck Anxiety Inventory (BAI; $\alpha=0.92$; $r=0.75$), the Positive and Negative Affect Schedule (PANAS; $\alpha=0.89$; $r=0.83$).

In both experiments we excluded students who had been diagnosed with a psychiatric disorder, who had high levels of trait anxiety or a history of head injury, and who had been affected by psychological trauma. Anyone answering 'yes' to the head trauma question or any of the four trauma questions, or scoring in the 'severe' category for anxiety or depression was not invited to participate in the in-person laboratory session. An additional exclusion criterion was used for Experiment 2, namely students who had a history of plagiarism at UCT were excluded. Only young adults, between 18 and 25, were invited to participate in the studies.

Experimental materials

The Maastricht Acute Stress Test

The MAST is a stress induction procedure that combines the Cold Pressor Test with mental arithmetic from the Trier Social Stress Test.[17] Participants alternate between immersing a hand in cold water and doing an arithmetic task in which they must count backwards in steps of 17

from 2043. The task is known to activate both the HPA and SAM axes, eliciting a comprehensive stress response.[17] This procedure was used to induce stress in Experiment 1, having previously been used in a similar experiment in Maastricht.[16] The control group goes through the Placebo MAST, which uses warm water and a simple counting task so as not to induce stress but to maintain the equivalent duration of the procedure.

The Self-Administered Interview

The Self-Administered Interview (SAI©) is a standardised, self-administered interview, based on the cognitive interview.[18] The interviewer asks generic questions applicable to various crimes; this test has been shown to enhance recall[18] and was used in Experiment 1.

The Vrije Universiteit Ambulatory Monitoring System

Heart rate (HR) and skin conductance (SC) were measured using the Vrije Universiteit Ambulatory Monitoring System (VU-AMS) in both experiments as part of the stress manipulation. Both HR and SC have been shown to increase when people experience stress.[18]

Line-ups

Target absent (TA) and target present (TP) colour line-ups were constructed for each target, all of whom were volunteers studying at UCT. Both experiments used two unique targets and so two separate TA and TP line-ups were constructed for each experiment. All line-ups consisted of six different faces photographed from the neck upwards against a pale grey background. Each individual in the line-up wore a black t-shirt. For Experiment 1, foils were selected to be similar to the target; for Experiment 2, foils were selected to match a verbal description of the target. These methods are both recommended in the literature and require similarity to be judged by the line-up constructor. Line-ups were constructed by the first author (M.G.) for Experiment 1 and the second author (T.J.) for Experiment 2, before being assessed for similarity and 'pop-out' effects by the study supervisor. Two versions of each line-up were constructed, with targets in different positions, to reduce the influence of a position effect. All targets and foils were women in their 20s; three were Caucasian and one (in Experiment 1) was East Asian. Photographs of foils were retrieved from the supervising author's database of line-up materials.

Procedure

Ethical clearance for both experiments was obtained through the Department of Psychology Ethics Committee at UCT.

Screening phase, Experiment 1

The STAI-Trait and PC-PTSD inventories were used to screen out (and thus protect) participants with high anxiety and those with symptoms of PTSD. These tests were administered online.

Laboratory phase, Experiment 1

Once consent forms had been signed, participants attached the VU-AMS device to themselves, and it was set to record. Participants then completed the BDI-II which was included in the modelling as a control. Participants were informed of the procedure for the MAST, or placebo-MAST for the controls, before commencing with the test. For the experimental group, the water was cooled with ice to 5 °C and for the control group it was warmed to 20 °C. After the MAST procedure, participants' state anxiety was measured with the STAI-State. Thereafter, participants were given a vocabulary test as a distractor task but told that it was the main test. This was done so as not to alert participants that the main test was in fact the mock crime. At this point, the experimenter left the room, leaving a phone on the VU-AMS container. A confederate then knocked on the door, entered the room and told the participant that they had left their phone behind during a previous session. The confederate then went to the table, took the researcher's phone, and left the room. Confederates always wore blue jeans and simple tops in dull colours.

When the researcher reappeared, he asked about the distractor task while searching for the phone, noted that the phone was missing and asked participants if they had seen anything. Once participants had told the researcher what had happened, they were informed that they had witnessed a mock crime and were debriefed. An email containing both a recall and recognition task, ensuring blind administration, was sent to participants the next day. A response period of 24 hours ensured that participants were only affected by stress during encoding, an important control (cf. Sauerland et al.[16]). The task sent to participants consisted of the SAI as well as two line-ups. Each participant first saw a TA line-up, and then a TP line-up, for the confederate/thief they saw in the lab. The repeat measure was used to increase power, given the small sample size. Both line-ups had unique foils, and participants did not know to expect two line-ups or that one would not include the target, thus mimicking real-world uncertainty. Instructions stated that 'the suspect may or may not be present in the line-up'. Participants were asked to rate their confidence in their line-up selection and had the option to say the suspect was not present or that they did not know who the suspect was. The information from the SAI was used to code for accuracy of descriptions by the witnesses. Confidence was not analysed in the present study.

Screening phase, Experiment 2

The same screening procedure was used as in Experiment 1, although the STAI was replaced with the BAI.

Laboratory phase, Experiment 2

Participants arrived for in-person testing group sessions with a maximum of 10 participants, only one of whom was asked to wear the VU-AMS. Each group was assigned to one of three stress conditions (control, moderate stress, high stress). Upon arrival at the laboratory, participants were seated in front of a computer with a consent form and a demographic information sheet, and reported their negative affect state using the PANAS. Once forms were completed, the participants selected for physiological monitoring attached themselves to the VU-AMS device. Participants were told that while they completed a task, a research assistant would be checking their academic plagiarism history. Each stress group received different instructions on this task: the control group directly copied the extract, the moderate stress group summarised it, and the high stress group summarised the extract but were told that it would be marked and compared with the rest of the group's. After this task, the first target emerged from a curtained area in the room and delivered a false plagiarism report: the control group was told their submissions were free of plagiarism; the moderate stress group was sternly told that plagiarism was present in some of the submissions but no further action was to be taken; and the high stress group was sternly told that plagiarism was present, that participants would be identified on their way out of the session, and that they were to appear in a student disciplinary tribunal to plead their case at a later date. This was the key point for the stress induction. Target one then went back into the curtained area after 90 seconds of exposure and interaction with participants.

Participants again reported their affective state, and the researcher informed participants that the plagiarism report was false, before giving them a distractor task. The second target was then introduced by the researcher. The target emerged from a curtained area in the room for 90 seconds and debriefed the participants on what the task had really consisted of. The targets were assigned to roles in counter-balanced order, across groups, which, along with the randomisation of experimental condition, was done through MS Excel. The researcher collected a final measure of affective state before debriefing participants. Participants were told that they would be required to identify the two targets in a photographic line-up the following day and were emailed a copy of the debriefing form and a link to two photographic line-ups 24 hours after their laboratory participation. Participants were instructed to identify the two targets to whom they were exposed previously and given the option to indicate if they did not know if the target was present, or to indicate that the target was not present in the line-up. The identifications were done online using simultaneous photographic line-ups and contained a TA line-up for one of the targets, and a TP line-up for the other target. Participants always saw one TA and one TP line-up.

Data management and statistical analysis

For both studies, data were analysed with the R statistical programming language[19], and the lme4 package[20]. All assumptions for analysis were met unless otherwise stated.

In Experiment 1, HR and SC were sampled between minutes two and three to obtain a baseline measurement of these physiological measures of stress. A second measure was taken between minutes 16 and 17 of the recording, by which time the MAST had been completed. A mixed-methods ANOVA was conducted to see if there were significant main or interaction effects between experimental manipulations and time. The STAI-State was administered in the second time interval, post-MAST, and a t-test was conducted to see if there was a significant difference in STAI-State between groups post-stress induction. A list of verifiable details was coded for details of the mock crime, descriptions of the perpetrator and the location of the event. Inter-rater reliability was obtained for the 28 coded statements on details of location, events and perpetrator descriptions. Descriptions were generated by the first author and checked with the confederates as in the original study we replicate[16] before being checked by a research assistant who was the second coder. Intra-class inter-rater reliability coefficients were 0.89, 0.93 and 0.82 for location, events and perpetrator details, respectively. An analysis of covariance (ANCOVA) was conducted for each category of recall, and on total details recalled, to analyse the effect of condition (stress vs control) on recall memory, using BDI-II as a covariate. For face identification, a logistic regression was conducted to compare the results between the stress and control groups for both TP and TA line-ups.

For Experiment 2, the negative affect score was calculated by using the negative items from the PANAS at the three time points at which physiological measures were also taken. A 3x3 mixed-designs ANOVA was conducted to ascertain whether a significant interaction occurred between the three stress groups at the different time intervals in self-report measures of negative affect. Furthermore, a planned contrast to check for the predicted differences between groups was run on negative affect measurements taken directly after the stressor to see whether the data matched the expected pattern of low, moderate and high reported stress. Once the raw data for line-up responses had been tabulated, the variables for each target line-up were reduced and coded as dichotomous variables with levels accurate and inaccurate (i.e. correct rejections in TA line-ups and correct identifications in TP line-ups were coded as accurate). It is important to note that 'don't know' responses were coded as 'inaccurate' for line-ups. A separate quadratic logistic regression was then run for each target in R to ascertain whether stress predicted accuracy in suspect identification in the expected inverted-U shape.[8]

Results

Experiment 1

Manipulation check

HR and SC were analysed for 35 participants (experimental group = 17) using a repeated measures ANOVA to check the stress manipulation on physiological responses. The STAI-Y1 scores of 40 participants (experimental group = 20) were used as a self-report manipulation check on psychologically experienced stress. Both groups completed the STAI-Y1 after doing the MAST or placebo MAST (for the control group). An independent samples t-test was conducted to see if there was a significant difference between the control and high stress groups post-MAST. Results are summarised in Table 1. Figure 3 shows the interaction effects of experimental group and time of measurement for each of HR and SC. This interaction between group and time on physiological measures shows opposite patterns in arousal for the stress and control groups. These results indicate that the stress induction procedure was successful.

Effect of stress on recall

In Experiment 1, recall data were collected from 28 participants (stress group = 14). Means and standard deviations are reported in Table 2 for each category of recall (events, location and participant description), and a total score for each group is also reported. Comparing the results for total recall across conditions suggests that the MAST group performed slightly worse than the control group. Separate ANCOVAs were conducted for memory of location, events and descriptions using depression as a covariate. For the analysis for memory of events, the more robust Welch was conducted to handle the homogeneity of variance assumption. The *p*-values and effect sizes for the ANCOVAs are reported alongside the descriptive statistics.

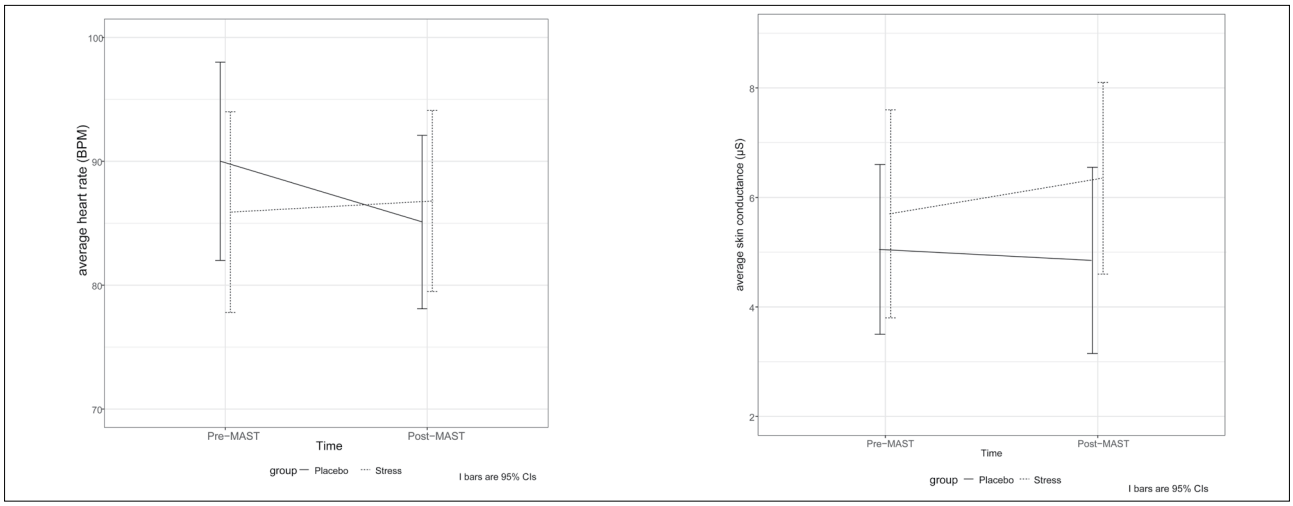


Figure 3: Disordinal heart rate x time and skin conductance x time interactions for the low (Placebo MAST) and high (MAST) stress groups in Experiment 1, showing averages and confidence intervals.

Table 1: Self-reported stress and depression scores by condition

	STAI-Y1			BDI(II)			Heart rate			Skin conductance		
	M	SD	d	M	SD	d	M	SD	d	M	SD	d
MAST	24.35	9.97	0.95	13.90	9.70	0.42	0.90	10.47	0.84*	0.20	2.25	0.92*
Placebo	14.30	8.94		10.35	6.45		-0.42	10.37		0.65	2.45	

Note: Means (M), standard deviations (SD) and effect sizes for manipulation checks. For heart rate and skin conductance, mean difference is reported.

**Indicates* $p<0.05$ *for the inferential test from which the effect size was produced.*

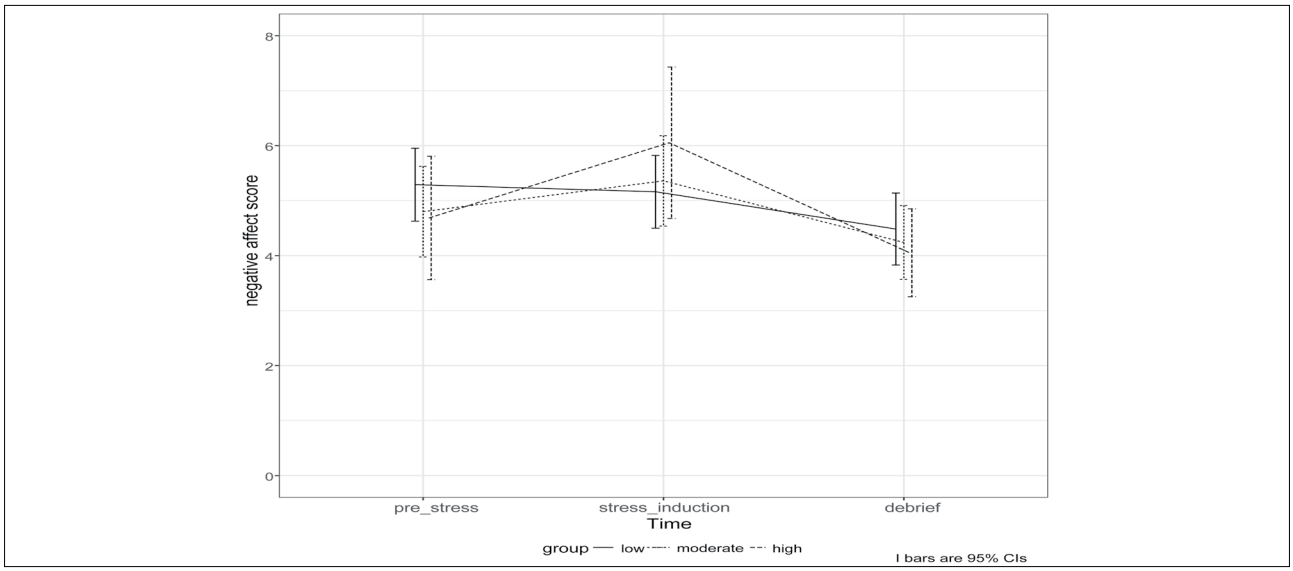


Figure 4: Time x group interaction of total negative affect scores from the PANAS, Experiment 2.

Table 2: Means and standard deviations for recall type by condition

	MAST		Placebo		ANCOVAs		
	M	SD	M	SD	F	*p*	d
Location details	4.86	3.08	6.50	2.98	1.85	0.190	0.54
Event details	5.21	2.63	5.21	1.19	0.01	0.934	0.03
Perpetrator description	3.93	2.37	4.36	2.92	0.12	0.646	0.18
Total	14.00	5.94	16.07	3.87	0.71	0.441	0.29

Note: Means (M) and standard deviations (SD) for each recall type for the stress and control group.

Effect of stress on face identification

Face identification data were collected from 28 participants (stress group = 14). From Table 3, which shows the response data for the face identifications, we can see that there is no difference between the groups. Of the 28 participants, 15 made identification attempts in the TP line-up across the groups, and 8 participants were successful. The results of the logistic regression show no difference between the high and low stress groups on TP (B=0.04, SE=1.10, p=0.974) or TA (B=0.97, SE=1.34, p=0.480) line-ups with no effect of depression on either task (p=0.718 and p=0.492, respectively). Thus, we can conclude that there were no meaningful differences in identification rate between the groups.

Table 3: Line-up responses by stress condition for target-absent and target-present line-ups

	MAST	Placebo
Target absent	*n*=14	*n*=14
Correct reject	3	2
False alarm	10	9
'Don't know' responses	1	3
Target present	*n*=14	*n*=14
Correct identification	4	4
Foil identification	3	4
False reject	5	6
'Don't know' responses	2	0

Note: For a target-absent line-up, rejecting the line-up is the best response, equivalent to a correct identification in the target-present line-up. False alarm in a target-absent line-up and foil identification are also equivalent.

Experiment 2

Manipulation check

HR was obtained and analysed from 10 participants (one from each session): three in the control group; three in the medium-stress group; and four in the high-stress group. The mean change values (see Table 4) indicate that HR increased for all groups throughout the experiment. However, due to the small sample for whom we collected physiological measures due to limitations of time and equipment, and the ensuing low statistical power, self-report measures were used for the manipulation check rather than the physiological measures. The PANAS was administered at three different points throughout the session: one prior to stress induction, one directly after stress induction, and one about 10 minutes after stress induction. The total negative affect score for each PANAS was captured for 75 participants. The descriptive statistics in Table 5 show that mean negative affect scores show the expected descriptive pattern based on the intended stress induction.

A multi-level linear model with a random intercept term was run on the negative affect PANAS data for the three stress groups at three time points. This accounted for individual differences in effect of the stress factor. Individual differences between participants accounted for 56% of the variance in negative affect observed. It also showed that negative affect was significantly lower at debrief than at pre-stress induction (B= -0.81, p=0.022), and that during stress induction, the high stress group was significantly more stressed than the control group (B=1.5, p=0.009), while the medium stress group was not significantly more stressed than the control group (B=0.69, p=0.192). This suggests that participants started and returned to similar stress levels, but that during stress induction they showed differences in their stress levels.

Table 4: Descriptive statistics for heart rate

Heart rate change	Group	Mean	SD	*n*
Baseline to stress induction	Control	3.75	0.22	3
	Moderate stress	4.38	1.02	3
	High stress	2.73	2.88	4
	Total	3.53	1.89	10
Stress induction to post-stress	Control	1.31	1.86	3
	Moderate stress	0.34	3.93	3
	High stress	1.23	4.78	4
	Total	0.99	3.47	10

Table 5: Descriptive statistics for self-report stress in Experiment 2 at three time points: pre-stress, stress induction and debrief

Time	Stress group	Mean	SD	*n*
Pre-stress	Low	5.29	1.88	31
	Moderate	4.80	2.10	25
	High	4.68	2.50	19
Stress induction	Low	5.16	1.88	31
	Moderate	5.36	2.10	25
	High	6.05	3.06	19
Debrief	Low	4.48	1.86	31
	Moderate	4.24	1.72	25
	High	4.05	1.78	19

A 3x3 multi-level linear model, for which the factors are the three levels of stress and three time points, showed a main effect for time ($F(2,144)=13.22$, $p<0.001$, $Eta^2=0.152$), but not for group ($F(2,72)=0.07$, $p=0.617$, $Eta^2=0.001$). The overall interaction between time x group approached significance ($F(4,148)=6.27$, $p<0.086$, $Eta^2=0.042$). This was investigated further, as ideally there would be a significant difference only at time point 2, when the stress was induced. A planned contrast of this point showed a non-significant trend of induced stress across the groups (B=0.89, SE=0.61, $p=0.144$) with the high stress group having greater negative affect than the low stress ($F(2,72)=1.68$, $p=0.199$) or moderate stress ($F(2,72)=0.82$, $p=0.370$) groups and the moderate stress group having greater negative affect than the low stress group ($F(2,72)=0.154$, $p=0.697$).

Line-up performance

Data from 75 participants were used to test the hypothesis that there would be differences in identification accuracy across the stress groups. Table 6 shows that the medium stress group had the best performance in both TA and TP line-ups.

Table 6: Descriptive statistics for Experiment 2 by group, line-up and perpetrator

	Group		
Line-up decision	**Control**	**Moderate stress**	**High stress**
Target (perpetrator)			
TP	(*n* = 6)	(*n* = 17)	(*n* = 12)
Correct identification	67% (4)	76% (13)	58% (7)
TA	(*n* = 25)	(*n* = 8)	(*n* = 7)
Correct rejection	40% (10)	88% (7)	43% (3)
Target (bystander)			
TP	(*n* = 25)	(*n* = 8)	(*n* = 7)
Correct identification	64 % (20)	87% (7)	71% (5)
TA	(*n* = 6)	(*n* = 17)	(*n* = 12)
Correct rejection	33% (2)	71% (12)	58% (7)
Target (both)			
TP	(*n* = 31)	(*n* = 25)	(*n* = 19)
Correct identification	64 (20)	80 (20)	63 (12)
TA	(*n* = 31)	(*n* = 25)	(*n* = 19)
Correct rejection	39% (12)	76% (19)	53 (10)

TP, target present; TA, target absent

This can be seen in the greater proportion of hits and lower proportion of false alarms for the medium stress group, as well as lower rates of rejection and 'don't know' responses in the TP line-ups, and higher rates of correct rejections in the TA line-ups. Logistic regressions of identification accuracy on stress were run using stress manipulation as a dummy variable, as well as a continuous variable, to assess whether effects would be linear or quadratic. As there was no significant difference between roles (B=0.25, SE=0.40, $p=0.527$) data were collapsed across perpetrator and bystander roles to increase statistical power.

The results of the regression analyses suggest a non-linear relationship. For the logistic regression using stress as a categorical predictor, one can see that the medium stress group always performed better than the low stress group (B=1.20, SE=0.43, $p=0.005$) and that there was no difference between the low and high groups (B=0.25, SE=0.42, $p=0.541$), as would be predicted from the negative quadratic relationship shown in Figure 2. Trend analyses showed that a negative quadratic model (B= -6.14, SE=2.28, $p=0.007$) was significant and the linear model was not (B=2.12, SE=2.02, $p=0.294$). A chi-squared model comparison showed that the quadratic model was significantly better than a linear one ($\chi^2 2(1, 147) = 7.86$, $p=0.005$). These two ways of looking at the data both confirm that the moderate stress group were the best performers, as expected under the Yerkes-Dodson relation. Table 7 shows the full sets of comparisons, where the non-linear models were consistently better than the linear ones.

Discussion

The experiments presented here investigated the effect of stress on eyewitness memory. Both used a live encoding event and a photographic line-up presented more than 24 hours later. Experiment 1 used a two-group (high/low) stress manipulation, with careful attention to methodological features to ensure successful stress induction, and found no significant effect of stress on witness memory, replicating Sauerland et al.[16] Experiment 2 used a three-level stress manipulation, inducing descriptively different stress levels, and found a non-linear relationship between stress intensity at encoding and line-up performance. While the power of the experiments is not sufficient to draw strong conclusions, the studies suggest that using only two levels of stress is not sufficient to show the effects of stress on witness memory. Although this is the current trend in the literature, a two-group design does not have the necessary resolution to reliably capture a non-linear relationship. Better methods of stress induction, using more than two groups, are absent in the literature and this absence needs to be addressed so as to avoid cumulation of null results or results of conflicting direction, due to low resolution modelling. Experiment 2, which found no difference between the low and high stress groups, confirms this, as without the moderate stress level, the pattern matches that of Experiment 1, in which there was no observable effect of stress at encoding. Figure 2 illustrates this problem, as depending on the level of induced stress, it is possible to find a positive, negative or null difference between stress groups (lines A and B). As inspection of the Yerkes-Dodson law[8] suggests, a non-linear relationship, measuring three or more points of stress, should become the standard practice.

Reasons for the non-linear relationship of stress and witness memory may be both physiological and behavioural. The SAM axis responses to stressors encourage the 'fight or flight' response, which may improve encoding: by focusing attention on the source of the stressor, it is more likely to be better remembered. A strong stress response will release hormones, which may impair encoding of witnessed events. These contrasting components should be better studied, and effects mapped so that the stress response can be viewed in terms of its separate physiological pathways. This could well improve the predictive utility of stress research, which extends to other research on memory as well as other physiological systems. Better manipulation checks using both physiological and standardised self-report measures will allow for clearer understanding of stress effects on witness memory. While this study showed a non-linear effect of stress intensity, there are still other factors which likely moderate the stress–performance relationship.

One of these factors might be the type of memory task utilised. In Experiment 1, there was a trend in the recall data that showed the low stress group performed slightly better than the high stress group – a trend which was not seen in the recognition data. This might indicate that

Table 7: Comparison of linear and non-linear models

Analysis	Condition	Comparison	Coefficient	SE	Z	*p*	AIC
Linear logistic regression	TP perpetrator	2 3	0.49 -0.36	1.04 1.05	0.47 -0.34	0.640 0.733	48.49
	TA perpetrator	2 3	2.35 0.12	1.14 0.87	2.01 0.14	0.040* 0.892	55.24
	TP bystander	2 3	1.37 0.34	1.15 0.93	1.20 0.37	0.232 0.715	53.08
	TA bystander	2 3	1.57 1.03	1.02 1.05	1.54 0.99	0.123 0.325	50.54
	TP both targets	2 3	0.79 -0.06	0.63 0.61	1.26 -0.10	0.207 0.923	96.35
	TA both targets	2 3	1.61 0.56	0.60 0.59	2.71 0.96	0.007* 0.337	101.22
	Collapsed TP and TA across targets	2 3	1.20 0.25	0.43 0.42	2.82 0.61	0.005* 0.541	196.3
Polynomial logistic regression	TP perpetrator	Linear Quadratic	-1.20 -1.91	2.08 2.22	-0.58 -0.86	0.564 0.392	48.49
	TA perpetrator	Linear Quadratic	1.98 -5.55	2.11 2.79	0.94 -1.99	0.349 0.047*	55.24
	TP bystander	Linear Quadratic	1.72 -2.90	2.23 2.82	0.77 -1.03	0.443 0.304	53.075
	TA bystander	Linear Quadratic	1.38 -3.03	2.08 2.14	0.66 -1.41	0.508 0.158	50.54
	TP both targets	Linear Quadratic	0.27 -3.31	2.08 2.36	0.13 -1.40	0.898 0.162	96.35
	TA both targets	Linear Quadratic	2.73 -5.37	2.03 2.24	1.34v-2.40	0.180 0.016*	101.22
	Collapsed TP and TA for both targets	Linear Quadratic	2.12 -6.14	2.02 2.28	1.05 -2.69	0.294 0.007*	196.3

TP, target present; TA, target absent; AIC, Akaike information criterion

tasks taxing different memory systems are affected differently by stress. However, it might also be an indication that task difficulty is an important moderating variable because the recall task was likely more difficult than the recognition task.[12] Although we included depressive mood as a covariate in our analyses, it was not found to predict any difference in performance, perhaps because the average level in our sample of students was low, and more severe depressive symptoms may be a moderating factor of the relationship between stress and memory.

Experiment 2 found no difference in recognition for the perpetrator who induced the stress and a bystander who appeared afterwards. This suggests that the effect of stress on facial encoding and recognition is generally as a result of the neurochemical response and not tied to the source of the stressor. While the stress induction for Experiment 2 did not show significant differences, it did identify the expected trend using both physiological and self-report measures. The stress induced was an inherent part of the eyewitness event, rather than induced through a laboratory manipulation as in Experiment 1. While such methods may not be as robust, they may be more meaningful to the participants. As stress is both subjective and physiological, the meaning and consequences of the stress may differ depending on its source. As such, the intensity, source and meaning of the induced stress should be considered when investigating the effect of stress on witness memory.

In conclusion, the studies presented here suggest that the stress–performance relationship should be considered non-linear. The main limitation in both of these studies is the relatively low sample size and consequent low statistical power. As such, we note that the studies are not conclusive but rather that the descriptive results suggest that the methods used for stress induction in the literature should be reconsidered. Experiments aiming to explore the effect of stress on eyewitness memory for faces and details of crimes should make use of designs with at least three stress groups. This will allow future studies to better investigate the stress–performance relationship along with any moderating effects.

Competing interests

We have no competing interests to declare.

Authors' contributions

M.G.: Conceptualisation, methodology, data collection, data analysis, data curation, writing – the initial draft, writing – revisions, project management. T.J.: Conceptualisation, methodology, data collection, writing – revisions. C.T.: Conceptualisation, methodology, writing – revisions, student supervision, project leadership.

References

1. Deffenbacher KA, Bornstein BH, Penrod SD, McGorty EK. A meta-analytic review of the effects of high stress on eyewitness memory. Law Hum Behav. 2004;28(6):687–706. https://doi.org/10.1007/s10979-004-0565-x
2. Gering MA, A systematic review and meta-analysis on the effects of stress at encoding on line-up decisions [master's thesis]. Cape Town: University of Cape Town; 2021.
3. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol. 1908;18(5):459–482. https://doi.org/10.1002/cne.920180503
4. Wolf OT, Atsak P, De Quervain DJ, Roozendaal B, Wingenfeld K. Stress and memory: A selective review on recent developments in the understanding of stress hormone effects on memory and their clinical relevance. J Neuroendocrinol. 2016;28(8):1–8. https://doi.org/10.1111/jne.12353

5. Gray J, Kogan J. Marrocco, McEwen, B. Genomic and epigenomic mechanisms of glucocorticoids in the brain. Nat Rev Endocrinol. 2017;13(11):661–673. https://doi.org/10.1038/nrendo.2017.97

6. Shields GS, Sazma MA, Yonelinas AP. The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. Neurosci Biobehav Rev. 2016;68(1):651–668. https://doi.org/10.1016/j.neubiorev.2016.06.038

7. Roozendaal B, McEwen BS, Chattarji S. Stress, memory and the amygdala. Nat Rev Neurosci. 2009;10(6):423–433. https://doi.org/10.1038/nrn2651

8. Drexler SM, Wolf OT. Stress and memory consolidation. In: Axmacher N, Rasch B, editors. Cognitive neuroscience of memory consolidation. Studies in Neuroscience, Psychology and Behavioral Economics. New York: Springer; 2017 p. 285–300. https://doi.org/10.1007/978-3-319-45066-7_17

9. Sandi C. Stress and cognition. Wiley Interdiscip Rev Cogn Sci. 2013;4(3):245–261. https://doi.org/10.1002/wcs.1222

10. Levine LJ, Edelstein RS. Emotion and memory narrowing: A review and goal-relevance approach. Cogn Emot. 2009;23(5):833–875. https://doi.org/10.1080/02699930902738863

11. Vakil E, Wasserman A, Tibon R. Development of perceptual and conceptual memory in explicit and implicit memory systems. J Appl Dev Psychol. 2018;57(1):16–23. https://doi.org/10.1016/j.appdev.2018.04.003

12. Het S, Ramlow G, Wolf OT. A meta-analytic review of the effects of acute cortisol administration on human memory. Psychoneuroendocrinology. 2005;30(8):771–784. https://doi.org/10.1016/j.psyneuen.2005.03.005

13. Gotlib IH, Joormann J. Cognition and depression: Current status and future directions. Annu Rev Clin Psychol. 2010;27(6):285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

14. Bearden CE, Glahn DC, Monkul ES, Barrett J, Najt P, Villarreal V, et al. Patterns of memory impairment in bipolar disorder and unipolar major depression. Psychiatry Res. 2006;142(3):139–150. https://doi.org/10.1016/j.psychres.2005.08.010

15. Wells GL, Kovera MB, Douglass AB, Brewer N, Meissner CA, Wixted JT. Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. Law Hum Behav. 2020;44(1):3–36. https://doi.org/10.1037/lhb0000359

16. Sauerland M, Raymaekers LH, Otgaar H, Memon A, Waltjen TT, Nivo M, et al. Stress, stress-induced cortisol responses, and eyewitness identification performance. Behav Sci Law. 2016;34(4):580–594. https://doi.org/10.1002/bsl.2249

17. Smeets T, Cornelisse S, Quaedflieg CW, Meyer T, Jelicic M, Merckelbach H. Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. Psychoneuroendocrinology. 2012;37(12):1998–2008. https://doi.org/10.1016/j.psyneuen.2012.04.012

18. Gabbert F, Hope L, Fisher RP. Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. Law Hum Behav, 2009;33(4)298–307.

19. Team RC. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

20. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48. https://doi.org/10.18637/jss.v067.i01

21. Diamond DM, Campbell AM, Park CR, Halonen J, Zoladz PR. The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. Neural Plas. 2007;2007(1):1–33. https://doi.org/10.1155/2007/60803

22. Baritaki S, De Bree E, Chatzaki E, Pothoulakis C. Chronic stress, inflammation, and colon cancer: A CRH system-driven molecular crosstalk. J Clin Med. 2019;8(10):1669–1695. https://doi.org/10.3390/jcm8101669

AUTHORS:
Colin Tredoux[1,2]
Ahmed M. Megreya[3]
Alicia Nortje[1]
Kate Kempen[1]

AFFILIATIONS:
[1]Department of Psychology, University of Cape Town, Cape Town, South Africa
[2]Cognition, Languages, Language, Ergonomics Laboratory, University of Toulouse – Jean Jaurès, Toulouse, France
[3]Department of Psychological Sciences, Qatar University, Doha, Qatar

CORRESPONDENCE TO:
Colin Tredoux

EMAIL:
colin.tredoux@uct.ac.za

DATES:
Received: 31 Aug. 2021
Revised: 25 Nov. 2022
Accepted: 25 Nov. 2022
Published: 29 Mar. 2023

HOW TO CITE:
Tredoux C, Megreya AM, Nortje A, Kempen K. Changes in the own group bias across immediate and delayed recognition tasks. S Afr J Sci. 2023;119(3/4), Art. #12126. https://doi.org/10.17159/sajs.2023/12126

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Leslie Swartz

KEYWORDS:
own group bias, own race bias, face recognition, encoding

FUNDING:
Programme for the Enhancement of Research Capacity (PERC), University of Cape Town; South African Centre for Digital Language Resources (SADILAR), Department of Science and Technology of South Africa as part of the South African Research Infrastructure Roadmap (SARIR); Oppenheimer Memorial Trust

Changes in the own group bias across immediate and delayed recognition tasks

Face recognition is biased in favour of in-group identity, particularly strongly for race or ethnicity but to some extent also for sex and age. This 'own group bias' (OGB) can have profound implications in practical settings, with incorrect identification of black suspects by white witnesses constituting 40% of criminal exonerations investigated by the Innocence Project. Although authors have offered several explanations for the OGB in face recognition, there is little consensus, apart from the acknowledgement that the bias must reflect perceptual learning history. One matter that is not currently clear is whether the bias occurs at encoding, or at retrieval from memory. We report an experiment designed to tease out bias at encoding, versus bias at retrieval. Black and white South African participants encoded 16 target faces of both the same and other race and gender, and attempted immediately afterward to match the target faces to members of photograph arrays that either contained or did not contain the targets. After a further delay, they attempted to identify the faces they had encoded from memory. Results showed a strong crossover OGB in the delayed matching task, but an asymmetrical OGB at retrieval (only white participants showed the OGB). Further investigation of recognition performance, considering only images correctly matched in the delayed matching task, showed a narrowly non-significant OGB at retrieval, but the investigation was likely not sufficiently powered to discover the effect, if it exists.

Significance:

- We demonstrate the presence of a crossover OGB in face recognition in a sample of black and white South Africans in a delayed matching task (a measure of encoding).
- Our findings show that the OGB may change rapidly. In the present study, the OGB took a crossover form at retrieval immediately after encoding, but was asymmetrical when assessed shortly afterwards.
- We used a novel approach for disentangling effects at encoding and at retrieval, but do not provide clear evidence to distinguish whether the OGB is a failure of encoding or of memory retrieval.

Introduction

The own group bias (OGB)[1] refers to a recognition advantage for faces of members of one's own group. The most common formulation posits an unequal advantage: groups that co-occupy a common environment will show an advantage for members of their own group over members of another group. However, the recognition advantage is typically asymmetrical in countries or regions where demographic or economic representation is unequal (e.g. the United Kingdom[2]). Members of the dominant group will usually show an OGB, but members of the subordinate groups often do not show it.[2,3] The practical implications of the OGB are profound. The Innocence Project in the USA[4] has shown, for instance, that over 40% of exonerations of people falsely imprisoned based on eyewitness identifications were convicted in part as a result of cross-race eyewitness identifications – a figure far out of proportion to the demography of the USA.

Explanations of the OGB in the literature are plentiful, but it is probably fair to say that no single account has won out empirically or theoretically. All theories appear to accept the notion that the OGB must be a consequence to some degree of differential exposure to one's own and other group faces.[5] Our interest is in what might be called the micro-chronology of the OGB. Whereas the OGB is often referred to in the literature as a memory bias – for instance Hugenberg and colleagues declaring their intent to "…explain the proliferation of own group biases in face memory"[6(p.1392)], and Yaros and co-authors[7] arguing that the OGB may emerge due to "tuned" memory mechanisms – researchers have pointed out for a number of years that the bias may stem from the differential ability to encode faces of other groups, rather than a reduced ability to recognise or retrieve them from memory[8]. Recent claims are that the OGB is likely due to the less efficient encoding of out-group faces in visual working memory[9,10], or to preferential distribution of attention during encoding[11].

Megreya and co-authors[12] have published research that questions the basis of the OGB in memory processes, arguing instead that the OGB may be entirely due to encoding difficulties. Participants in Megreya et al.'s study had difficulty in matching the face of a 'target' in a film still to digital photographs of the same person. This finding was more pronounced when participants attempted to match faces from a group different from their own. This outcome has important implications, especially for theoretical explanations of the OGB bias: many theoretical accounts have focused on memory retrieval operations or are at least unclear on the relative role of encoding and retrieval in the provenance of the OGB, and this should be addressed if the phenomenon to be explained is entirely about encoding rather than retrieval. But this may be somewhat premature: at this stage in the development of thinking about the OGB, it is not clear what the relative contribution of each process is. Indeed, several authors have offered evidence that either questions whether the deficit is in any way encoding based, or that posits that it is in part retrieval based. Papesh and Goldinger[13] were able to show that disruptions during retention intervals modulated the OGB for out-group faces, but not for in-group faces. Stelter et al.[14] showed that eye movement activity differed across in-group and out-group targets during a recognition, but not an encoding phase. They also showed that the OGB was a function of poor performance in recognition of new faces specifically.

Our aim was to explore the relative contributions of encoding and retrieval in the manifestation of the OGB in face recognition. It is important at the outset to acknowledge that it is difficult to disentangle encoding from recognition, and in this respect the difficulty appears to lie primarily in establishing that a stimulus has been encoded. We offer one approach, which is to test recognition with a delayed matching task, and then after a longer period with a recognition memory task. In a delayed matching task, participants view a stimulus, and then immediately attempt to match the stimulus in a test array once it has been removed from view. If participants are accurate in the delayed matching task, this suggests successful encoding. Note that Megreya et al.[12] did not use a delayed matching task, but an in-view matching task. In other words, they attempted to match a target image to one embedded in an array of images while both the target and the array were in full view. However, we believe that in-view matching tasks do not demonstrate encoding as clearly as delayed matching tasks because matching targets in-view might be accomplished without encoding faces at all. Participants may be able to use non-face properties of stimuli, such as colour temperature, or other image artefacts to effect the match, or might be able to do so with minimal encoding of the target image, e.g. by relying on specific, transient and localised characteristics of faces, such as skin blemishes, and hairstyle oddities. This distinction between face identity encoding and face image encoding was powerfully made over 40 years ago.[15] For this reason we departed from the procedure used by Megreya et al.[12] There were some further departures, too, that we thought important. In the first instance, we wished to see whether the results they obtained replicate across different viewpoints and with different stimulus materials. They used only frontal views of participants, in both matching and target views, and they also appeared to have made their matching arrays very difficult (the faces in the array were the most similar in a moderately large set). Repeating their experiment with different viewpoints of the faces (e.g. in profile vs. frontal views) would test whether the memory trace of the face is well enough encoded to match and recognise it over different transformations (i.e. different face poses). Furthermore, repeating their experiment with line-up arrays that are constructed around faces that are not specifically selected for high similarity seems important to us, especially as one of the most significant potential applications of knowledge about the OGB is in the eyewitness domain, where line-ups are unlikely to be constructed in this way (see Wells et al.[16] for a discussion of why this is not a good idea).

However, the most important goal in our experiment was to extend Megreya et al.'s[12] original experiment and to attempt to separate out encoding and retrieval processes, and to estimate the relative importance of each in the manifestation of OGB bias. Thus, in our experiment, participants were tested for delayed face matching ability, and, after a suitable delay, for their ability to recognise the face to which they were exposed in the delayed matching task. We used a line-up task, designed to permit the calculation of the signal detection theory (SDT) measures *d'* (discrimination) and *c* (criterion). Through this, and by calculating successful recognition of faces that were incorrectly matched in the first part of the procedure, we attempted to assess the relative contributions of encoding and retrieval deficits in the OGB effect.

Method

Target stimuli

Stimuli were digital images taken from a large database of black South African and white South African faces (containing multiple views of over 1000 individuals) and curated by the first author. Sixteen target faces were used in this experiment: four male and four female black South African faces, and the same again for white South African faces. The target faces were randomly sampled by the authors from the total collection, and target ratings were obtained to ascertain whether there were any idiosyncrasies present in the target images that may have resulted in them being more memorable than the other target faces. A total of 22 (15 female) (M_{age} = 20.52 years; SD_{age} = 1.21 years) participants rated the target faces on a number of dimensions that are known to affect face recognition: typicality, distinctiveness, attractiveness, perceived criminality, age, wealth, memorability, and familiarity. Each face was rated on a scale from 0 (not at all) to 8 (extremely), whereas age was estimated numerically. Two images of the same target were presented side-by-side during the rating task: one photograph was of the target in a neutral three-quarter side pose, the other image was of the target in a frontal casual/smiling pose. Each image pair was presented one at a time, along with a randomised order of rating dimensions. The size of each face image was approximately 7.94 cm in width and 10.35 cm in height, with a resolution of 300 x 391 pixels (8-bit). The image background and the target clothing were edited to be standard and consistent across all the target images.

Line-up construction

Target present (TP) and target absent (TA) photographic line-ups were constructed for the target images. TP line-ups contained the target and five foils. In the TA line-ups, the target was replaced with a foil that was randomly selected by the authors, resulting in a six-foil line-up (i.e. there was no designated suspect). No foils were repeated or appeared in any of the other line-ups. The line-up members were selected to be subjectively moderately similar in appearance to the target (i.e. following the principle set down in Wells et al.[16] that line-up members should be matched to the target on some but not all characteristics). Three of the present authors independently selected ten possible foils for each target face from a database of hundreds of black South African and white South African faces. The most frequently chosen foils were used as the final images in the line-ups.

Corresponding to the 16 target faces, each participant saw 16 line-ups. The line-ups appeared in one of two orders: eight of the line-ups were TP line-ups, and the other eight were TA line-ups. Two line-ups were created for each target face: a frontal pose line-up, and three-quarter pose line-up. The frontal neutral pose and the three-quarter pose were used to control for picture recognition and to ensure that identification was made on memory for the target and not the target photograph. The line-up photographs were in colour and standardised. The backgrounds were edited to remain consistent across all the images. All clothing, jewellery, and distinctive markings were digitally removed from the images. The size of each face image was approximately 6 cm x 7.8 cm, with a resolution of 227 x 296 pixels (8-bit). The image background and the target clothing were edited to be consistent across the target images.

Participants

A total of 64 (53 female) participants from the University of Cape Town participated in the study in exchange for course credit; 32 (50%) of the participants identified themselves as white South Africans, and 32 (50%) of the participants identified themselves as black South Africans. All participants reported normal or corrected to normal vision.

Procedure

Encoding and delayed matching phase

The study was conducted at the University of Cape Town, in a quiet computer laboratory. The experiment was presented on computer using E-Prime 2.0, at a resolution of 1024 x 768 pixels.

After providing demographic information, participants were informed that they would be presented with a series of target faces, one at a time, for five seconds. One of the target faces appeared on the screen, in a frontal casual/smiling pose. The size of the target image was 12 cm x 15.7 cm, at a resolution of 456 x 594 pixels.

After five seconds, the face disappeared, and a six-member simultaneous TP or TA line-up was immediately displayed, in a delayed matching task. Participants either saw a three-quarter line-up or a frontal line-up. Participants were informed that the target they had just seen may or may not be present in the line-up. They were cautioned that the clothing and background of the target image may be different from the image they had studied. If participants recognised the target, they had to indicate the corresponding number above the line-up member on the keypad; if they thought the target was not present, they had to indicate '0' on the keypad. This delayed matching procedure was repeated for the

remaining 15 target faces; the order of the target faces and matching line-up pairs was randomised across participants.

When the line-up delayed matching phase was finished, participants completed a filler task for 10 minutes (a Sudoku game).

Recognition phase

Following the distractor task, participants were told that they would view a series of six-member line-ups and that their task was the same as before – to select the target face that they had studied earlier. They were informed that the line-ups may be different from the line-ups they had seen in the delayed matching session. Participants were again cautioned that the target face may or may not be present in the line-up. It was emphasised that the line-ups would appear in a random order and would not necessarily correspond with the order in which the participants had studied the target faces.

Participants attempted to identify the targets from the same line-ups used in the delayed matching procedure. For example, if they saw a TA three-quarter pose line-up at delayed matching, they saw the same line-up at recognition. The position of the line-up members was rearranged between delayed matching and recognition to guard against commitment effects – if participants recognised the line-up from the delayed matching session, they may have been tempted to select the same line-up position as before.

If participants recognised the target, they indicated the corresponding number above the line-up member on the keypad; if they thought the target was not present, they selected the appropriate option on the keypad. This recognition procedure was repeated for the remaining 15 line-ups; the order of arrays was randomised across participants.

When all recognition trials were completed, participants who responded '0' (not present) to a TP line-up – i.e. who had incorrectly rejected the line-up – were given another chance to respond to the line-up. They were shown the same line-ups to which they had responded 'not present' and were then required to make a selection (i.e. a forced choice; however, data from the forced choices were not included in the present analyses). The order of these line-ups was randomised.

Once participants had completed their forced-choice line-up tasks, they were thanked, debriefed, and dismissed from the study. Please note that further details regarding preparation of stimuli, as well as descriptive statistics of task performance, appear in Figure 1.

Results

Hit and false alarm scores were computed for each participant and were used to derive discrimination (*d'*) and criterion (*c*) signal detection measures. Because we did not use designated suspects in our TA line-ups, false alarms were computed as the total number of identifications of foils in TA line-ups. (Although eyewitness researchers often divide the total number of foil identifications to estimate the false alarm rate, this constrains the false alarm rate to a maximum of 1/number_of_line-up members. This may be appropriate for a perfectly fair line-up, but very few line-ups are perfectly fair.[17] We have not divided by the number of foils. The consequence is that false alarms are more probable than hits, and this can be expected to decrease *d'*, but as long as one recognises that it is comparisons across *d'* and *c* that are important, not absolute levels, it should make little difference. Judging absolute levels might lead one to believe that participants in this study had poor discrimination in some conditions (indeed, $d' < 0$ in some conditions), but that is not a real reflection of ability).

Our analyses are generally based on these SDT measures, rather than on raw hit, false Alarm, or other untransformed accuracy measures. Figure 2 shows raw measures for the information of interested readers.

We conducted a four-way linear mixed model analysis of the experimental data, that is stimulus group (black vs. white) X line-up face view (frontal vs. three-quarter) X task condition (delayed matching vs. recognition) X participant group (black vs. white), using the package LME4[18], within R[19]. 'Stimulus group' and 'task condition' were within-participant factors. The analysis of variance revealed several significant effects, as shown in Table 1.

At the level of main effects, we observed differences in *d'* across 'line-up face view', and for 'task condition' ($p < 0.05$). We found two-way interactions for both 'task condition and stimulus group', and for 'stimulus group and participant group'. Finally, we found three-way interactions for 'participant group, stimulus group, and task condition', as well as for 'stimulus group, task condition and line-up face view'.

The most important findings from the analysis for our concerns here are the classic two-way interaction of 'stimulus group and participant group' (the OGB effect), and the three-way interaction of 'stimulus group,

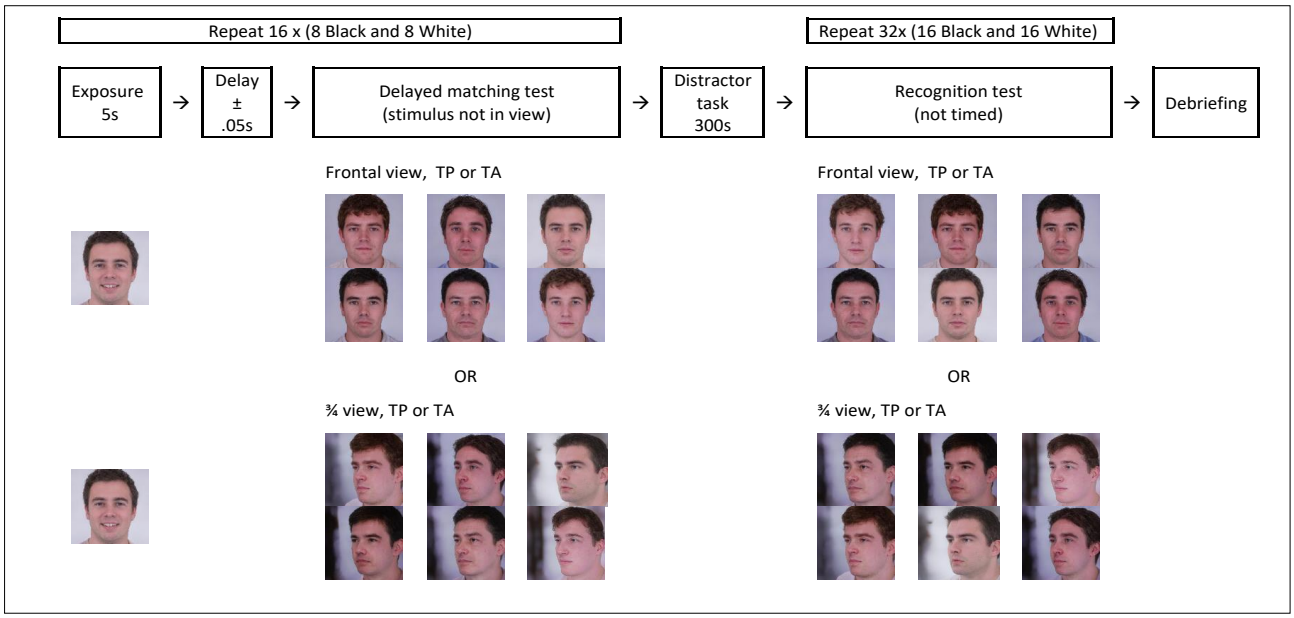


Figure 1: Experimental procedure, showing timeline and other details. For display purposes, faces are synthetic, i.e. not of real people. Assignment to target present (TP) or target absent (TA), and to frontal or three-quarter view, was random. Faces at exposure were casual or smiling, and neutral at delayed matching and/or recognition.

participant group, and task condition'. The latter effect is particularly important to our enquiry – it tells us that the OGB effect in the present sample differs between the two types of task condition (delayed matching and recognition).

Table 1: Mixed linear model ANOVA on discrimination (*d'*), and criterion (*c*), by participant group, stimulus group, task condition (delayed match vs recognition), and view (frontal vs three-quarter). Only fixed effects are shown; for random effects ICC = 0.49 for *d'*, and 0.48 for *c*. β coefficients are standardised. All $df_1 = 1$, and $df_2 = 60$.

	Discrimination (*d'*)			Criterion (*c*)		
Effect	**F**	***p*-value**	**| β |**	**F**	***p*-value**	**| β |**
Participant group	0.03	0.869	0.87	3.40	0.070	0.53
Stimulus group	0.01	0.905	0.55	0.50	0.484	0.54
Task condition	34.37	0.001	0.93	12.58	0.001	0.00
View	11.50	0.001	0.89	2.73	0.104	0.10
Participant : Stimulus	26.29	0.001	1.08	5.70	0.020	1.25
Participant : Task	0.20	0.653	0.29	0.01	0.912	0.81
Stimulus : Task	5.05	0.028	0.30	0.00	0.958	0.72
Participant : View	2.38	0.128	0.49	0.37	0.546	0.81
Stimulus : View	0.24	0.627	0.53	0.61	0.439	0.40
Task : View	9.42	0.003	0.28	0.03	0.871	0.51
Participant : Stimulus : Task	4.10	0.047	0.41	3.91	0.053	1.19
Participant : Stimulus : View	0.12	0.727	0.36	0.89	0.350	0.87
Participant : Task : View	0.89	0.349	0.08	0.51	0.476	0.86
Stimulus : Task : View	2.74	0.103	0.76	0.45	0.506	0.74
Participant : Stimulus : Task : View	0.53	0.467	0.47	2.09	0.154	1.01

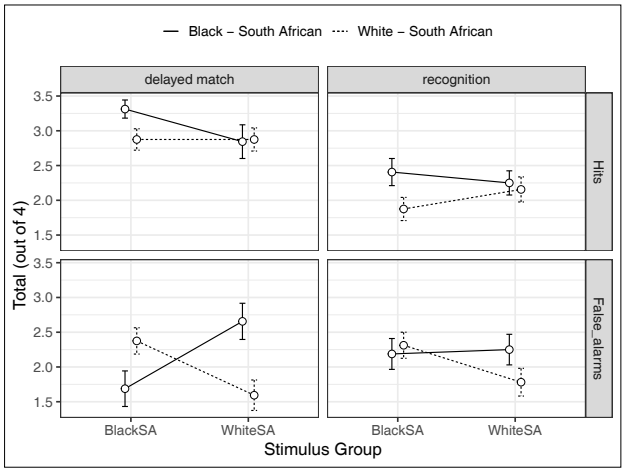


Figure 2: Hits and false alarms as a function of stimulus group, race of respondent, and task type. Error bars are standard errors of the mean.

A classic bias in face recognition ability is for own group members to recognise faces belonging to their own group with greater facility than members of other groups. Megreya et al.[12] have argued though that this facility is also present when participants are asked to match faces to arrays: own-group faces are matched more accurately than out-group faces. This raises the question of whether there is a retrieval component at all in the own-group bias in face recognition – the effect could be entirely due to poorer encoding of out-group faces. In the present experiment, we observed the classic crossover interaction of stimulus and participant group, on *d'*, as shown in Figure 3. White participants were better at delayed matching and recognising white faces (mean *d'*=0.52, SD=0.95) than they were at delayed matching and recognising black faces (mean *d'*=0.02, SD=0.89), and black participants were better at delayed matching and recognising black faces (mean *d'*=0.56, SD=1.22) than they were at delayed matching and recognising white faces (mean *d'*=0.04, SD=1.12). A set of linear contrasts showed that white South Africans performed better, overall, with white South African faces than with black South African faces (t(60)=3.54, SE=0.14, $p<0.001$, Cohen's d=0.52. Cohen's d was estimated by dividing the mean difference by the square root of the sum of the variances of the random, and residual effects[20]), and that black South Africans performed better with black South African faces than with white South African faces (t(60)=3.71, SE=0.14, $p<0.001$, Cohen's d=0.54). As we were interested in testing whether viewing faces in frontal vs three-quarter view would impact the OGB, it is worth noting that the 'view' factor was not implicated in any interaction involving either participant group or stimulus group, but it was involved in a significant interaction with task condition, showing that participants found it difficult to match or recognise faces across frontal and three-quarter views, but this effect was greater for the delayed matching condition than for the recognition condition ($\Delta d'$=0.89 vs 0.34).

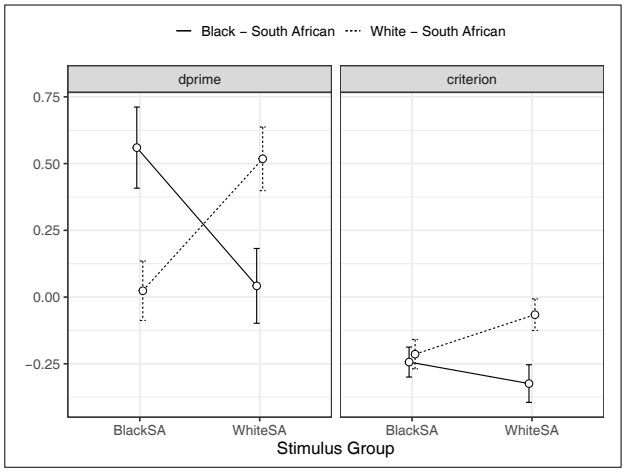


Figure 3: Own group bias in discrimination (*d'*) and criterion (*c*), collapsed over delayed matching and recognition tasks. Error bars are standard errors of the mean. Both interactions shown here are statistically significant.

Most pertinent to our concerns was the three-way interaction of participant group, stimulus group, and task condition – or in other words, the two-way interaction of participant group and stimulus group (the OGB effect) considered as a function of which task was being completed. As Table 1 shows, this effect was significant, and Figure 4 shows the effect graphically. Follow-up contrasts showed that, in the delayed matching task, both groups of participants were better at matching their own-group faces: black South Africans were better at matching black South African faces than white South African faces ($M_{black\text{-}black}$=0.98 vs $M_{black\text{-}white}$=0.1, $SD_{black\text{-}black}$=1.02 vs $SD_{black\text{-}white}$=1.35, t(118)=4.78, $p<0.001$, Cohen's d=0.92), and white South Africans were better at matching white South African faces than black South African faces ($M_{white\text{-}white}$=0.79 vs $M_{white\text{-}black}$=0.31, $SD_{black\text{-}black}$=0.93 vs $SD_{black\text{-}white}$=0.88, t(118)=2.57. $p<0.012$, Cohen's d=0.50).

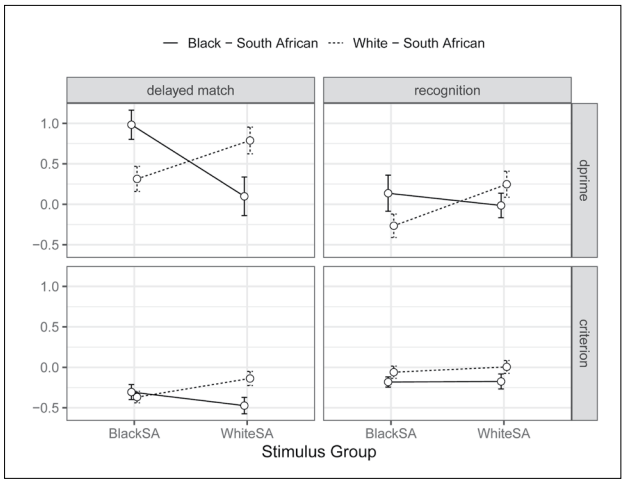


Figure 4: Own group bias in discrimination (*d'*) and criterion (*c*), across delayed matching and recognition tasks. Error bars are standard errors of the mean. At matching, both black and white participants show the OGB in *d'*, but only the white participants show the OGB at recognition. The interactions between stimulus group and participant group are not significant.

In the recognition task, however, white South Africans showed the OGB ($M_{white\text{-}white}$=0.25 vs $M_{white\text{-}black}$= -0.27, $SD_{white\text{-}white}$=0.91 vs $SD_{white\text{-}black}$=0.82, t(118)=2.8, p<0.007, Cohen's d=0.54), whereas black South Africans did not ($M_{black\text{-}black}$=0.14 vs $M_{black\text{-}white}$= -0.01, $SD_{black\text{-}black}$=1.26 vs $SD_{black\text{-}white}$=0.87, t(118)=0.82. p<0.416, Cohen's d=0.16).

In addition to *d'*, we computed criterion for each participant in the experiment. The analysis of criterion as the dependent variable for the full four-way design produced a significant two-way interaction between participant group and stimulus group (F(1, 60)=5.70, p<0.02). This interaction shows that white and black participants were both more conservative when considering faces of their own group than when considering faces of the other group ($M_{white\text{-}white}$= -0.07 vs $M_{white\text{-}black}$= -0.21, $SD_{white\text{-}white}$=0.47 vs $SD_{white\text{-}black}$=0.44, and $M_{black\text{-}black}$= -0.24 vs $M_{black\text{-}white}$=-0.32, $SD_{black\text{-}black}$=0.45 vs $SD_{black\text{-}white}$=0.56). The three-way interaction between participant group, stimulus group, and task condition was narrowly not significant (F(1, 60)=3.91, p<0.054). We show the marginally non-significant three-way interaction in Figure 4, where it seems that (1) there is a typical OGB for criterion present, and that (2) the OGB may be apparent in the delayed matching task, but not in the recognition task.

Thus far the results have shown a clear crossover OGB in the delayed matching task, as well as an asymmetrical OGB in the recognition task. It is not clear that the OGB in the recognition task can be said to be independent of that in the delayed matching task – that is, it seems evident that faces that are not matched or recognised accurately in the delayed matching task will not be recognised accurately in the later task either. Failure in the matching task may also be an indication of encoding failure. In order to test the point that Megreya et al. make about the possible entire dependency of the OGB on encoding processes, we reduced our data set to only those faces that had been accurately matched in the delayed matching phase and re-analysed the recognition data for just those faces. The idea here was to choose faces that we were confident had been encoded successfully.

One implication of selecting only faces that were correctly matched in the delayed matching task, is that the OGB shown in the delayed matching task was effectively controlled for – as all faces were correctly matched, there could be no bias. This reduction left us with 78% of our original participants (selecting those who made at least one correct decision at delayed matching), but as some participants performed better than others in the delayed matching task, this resulted in an uneven distribution of stimuli across conditions. Thus, to take the extreme conditions, whereas 86% of the original stimuli were taken into account for black participants viewing white South African faces in TP conditions, 46% of stimuli were taken into account for white participants viewing black South African faces in TA conditions. This 'attrition' undoubtedly affected the potential power of analyses of recognition of stimuli that had been successfully matched in the delayed matching task.

A linear mixed model testing discrimination (*d'*) for faces correctly matched in the delayed matching task showed a non-significant effect for the key two-way interaction of interest (between participant group and stimulus group), that was significant in the earlier model (F(1, 49.35)=2.80, p<0.104). Although the interaction effect was not significant, it seemed appropriate to us to follow up with focused contrasts directly exploring a potential OGB, and these showed that, in the recognition task, testing recognition of only faces successfully matched in the delayed matching task, neither white nor black South Africans were better at their own-group faces, although this was narrowly not the case for white South Africans (for black South Africans: $M_{black\text{-}black}$=0.38 vs $M_{black\text{-}white}$=0.29, $SD_{black\text{-}black}$=0.59 vs $SD_{black\text{-}white}$=0.46, t(64.8)=0.18, p<0.86, Cohen's d=0.18, and for white South Africans: $M_{white\text{-}white}$=0.33 vs $M_{white\text{-}black}$=0.09, $SD_{white\text{-}white}$=0.43 vs $SD_{white\text{-}black}$=0.47, t(43.5)=1.94. p<0.059, Cohen's d=0.48). While the asymmetrical OGB seen in the recognition test using all stimuli was not technically significant in the recognition test using only faces successfully encoded (matched in the delayed matching task), it was very close to being so. Of course, this analysis is based of necessity on a smaller sample, and has less statistical power, so it is not strong evidence that controlling for the OGB in the matched delay task (our operationalisation of encoding) eliminates the OGB at recognition. We also conducted a mixed linear model analysis on criterion scores, but did not find any significant results, or any suggestion of an effect (all p>0.48).

Discussion

The OGB in face recognition is a well-established phenomenon, with serious consequences in applied contexts, especially when recognition is treated as person identification, as one finds in law enforcement procedures. Despite often being the subject of empirical and theoretical investigation, not much is known about the cognitive processes underlying the phenomenon. One important question concerns whether the OGB is an encoding or a recognition phenomenon, and we have brought results from an empirical study to bear on this question, or at least on the micro-chronology of the OGB.

Black and white South African participants in our study were asked to match black or white target faces to corresponding images in arrays of same group faces immediately after viewing them. They were then asked, after an intervening period in which they completed a distractor task, to recognise the target faces from memory, in the same arrays. Our results show a strong OGB in discrimination accuracy for both black and white participants at the delayed matching phase of the study, but only white participants showed the bias at the recognition memory phase of the study. We found similar results for the measure of response criterion that we computed, although the result for the important interaction of participant group, stimulus group, and task condition was narrowly not significant. Our finding of an asymmetry in the delayed recognition task for discrimination, rather than in the delayed matching task, is partly novel, and partly in line with extant research. Most studies that have been conducted on the OGB (for race) in South Africa have reported an asymmetric OGB at recognition. Usually, white participants show a strong OGB, whereas black participants rarely show an OGB at all (but see Wittwer et al.'s Study 6[21] for an example of a crossover), and sometimes show an inverse OGB, recognising white faces with more accuracy than they do black faces[22]). Some authors, including ourselves, have ascribed this to the way in which the social and political context have structured intergroup contact in South Africa (and other countries with similar intergroup histories), and therefore perceptual contact. It is not easy, though, to explain the finding of a crossover OGB among black participants at encoding (the delayed matching task) that fails to materialise in the recognition test. One possibility is that performance on the recognition task bottomed out, approaching basement level, but this does not seem to have occurred differentially for black vs white participants, and is not a convincing explanation. It is possible that it could reflect a real difference in memory consolidation of out-group faces by black participants. What does seem clear to us is that the argument that

the typical asymmetric OGB between majority and minority groups is due to socio-political factors – e.g. that social and political inequality might impact perceptual learning of out-group faces – is made more difficult to sustain by our results. If we had considered only data from the recognition task we might well have reported another instance of an asymmetric OGB, but in looking at the encoding/delayed matching task, we noted a crossover OGB. A source of evidence that is sometimes marshalled in an attempt to understand the OGB as it manifests in particular contexts is the perceptual contact history of participants. Methods of assessing past perceptual history generally rely on self-reporting, and in the meta-analysis reported in Meissner and Brigham[1] were only poorly correlated with the OGB ($r=0.13$). We did not collect self-reported data on perceptual contact history, but do not believe it would have shed any light on the nature of the OGB in our study, given its poor predictive record reported in Meissner and Brigham[1]. We thus do not have a good sense of why we observed an OGB at encoding but not at recognition for black participants. An anonymous reviewer has suggested that it could be a task-dependent difference, possibly interacting with social-motivational factors, which we think is possible. On the more general point of why there is often an asymmetric pattern for the OGB, with disadvantaged groups showing a weaker or non-existent OGB when compared to advantaged groups, we think that Malpass'[23] notion of a 'social interaction utility' is useful: members of the disadvantaged group have a positive utility associated with interaction and person recognition (often being economically dependent on the advantaged group, it is important to interact with, and recognise members of that group), whereas members of the advantaged group have a negative utility (there is no clear advantage to interaction and recognition).

An important question in research on the OGB concerns the degree to which the OGB is a function of encoding, or of recognition. The use of delayed matching and recognition tasks allowed this to be addressed in our study. We conducted a second analysis on images correctly matched (therefore showing no OGB), and although we did not find an OGB for these images at recognition using a statistical significance test as the criterion, the size of the effect we observed in this task ($d=0.48$) was very similar to that observed in the recognition task ($d=0.52$). Our results likely have equivocal bearing in the end on whether the OGB is an encoding phenomenon, rather than a recognition memory phenomenon. What is most interesting about our findings, though, and which has nothing to do with the final task we used, is that we observed a strong crossover OGB at delayed matching, which is not at all usual in South African studies, and after a brief delay of 5 minutes, the OGB manifested in a recognition memory task as an asymmetrical effect, which is typically the form other studies in South Africa have previously reported.

The claim that the OGB is an encoding phenomenon makes sense from a perceptual learning viewpoint, especially in line with Valentine's multidimensional face space model[24], and his explanation of the OGB in terms of the model[25]: the dimensions available to represent out-group faces are fewer, and less well developed than for in-group faces, and encoding of out-group faces will be less well differentiated from other exemplars of that group. Of course, our method for separating encoding and recognition processes is admittedly ad hoc: we tested recognition of faces that had been correctly matched, thus attempting to control encoding processes. We did not directly show that own-group faces were better encoded than out-group faces, although in demonstrating an OGB at encoding we think that differences were implicitly demonstrated. Face encodings are complex patterns of electrical activation across multiple brain regions, and not directly accessible, although there are some clues to the neural underpinnings of the OGB and its connections to socio-affective processes.[26] Some researchers have shown differential event-related potentials to own- and other-group faces[27], with a potentially special role for the P200 visual component of the event-related potentials[28]. These investigations suggest that we are making progress toward a better understanding of the brain mechanisms underlying the OGB.

Apart from the obvious limitation of not being able to assess encoding proficiency directly, there are other limits to what we are able to conclude from our study. An important methodological limitation may be our method of checking that a stimulus had been encoded, and the knock-on effect of this for the accuracy of our recognition test. As a reminder, we assessed encoding with a delayed matching task: participants attempted a match immediately after viewing a target stimulus. After a further delay of 5 minutes, we assessed recognition memory by asking participants to choose the target image from a line-up, which either contained or did not contain the target. Our recognition test could thus be said to have exposed participants to the target stimulus twice, and this taints the comparison between encoding and recognition in which we are interested. In other words, although the delayed matching task tests memory for a stimulus seen once, the recognition task tests memory for a stimulus seen once on its own, and once within an array. It is important to note that participants were not told the position of the target in the array in the delayed matching task, and that the photographs of the target changed from the original presentation (frontal casual/smiling view) to that in the line-up arrays used in the delayed matching and recognition tasks (three-quarter profile, or neutral/passport style view). It is still possible that there was some strengthening of the memory assessed in the recognition memory task, but our results show that memory performance decreased considerably between the delayed matching task and the recognition memory task (see Figures 3 and 4), so it does not seem to have counteracted that decrease, if at all. It would also have been unlikely to affect our white and black participants differentially, and because our key effect of interest was the nature of the OGB at the two time points, it does not appear to us to be confounded with the two- or three-way interactions in which we were interested.

Whereas we studied matching and recognition of two-dimensional images of faces, face recognition in real contexts is of three-dimensional surfaces that change locally, and globally, over time, in interaction with perceivers. We also used a short delay period between encoding (matching), and recognition, unlike many face recognition tasks 'in the wild'. However, it is likely that the OGB would be higher in more naturalistic contexts, because there are known additional biases across groups for emotion[29] and age[30], among other biases, and we do not see a reason to believe that this would differentially affect encoding and recognition deficits already present. There is an extensive and convincing argument in support of the view that one expects real witnesses to perform far worse than participants in laboratories.[31]

In conclusion, although we have not reported clear evidence in favour of the view that the OGB in face recognition is likely a consequence of poor encoding of other group faces, we have identified an interesting, rapid change in the manifestation of the OGB. Whilst it took a strong crossover form at encoding (delayed matching), it reverted to an asymmetric form after a brief delay. There are some applied implications of this result, although the evidence is at this stage too slight to base strong recommendations on. If the OGB is a failure of encoding then there may be little justification for devising methods that focus on improving recall or recognition of out-group faces, and it may be better to develop training programmes that focus on encoding processes. There is some evidence that in-group members focus on different face regions when encoding out-group faces than out-group members do[32], although it is not yet clear that reshaping cross-group encoding will work[33]. On the other hand, if the OGB is a failure of retrieval, some methods such as the Cognitive Interview[34] and the Person Description Interview[35] may be useful interventions when recovering information from memory about face appearance and identity – although it should be borne in mind that these methods are typically good for improving recall memory, but not recognition memory[36]. It could also be that the OGB is a function of both encoding and retrieval processes, and we are presently considering ways of adapting our procedure to accommodate this possibility.

Ethical considerations

Ethical clearance to conduct the study was granted by the Department of Psychology Ethics Committee of the University of Cape Town. All participation was voluntary and informed consent was obtained before study commencement.

Competing interests

We have no competing interests to declare.

Authors' contributions

C.T.: Conceptualisation, method, data analysis, data curation, writing, project leadership, funding acquisition, student supervision. A.M.M.: Conceptualisation, method, writing, project leadership. A.N.: Conceptualisation, method, data collection, data analysis, writing – revision. K.K.: Conceptualisation, method, data collection, writing – revision.

References

1. Meissner CA, Brigham JC. Thirty years of investigating the own-race bias in memory for faces. Psychol Public Policy Law. 2001;7(1):3–35. https://doi.org/10.1037/1076-8971.7.1.3
2. Walker PM, Hewstone M. A perceptual discrimination investigation of the own-race effect and intergroup experience. Appl Cogn Psychol. 2006;20(4):461–475. https://doi.org/10.1002/acp.1191
3. Malpass RS, Kravitz J. Recognition for faces of own and other race. J Pers Soc Psychol. 1969;13(4):330–334. https://doi.org/10.1037/h0028434
4. Innocence Project. How eyewitness misidentification can send innocent people to prison [webpage on the Internet]. c2020 [cited 2023 Feb 03]. Available from: https://innocenceproject.org/how-eyewitness-misidentification-can-send-innocent-people-to-prison/
5. Rhodes G. Face recognition. In: Reisberg D, editor. The Oxford handbook of cognitive psychology. New York: Oxford University Press; 2013. p. 46–68.
6. Hugenberg K, Wilson JP, See PE, Young SG. Towards a synthetic model of own group biases in face memory. Vis Cogn. 2013;21(9–10):1392–1417. https://doi.org/10.1080/13506285.2013.821429
7. Yaros JL, Salama DA, Delisle D, Larson MS, Miranda BA, Yassa MA. A memory computational basis for the other-race effect. Sci Rep. 2019;9(1):19399. https://doi.org/10.1038/s41598-019-55350-0
8. Bornstein BH, Laub CE, Meissner CA, Susa KJ. The cross-race effect: Resistant to instructions. J Criminol. 2013;2013:745836. https://doi.org/10.1155/2013/745836
9. Zhou X, Mondloch CJ, Emrich SM. Encoding differences affect the number and precision of own-race versus other-race faces stored in visual working memory. Atten Percept Psychophys. 2018;80(3):702–712. https://doi.org/10.3758/s13414-017-1467-6
10. Stelter M, Degner J. Investigating the other-race effect in working memory. Br J Psychol. 2018;109(4):777–798. https://doi.org/10.1111/bjop.12304
11. Brown TI, Uncapher MR, Chow TE, Eberhardt JL, Wagner AD. Cognitive control, attention, and the other race effect in memory. PLoS ONE. 2017;12(3), e0173579. https://doi.org/10.1371/journal.pone.0173579
12. Megreya AM, White D, Burton AM. The other-race effect does not rely on memory: Evidence from a matching task. Q J Exp Psychol (Hove). 2011;64(8):1473–1483. https://doi.org/10.1080/17470218.2011.575228
13. Papesh MH, Goldinger SD. Deficits in other-race face recognition: No evidence for encoding-based effects. Can J Exp Psychol. 2009;63(4):253–262. https://doi.org/10.1037/a0015802
14. Stelter M, Rommel M, Degner J. (Eye-) tracking the other-race effect: Comparison of eye movements during encoding and recognition of ingroup faces with proximal and distant outgroup faces. Soc Cogn. 2021;39(3):366–395. https://doi.org/10.1521/soco.2021.39.3.366
15. Bruce V. Changing faces: Visual and non-visual coding processes in face recognition. Br J Psychol. 1982;73(1):105–116. https://doi.org/10.1111/j.2044-8295.1982.tb01795.x
16. Wells GL, Rydell SM, Seelau EP. The selection of distractors for eyewitness lineups. J Appl Psychol. 1993; 78(5):835–844. https://doi.org/10.1037/0021-9010.78.5.835
17. Fitzgerald R, Tredoux CG, Juncu S. Estimation of eyewitness identification error rates in fair and biased lineups. Paper presented at: American Psychology and Law Society Annual Meeting; 2022 Mar 16–20; Denver, CO, USA.
18. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48. https://doi.org/10.18637/jss.v067.i01
19. R Core Team. R: A language and environment for statistical computing [software on the Internet]. Vienna: R project; 2018. Available from: https://www.R-project.org /
20. Westfall J, Kenny DA, Judd CM. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. J Exp Psychol Gen. 2014;143(5):2020–2045. https://doi.org/10.1037/xge0000014
21. Wittwer T. The own group bias in face processing: The effect of training on recognition performance [dissertation]. Cape Town: University of Cape Town; 2020.
22. Wright DB, Boyd CE, Tredoux CG. Inter-racial contact and the own-race bias for face recognition in South Africa and England. Appl Cogn Psychol. 2003;17(3):365–373. https://doi.org/10.1002/acp.898
23. Malpass RS. An excursion into utilitarian analysis. Behav Sci Res. 1990;24(1–4):1–15. https://doi.org/10.1177/106939719002400101
24. Valentine T. Face-space models of face recognition. In: Wenger MJ, Townsend JT, editors. Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. New York: Lawrence Erlbaum Associates Publishers; 2001. p. 83–113.
25. Valentine T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. Q J Exp Psychol A. 1991;43(2):161–204. https://doi.org/10.1080/14640749108400966
26. Natu V, O'Toole AJ. Neural perspectives on the other-race effect. Vis Cogn. 2013;21(9–10):1081–1095. https://doi.org/10.1080/13506285.2013.811455
27. Herzmann G, Willenbockel V, Tanaka JW, Curran T. The neural correlates of memory encoding and recognition for own-race and other-race faces. Neuropsychologia. 2011;49(11):3103–3115. https://doi.org/10.1016/j.neuropsychologia.2011.07.019
28. Stahl J, Wiese H, Schweinberger SR. Learning task affects ERP-correlates of the own-race bias, but not recognition memory performance. Neuropsychologia. 2010;48(7):2027–2040. https://doi.org/10.1016/j.neuropsychologia.2010.03.024
29. Halberstadt AG, Cooke AN, Garner PW, Hughes SA, Oertwig D, Neupert SD. Racialized emotion recognition accuracy and anger bias of children's faces. Emotion. 2022;22(3):403–417. https://doi.org/10.1037/emo0000756
30. Dehon H, Brédart S. An 'other-race' effect in age estimation from faces. Perception. 2001;30(9):1107–1113. https://doi.org/10.1068/p3122
31. Davis D, Loftus EF. Eyewitness science in the 21st century. In Wixted J, editor. Stevens' handbook of experimental psychology and cognitive neuroscience. 4th ed. Hoboken, NJ: John Wiley & Sons; 2018. p. 1–38. https://doi.org/10.1002/9781119170174.epcn116
32. Hills PJ, Lewis MB. Reducing the own-race bias in face recognition by shifting attention. Q J Exp Psychol (Hove). 2006;59(6):996–1002. https://doi.org/10.1080/17470210600654750
33. Wittwer T, Tredoux CG, Py J, Paubel PV. Training participants to focus on critical facial features does not decrease own-group bias. Front Psychol. 2019;10:2081. https://doi.org/10.3389/fpsyg.2019.02081
34. Fisher RP, Geiselman RE. Memory enhancing techniques for investigative interviewing: The cognitive interview. Springfield, IL: Charles C Thomas Publisher; 1992.
35. Demarchi S, Py J. A method to enhance person description: A field study. In: Bull R, Valentine T, Williamson T, editors. Handbook of psychology of investigative interviewing: Current developments and future directions. Chichester: Wiley-Blackwell; 2009. p. 241–256. https://doi.org/10.1002/9780470747599.ch14
36. Memon A, Meissner CA, Fraser J. The cognitive interview: A meta-analytic review and study space analysis of the past 25 years. Psychol Public Policy Law. 2010;16(4):340–372. https://doi.org/10.1037/a0020518