frontiers | Research Topics

Methodological issues in consciousness research, volume II

Edited by
Antonino Raffone, Morten Overgaard, Axel Cleeremans
and Luca Simione

Published in
Frontiers in Psychology

ISSN 1664-8714
ISBN 978-2-8325-6243-7
DOI 10.3389/978-2-8325-6243-7

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Methodological issues in consciousness research, volume II

Topic editors
Antonino Raffone — Sapienza University of Rome, Italy
Morten Overgaard — Aarhus University, Denmark
Axel Cleeremans — Université libre de Bruxelles, Belgium
Luca Simione — UNINT - Università degli studi Internazionali di Roma, Italy

Citation
Raffone, A., Overgaard, M., Cleeremans, A., Simione, L., eds. (2025). *Methodological issues in consciousness research, volume II*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-6243-7

Table of contents

04 Editorial: Methodological issues in consciousness research, volume II
Luca Simione, Antonino Raffone, Morten Overgaard and Axel Cleermans

07 Does report modality modulate psychophysical sensitivity? The jury remains out
Oliver J. Hulme, Barrie Roulston and Morten Overgaard

16 The heavy-tailed valence hypothesis: the human capacity for vast variation in pleasure/pain and how to test it
Andrés Gómez-Emilsson and Chris Percy

33 When philosophical nuance matters: safeguarding consciousness research from restrictive assumptions
Marius Usher, Niccolò Negro, Hilla Jacobson and Naotsugu Tsuchiya

48 Study protocol: Cerebral characterization of sensory gating in disconnected dreaming states during propofol anesthesia using fMRI
Benedetta Cecconi, Javier Montupil, Sepehr Mortaheb, Rajanikant Panda, Robert D. Sanders, Christophe Phillips, Naji Alnagger, Emma Remacle, Aline Defresne, Melanie Boly, Mohamed Ali Bahri, Laurent Lamalle, Steven Laureys, Olivia Gosseries, Vincent Bonhomme and Jitka Annen

62 Approaching the nature of consciousness through a phenomenal analysis of early vision. What is the explanandum?
Bruno Forti

75 Validation of the Awareness Atlas—a new measure of the manifestation of consciousness
Yuane Jia, Margaret Schenkman, Hester O Connor, Krishnmurthy Jayanna, Rosalind Pearmain, Annelies Van't Westeinde and Kamlesh D. Patel

90 The hidden structure of consciousness
Bruno Forti

104 Development of emergent processes and threshold of consciousness with levels of processing
Ryoichi Watanabe and Yusuke Moriguchi

120 A novel method for objectively classifying sequential emotion within dreams: a proof-of-concept pilot study
Mariza van Wyk, Mark Solms and Gosia Lipinska

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 02 April 2025
DOI 10.3389/fpsyg.2025.1585426

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Christopher Gutland,
Zhejiang University, China

*CORRESPONDENCE
Luca Simione
✉ luca.simione@unint.eu

RECEIVED 28 February 2025
ACCEPTED 24 March 2025
PUBLISHED 02 April 2025

CITATION
Simione L, Raffone A, Overgaard M and Cleermans A (2025) Editorial: Methodological issues in consciousness research, volume II. *Front. Psychol.* 16:1585426.
doi: 10.3389/fpsyg.2025.1585426

Editorial: Methodological issues in consciousness research, volume II

Luca Simione[1,2]*, Antonino Raffone[3], Morten Overgaard[4] and Axel Cleermans[5]

[1]Dipartimento di Scienze Umanistiche e Sociali Internazionali, Università degli Studi Internazionali (UNINT), Rome, Italy, [2]Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche (CNR), Rome, Italy, [3]Department of Psychology, Sapienza University of Rome, Rome, Italy, [4]Center of Functionally Integrative Neuroscience (CFIN), Department of Clinical Medicine, Aarhus University, Aarhus, Denmark, [5]Center for Research in Cognition and Neuroscience (CRCN), ULB Neuroscience Institute (UNI), Université Libre de Bruxelles, Brussels, Belgium

KEYWORDS

awareness, consciousness, methodological challenge, metacognition, perceptual awareness

Editorial on the Research Topic
Methodological issues in consciousness research, volume II

The study of consciousness spans a vast array of domains, including perceptual awareness, cognition and metacognition, executive control, selfhood, sleep and dreaming, emotional competence, and empathy. It concerns both healthy states (e.g., meditation, aging, spiritual experiences) and pathological conditions (e.g., epilepsy, neglect syndromes, locked-in syndrome, minimally conscious states, anesthesia). Despite decades of interdisciplinary research, the fundamental nature and mechanisms of consciousness remain elusive (Dehaene, 2017; Seth and Bayne, 2022). Several key theoretical distinctions continue to fuel debate. For instance, the differentiation between phenomenal and access consciousness (Block, 1995), the pre-reflective (minimal) and the reflective (narrative) self (Gallagher and Zahavi, 2008), or between graded and all-or-none processing (Overgaard and Sandberg, 2021) remain unsettled. Similarly, methodological controversies persist: how to best measure awareness, how to establish its absence, and how to isolate the neural correlates of consciousness (Mashour et al., 2020). Recent debates also highlight the limitations of current paradigms in distinguishing necessary from incidental neural correlates (Koch et al., 2016). This Research Topic gathered recent contributions that address these theoretical and methodological challenges from diverse perspectives.

A core challenge in consciousness research is developing reliable measures that capture different levels and manifestations of awareness. Jia et al. introduce the Awareness Atlas, a novel self-report scale aimed at assessing what they called the "manifestations of consciousness", i.e., behavioral, cognitive, and affective effects of different levels of awareness. This approach emphasizes the practical implications of measuring consciousness beyond its theoretical construct, particularly in areas like meditation research and clinical interventions. Watanabe and Moriguchi contribute to the long-standing debate on graded vs. all-or-none consciousness. By applying the Perceptual Awareness Scale (PAS) to an online discrimination task in children and adults, they provide evidence supporting a gradual emergence of conscious content. Notably, their findings suggest that while age does not significantly alter the emergence of subjective awareness, the gap between subjective experience and objective discrimination narrows over time.

Hulme et al. investigated whether report modality influences psychophysical sensitivity, a crucial issue in consciousness research that has received limited attention. Their study examines different report modalities in a perceptual discrimination task and reanalyzes previous data (Overgaard and Sørensen, 2004) to determine whether changes in report format affect perception itself. While their findings remain inconclusive, we advocated (together with the authors) for the necessity of further research into the relationship between report modality and conscious perception. Understanding if and how report modality interacts with perception mechanisms can deepen our general understanding of the perceptual conscious experience, allowing for a deeper (re-)evaluation of the widely accepted paradigm of the sensorimotor arc in favor of an alternative models, such as the one supported by report-dependent perceptual phenomena, in which different types of report manifest perceptual consequences.

Understanding how consciousness fluctuates across different states—such as wakefulness, sleep, and anesthesia—provides crucial insights into its mechanisms. Cecconi et al. propose a novel fMRI protocol to investigate sensory gating during disconnected dreaming states under propofol anesthesia. By combining neuroimaging with serial awakenings, their study offers a promising approach to identifying neural markers of disconnected vs. connected consciousness, with potential applications in anesthesia and disorders of consciousness.

The intersection of consciousness and emotion remains a crucial but overlooked area of the human experience. To fill this gap, van Wyk et al. employ sentiment analysis techniques—a branch of Natural Language Processing—to study emotional fluctuations in dreams. This methodological innovation provides an objective way to track the interplay between cognitive and emotional elements in dreaming, moving beyond traditional self-report approaches. With this new approach, they demonstrated how the emotional tone of dream content exhibits peaks and troughs across different dream segments. Instead, exploring the intensity of emotions and how it could shape our experience, Gómez-Emilsson and Percy pointed out the importance to take into consideration more seriously in research the incredible range of highs and lows which characterized such emotional experience. They challenge conventional models of emotional experience with their Heavy-Tailed Valence (HTV) Hypothesis. Contrary to standard models that assume a constrained valence range, their research suggests that the most intense emotional experiences (both pleasurable and painful) are orders of magnitude more extreme than previously assumed. This perspective has broad implications for research on wellbeing, self-reported happiness, and affective neuroscience.

Theoretical models of consciousness often hinge on specific philosophical assumptions, which can shape empirical research in subtle but profound ways. Usher et al. critique the Unfolding Argument (UA) against causal structure theories, arguing that it imposes restrictive constraints that may hinder scientific progress. They advocate for a phenomenology-centered approach to consciousness studies, emphasizing the primacy of subjective experience in grounding empirical investigations. Put phenomenology at the center of the scientific exploration of consciousness is not only important, but necessary for the authors, and we agree with them in this consideration on the limits of methods in consciousness studies which fail in linking human experience with neurophysiological and behavioral data. Consistent with this claim, Forti (a, b) contributes to the ongoing theoretical discourse on consciousness with two insightful papers that examine the structure of conscious experience from a phenomenological perspective. In the first paper [Forti (a)], he argues that the hierarchy of spatial belongings underlies the cohesive perception of early vision, proposing that conscious experience is organized in a way that mirrors brain structures. This hierarchical framework challenges reductionist accounts of perception and suggests that consciousness is best understood through the intrinsic relationships within experience itself. Similarly, in his second paper, Forti (b) critiques the traditional focus on qualia and the subjective "what it is like" focus of consciousness studies. He advocates shifting the explanatory target toward the structural and relational properties of phenomenal experience, particularly in early visual processing. By doing so, he provides a fresh perspective on the long-standing debate between higher-order and first-order theories of consciousness.

Conclusion

These contributions align with the broader methodological challenges discussed in this editorial, particularly the difficulties in operationalizing and measuring consciousness beyond subjective reports. The studies collected in this Research Topic highlight the need for refining conceptual models that account for the intrinsic organization of conscious perception, reinforcing the idea that empirical research on awareness must be complemented by rigorous phenomenological analyses. Moreover, they reflect the diverse and interdisciplinary nature of contemporary consciousness research. From novel measurement tools and state-dependent investigations to emotion-consciousness interactions and theoretical refinements, these contributions advance our understanding of one of the most complex scientific challenges. While fundamental questions remain unresolved, these works illustrate the ongoing evolution of methodologies and theoretical perspectives necessary to tackle the enigma of consciousness. Future research will benefit from continued interdisciplinary dialogue, integrating insights from neuroscience, psychology, philosophy, and computational modeling to refine our grasp of awareness and its mechanisms.

Author contributions

LS: Writing – original draft, Writing – review & editing. AR: Writing – original draft, Writing – review & editing. MO: Writing – original draft, Writing – review & editing. AC: Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188

Dehaene, S. (2017). *Consciousness and the Brain: Deciphering How the Brain Codes our Thoughts*. New York City: Penguin Random House.

Gallagher, S., and Zahavi, D. (2008). *The Phenomenological Mind*. Oxfordshire: Routledge.

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22

Mashour, G. A., Roelfsema, P. R., Changeux, J. P., and Dehaene, S. (2020). The neural correlates of consciousness: progress and future directions. *Sci. Adv.* 6:eaba8034. doi: 10.1016/j.neuron.2020.01.026

Overgaard, M., and Sandberg, K. (2021). The Perceptual Awareness Scale - recent debate. *Neurosci Conscious*. 2021:niab044. doi: 10.1093/nc/niab044

Overgaard, M., and Sørensen, T. A. (2004). Introspection distinct from first order experiences. *J. Conscious. Stud*. 11, 77–95.

Seth, A., and Bayne, T. (2022). Theories of consciousness. *Nat. Rev. Neurosci.* 23, 297–311. doi: 10.1038/s41583-022-00587-4

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 19 October 2023
DOI 10.3389/fpsyg.2023.1226588

Check for updates

OPEN ACCESS

EDITED BY
Valerio Santangelo,
University of Perugia, Italy

REVIEWED BY
Michał Wierzchoń,
Jagiellonian University, Poland
Renzo Lanfranco,
Karolinska Institutet (KI), Sweden

*CORRESPONDENCE
Morten Overgaard
✉ Morten.storm.overgaard@cfin.au.dk

RECEIVED 21 May 2023
ACCEPTED 18 September 2023
PUBLISHED 19 October 2023

CITATION
Hulme OJ, Roulston B and Overgaard M (2023) Does report modality modulate psychophysical sensitivity? The jury remains out. *Front. Psychol.* 14:1226588.
doi: 10.3389/fpsyg.2023.1226588

Does report modality modulate psychophysical sensitivity? The jury remains out

Oliver J. Hulme[1,2,3], Barrie Roulston[4] and Morten Overgaard[5]*

[1]Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Amager and Hvidovre, Copenhagen, Denmark, [2]London Mathematical Laboratory, London, United Kingdom, [3]Department of Psychology, University of Copenhagen, Copenhagen, Denmark, [4]Independent Researcher, London, United Kingdom, [5]CNRU, Center for Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark

Scientific studies of perception use motoric reports as the principal means of communicating subjective experience. In such experiments, a widely held and implicit assumption is that the motor action conveys but does not tamper with perceptual experience. We tested nine observers on a luminance detection task in a cross-over repeated measures design. In separate conditions, observers reported their detection via movements of either their hands or eyes. We found only anecdotal evidence for any modality-dependent effect on psychophysical sensitivity. We also reanalyzed an existing dataset from which deployed a similar detection paradigm involving hand and eye reports. In the four paradigm variants tested, we again only found anecdotal evidence for the effect of report modality on psychophysical sensitivity. Both studies reported here provide only anecdotal evidence; thus, whether we can replicate report-dependent perceptual effects still needs to be resolved. We argue why this remains an important question for consciousness research and why it deserves more rigorous and high-powered replication attempts.

KEYWORDS

report, conscious, modality, action, methodology

Introduction

In the cognitive and perceptual sciences, it is common to instruct observers to disclose motoric reports as an index of their subjective experience (e.g., Friston et al., 1995; Leopold and Logothetis, 1996; Genovese et al., 2002; Lamme, 2003; Hulme et al., 2009; Overgaard and Grünbaum, 2011; Overgaard and Sandberg, 2012; Tsuchiya et al., 2015; Andersen et al., 2016; Skewes et al., 2021; Andersen et al., 2022). A widely held and implicit assumption is that this sensorimotor arc begins with stimulus input, unfolds as a perception, and results in a decision that culminates in a motoric report (Figure 1A). Here, the report is the final stage conveying the semantic information pre-specified by the task. In such stage models of cognition and consciousness (Overgaard and Mogensen, 2017), one can compare results from experiments that have used different report modalities since the means of the report should not influence the earlier perceptual stages (Figure 1B). Under this model class, the assumption of report-modality invariance follows intuitively from the fact that reports are "temporally and logically posterior to the perceptions they describe" (Marcel, 1993).[1] In conflict with this view, Marcel (1993) reported that observers' psychophysical

1 The original citation year is 1993, the online version of the same book chapter is dated 2007.

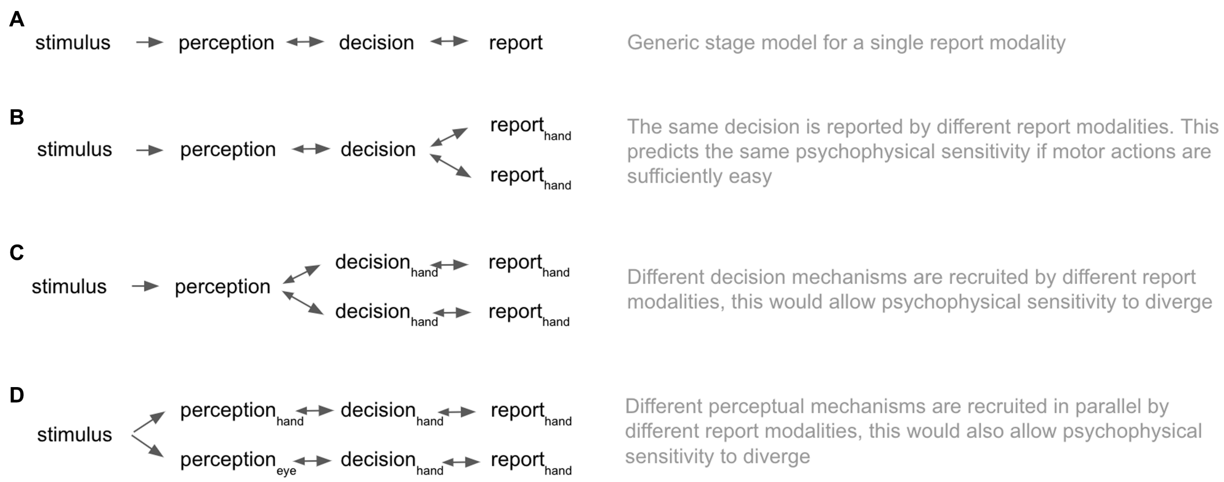

FIGURE 1
Stage models of conscious perception. **(A)** Illustrates a generic stage model for a simple detection task with a single report modality. **(B)** Illustrates a stage model with two report modalities, which report on the content of a unified decision mechanism. **(C)** Illustrates a stage model in which separate decision processes and downstream reports form in parallel. **(D)** Illustrates a stage model in which the perceptual, decision and report processes are parallel.

sensitivity in a simple speeded detection task varied according to the report modality. Observers performed a visual detection task, simultaneously reporting via three different modalities. Eye blinks were the most sensitive (mean d' 0.91), followed by finger-presses (mean d' 0.85), and then verbal reports (mean d' 0.18).[2] In a second condition where subjects were instructed to guess, Marcel reports the same ordering in the sensitivity of report modalities (eye-blink mean d' 2.07; button-press mean d' 1.76; verbal mean d' 1.52). These results would be predicted under either a model in which different report modalities recruit different decision-making pathways, which could be subject to different noise levels (Figure 1C) or a model in which perception itself is somehow contingent on the report modality (Figure 1D). Despite these original findings being theoretically intriguing, it is important to note that the author reported no inferential statistical tests in the original paper. Thus, the interpretation rests entirely on differences between descriptive statistics. Based on the original report, we could not recover the information necessary to calculate inferential statistics.

Report-dependent perceptual phenomena (RDPP) such as that claimed in Marcel (1993) echo earlier clinical studies showing that patients may have access to different levels of sensory information depending on which response modality they use. For instance, patients with visual extinction after brain damage to the right hemisphere were significantly worse at detecting contralesional stimulus when using a keypress report than when using a verbal report (Bisiach et al., 1989). Joanette et al. (1986) report how the subjective report of visual stimuli in hemispatial neglect patients depends on the hand used to report. All patients in their study reported more stimuli when reporting with the left hand (controlled by the contralesional hemisphere) than the right. In contrast, psychophysical sensitivity was the same for both hands. Related findings have been reported in a patient with a focal anterior cingulate cortex lesion whose performance on Stroop and divided attention tasks depended on the response modality used (Swick and Turken, 1999). The patient exhibited impairment in manual responses but not vocal responses under the same task requirements. Beyond neuropsychology, Gomi et al. (2013) showed that visual motion coding responds differently according to particular downstream motor outputs, with hand reports being less sensitive to the central occlusion of visual motion than eye reports. In all these cases, motor outputs modulate visual thresholds. Further to this neurobiological plausibility, a large body of theoretical work on action-perception loops (Gibson, 1979; Clark, 2015) conjectures a dynamic and bidirectional relationship between perception and action as fundamental to cognition. Such models would naturally predict a strong dependence between the action networks that are deployed to report a percept and the perceptual networks responsible for the perception. Closely related to this is the phenomenon of action-specific perception, whereby the action for which a perception is needed has a qualitative effect on the perceptual content (Witt, 2011). A commonly reported example is that softball players who are hitting better see the ball as bigger (Witt and Proffitt, 2005). As mentioned above, the possibility that reporting itself may confound the neural correlates of consciousness in experiments – or even alter the experience itself – has been expressed. In previous research (Sandberg and Overgaard, 2015; Overgaard and Sandberg, 2021), we have argued that the reporting method may influence results in the sense that people will act differently when presented with a dichotomous scale, a 100-point scale or a scale with defined scale points such as PAS. However, the experiments mentioned above indicate that this is more than a methodological concern. They suggest that there might be perceptual consequences to different types of report. In recent years, no-report paradigms have appeared as a reaction against some of the methodological challenges related to reporting, i.e., paradigms

2 Note that Marcel (1993) did not explicitly compute d', however these were retrospectively calculated and reported in Overgaard and Sørensen (2004). Because the data cannot be fully reconstructed from the original paper, the statistical tests on the differences in d' cannot be computed.

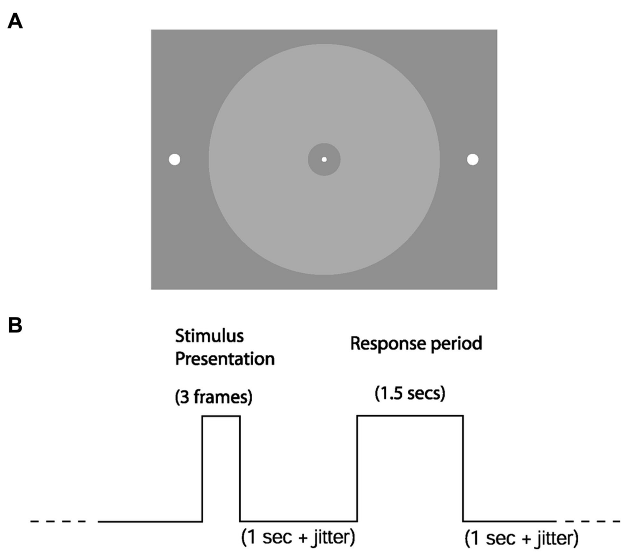

FIGURE 2
Stimuli and trial structure. **(A)** Observers were required to detect the luminance-defined disc in their parafovea. In the eye-reporting condition, they would do so by moving their eyes from the central fixation dot to either of the lateralized target dots (left to report "present," right to report "absent"). **(B)** Shows the temporal structure of a single trial. The three frames for the stimulus presentation take 48 ms.

where some objective behavior is measured instead of using a report (Tsuchiya et al., 2015). No-report paradigms have been recognized as a methodologically important supplement to consciousness research but have been challenged as a "stand-alone" approach. If one wishes to study subjective experience, it is unclear which objective measure to use. How can we know that correct identification or any other measure of performance is also a measure of consciousness? It seems the only knowledge we could have comes from previous experiments finding correlations between the behavior in question and introspective observation followed by a report. Thus, the behavioral measure cannot have any higher precision than the introspective observation/report, and it depends on i.

Despite several published examples of putative RDPP, it is far from established as a known phenomenon. Given the volume of experimental data on visual psychophysics, we might expect that if RDPP does exist, researchers should report it more commonly. However, as the stage model implicitly underlies most consciousness research, it is rare that more than one motoric modality is tested within the same experiment, and even rarer that researchers directly compare report modalities. Considering the potential impact of this debate on current models of perception, it is notable (publication bias notwithstanding) that the field needs to direct effort toward replicating or generalizing RDPP. Were the findings obtained by Marcel (1993) to be verified and generalized, consciousness research would face two problems: (1) accumulated knowledge must account for the fact that evidence collected using different report methods can no longer be directly compared, and (2) the assumption that that perception is prior to and independent of the report must be reconsidered.

Toward this end, we do two things in this paper. First, we present data from a repeated measures experiment designed to test for the putative existence of RDPP. We measure psychophysical sensitivity in a visual detection task, comparing a condition in which subjects report with their hands versus a second condition in which they report with their eyes via left or right saccades. Second, we reanalyze data from Overgaard and Sørensen (2004), which measured psychophysical sensitivity in four different experiments, each comparing hand reports against eye reports. We perform Bayesian statistics for both datasets to compute evidence levels for or against effects.

Methods

Observers

Twelve healthy observers (4 women; mean age 23, range ± 0.6 years) with normal or corrected to normal vision participated in the study after giving written consent. Data from three observers were discarded due to technical problems with eye-tracking, leaving nine observers for whom we can compare data from the two conditions. The National Hospital for Neurology and Neurosurgery Ethics Committee, London, UK, granted ethical approval for the study. All observers gave informed consent as per the declaration of Helsinki.

Stimuli and task

We made all stimuli using COGENT 2000 Graphics[3] running in MATLAB.[4] The stimuli were presented centrally and projected onto the screen using an LCD projector (60 Hz refresh rate). A continuous trace of horizontal eye position, vertical eye position, and pupil areas was recorded at 120 Hz using an ASL 5000 eye tracker. Observers performed two psychophysical testing sessions, one out-of-scan followed by an in-scan session, otherwise identical. We analyzed both sessions together in this manuscript. In each session, observers performed a simple detection task to obtain their psychometric performance as a function of stimulus luminance for both the saccade (eye-reporting) and keypress (hand-reporting) conditions. Each report modality was performed in separate blocks of 220 trials using the same paradigm. We permuted an equal number of (target) present and absent trials within each session. Two white dots (radius: 0.25° visual angle) were present throughout the trial on the left and right periphery. The appearance of a central white fixation point on an achromatic background indicated the start of the trial (see Figure 2 for stimulus configuration). After 800 ms plus random jitter (on the interval of 0–1 s, uniform probability distribution), either a circular achromatic ring of variable luminance (radius of ~6.2°, with an inner circular gap of ~0.3°) appeared for 48 ms ("present" trial) or did not appear at all ("absent" trial). The stimulus's exact luminance was unknown because we could not use a photometer sufficiently near the high magnetic strength of the scanner for safety reasons. 1 s after stimulus presentation (plus jitter 0.0–0.5 s, uniform

3 www.vislab.ucl.ac.uk
4 www.mathworks.com

probability distribution), the fixation dot disappeared, and observers had a 1.5 s period in which to give their response ("present" or "absent"). Following that was the intertrial interval of 1 s (plus jitter on the interval of 0–0.5 s, uniform probability distribution). The keypress condition consisted of a left-hand or right-hand button press on a keypad with the index fingers of the left and right hands, respectively. For the saccade condition, observers moved their eyes to the left or right peripheral target dots before returning to the fixation dot (Figure 2A). The side representing each response ("present" or "absent") was fixed within observers but randomly counter-balanced across observers. After fitting the psychometric function for both report modalities of each subject (separately for the in-scanner and out-of-scanner sessions), we calculated the threshold luminance value (75% detection accuracy) of the manual keypress condition.

Data types reported

Noted there is functional neuroimaging data acquired during the in-scanner sessions. We do not report this data due to this data not being adequately analyzed when first collected and not being adequately archived to allow the data to be salvaged and reanalyzed. The primary aim of this paper rests entirely on the behavioral data acquired.

Results

Figure 3 shows psychophysical functions from 4 representative observers. As seen from the overlapping fits, in these four observers, the fitted psychophysical functions appear effectively the same for the two report modalities. Table 1 contains the hit rates, false alarm rates, and other signal detection measures for each participant. It can be noted that the hit rate for all participants was relatively high, approximately 90%, with a low false alarm rate below 0.01 for all participants except number 7. We do not know why this participant had a higher false alarm rate than the others. We performed a Bayesian paired *t*-test testing the null hypothesis (H0) of no difference in d' against an alternate hypothesis (H1) of there being a difference in either direction, with default Cauchy priors with a scale parameter of 0.707 (12). We found anecdotal evidence in favor of the null hypothesis (BF10 = 0.466, BF01 = 2.144), with a median effect size of −0.246 and 95% Bayesian credibility interval (BCI95) of [−0.870, 0.330] (Figure 4 upper panel). Statisticians typically describe this level of evidence as "barely worth mentioning." The evidence levels proved robust to variations in the width of the priors. However, we note that ultrawide priors resulted in a moderate evidence level in favor of the null hypothesis (BF10 = 0.2789, BF01 = 3.586) (Figure 4, middle). The evolution of evidence can be seen in Figure 4 (lower) as the sample of observers increases from 1 to 9 (Table 2).

We explored bias as another possible difference in psychophysical performance that report modality could influence. To test for differences in perceptual bias, we performed a Bayesian paired *t*-test testing the null hypothesis of no difference in criterion value against an alternate hypothesis of a difference of either sign, again with default Cauchy priors with a scale parameter of 0.707. We found anecdotal evidence in favor of the null hypothesis (BF10 = 0.737, BF01 = 1.358), with a median effect size of −0.385 and Bayesian credibility intervals (BCI) of [−1.052, 0.215] (Figure 5, upper). This level of evidence proved to be robust to variations in the width of the priors (Figure 5, middle). Again, the evolution of evidence as observers increase from 1 to 9 is shown in Figure 5, lower.

As mentioned in the introduction, there has (to our knowledge) been only one previous replication attempt of the Marcel (1993) paper (Overgaard and Sørensen, 2004). This study did not report the group statistics relevant to our question. We have taken the opportunity to perform a simple reanalysis of this data. Table 3 displays the d-prime values calculated by Overgaard and Sørensen, 2004, who found a d' value of 0.91 for eye blinks, 0.85 for button press and 0.18 for verbal response. The only difference between experiments 3 and 4 is the wait time between the stimulus and the report to account for a memory decay effect. Within each experiment, there is a comparison between a pre-cue, where they cued which report modality to use ahead of the stimulus, and a post-cue, where they cued the report modality after the stimulus. We performed the equivalent Bayesian paired two-sided *t*-tests on all four paradigm variants. We obtained anecdotal evidence for all variants (Figures 6A–C), except for experiment 4 in the pre-cue condition, which revealed moderate evidence (Figure 6D, BF01 = 3.2) in favor of the null hypothesis of no difference in d-prime. Both datasets testing for differences in d-prime between eye and hand report conditions are effectively anecdotal or close to anecdotal.

Finally, to integrate the two datasets, we averaged the d-primes across the four paradigm variants, creating a subject-specific d-prime for hand and eye reports. We appended the data obtained in the experimental data first reported here to this averaged data. This analysis gave a group of 19 subjects. Again calculating the same *t*-test,

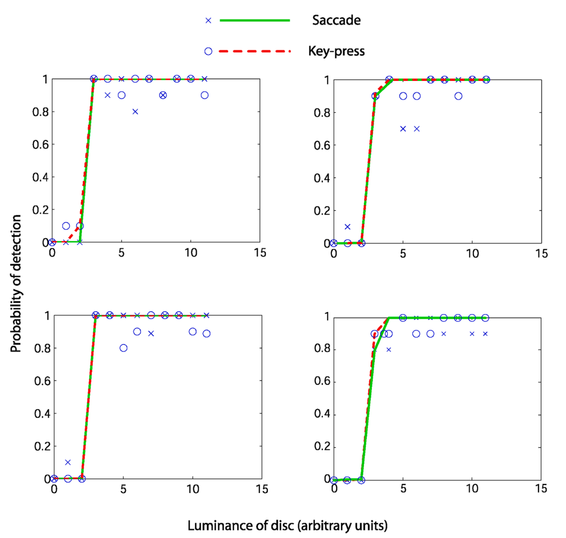

FIGURE 3
Psychometric functions from four representative observers. Crosses and circles represent real data from the saccade and keypress conditions. Solid lines represent the fitted cumulative Weibull function.

TABLE 1 Performance measures for both report modes during scanning sessions.

Subject number	Saccade				Key-press			
	Hit	False alarm	d'	c	Hit	False alarm	d'	c
1	0.991	0.009	4.720	0.002	0.991	0.009	4.740	0.000
2	0.933	0.009	3.860	0.434	0.938	0.009	3.910	0.415
3	0.990	0.018	4.440	−0.119	0.939	0.077	2.970	−0.060
4	0.846	0.011	3.310	0.635	0.938	0.009	3.900	0.413
5	0.944	0.070	3.060	−0.058	0.958	0.009	4.100	0.317
6	0.891	0.018	3.340	0.437	0.867	0.018	3.220	0.499
7	0.917	0.270	2.000	−0.387	0.878	0.160	2.160	−0.086
8	0.762	0.129	1.840	0.209	0.899	0.025	3.230	0.340
9	0.861	0.094	2.400	0.116	0.911	0.039	3.110	0.212

Hit indicates the hit rate, False indicates the false alarm rate, c indicates an estimate of the decision criteria, and d' is the estimate of psychophysical performance.

we find only anecdotal evidence for the alternate hypothesis (Figure 6E, BF01 = 2.934).

Bayes factor design analysis for future studies

In light of the inconclusive findings above, it is important to assess how much data would be required to adequately test for the effect of interest. We performed a Bayes factor design analysis (Stefan et al., 2019), as implemented by the Bayes factor design analysis R package (Schönbrodt and Stefan, 2019; R Core Team, 2021). We ran two simulations for two-sided Bayesian paired *t*-tests, one for data generated under the null model (H0 that there is no report modality effect) and data generated under the alternate model (H1 that there is an effect of report modality on psychophysical sensitivity). For the H0 simulation, the effect size was set to 0, with default Cauchy priors on the effect size (sqrt(2)/2), with a minimal number of subjects set at 20 and a maximal number of subjects set at 150. An equivalent simulation was run for H1, where the effect size was set to medium (Cohen's d = 0.5). For the H0 simulation, the average stopping point was 78, defined as the average number of subjects sampled before hitting a strong evidence threshold of BF = 10 for or against H0. In other words, assuming no effect exists, a strong evidence threshold is reached on average after 78 subjects. 76% of simulated studies correctly hit the strong evidence bound for H0. Only 2% incorrectly hit the strong evidence boundary for H1 (Figure 7A). For the H1 simulation, the average stopping point was 36, meaning that on average, if the true effect size was 0.5, then an average of 36 subjects would be sampled before hitting a strong evidence threshold. 100% of simulated studies correctly hit the strong evidence bound for H1 (Figure 7B). The estimated sample size for 90% of correct detection of H1 was 73 (Figure 7C). This provides some perspective on why the evidence is inconclusive for the small samples presented in the experiments above. Based on simulating different designs, we recommend an optional stopping design, with a minimum of 20 subjects and a maximum of 150 subjects, stopping data acquisition whenever a strong evidence threshold is reached. This policy provides a very large chance of correctly detecting a medium-sized effect if it exists (>99%) and a defensible chance of correctly detecting the absence of an effect (>75%).

Discussion

Summary of results

We present the results of two studies that asked if report modality impacts psychophysical sensitivity. Comparing hand and eye reports in similar ways, the results of both studies failed to show even moderate evidence for or against the hypothesis for such an effect. The results presented here remain largely inconclusive; as such, they offer no update to our credences for or against the existence of these report-dependent phenomena. We discuss several limitations of the experiments analyzed, and we end with recommendations for future experiments.

Sample size and Bayes factor design analysis

An obvious limitation is the small sample size that was obtained. Unfortunately, this is a dataset that was collected a long time ago. Otherwise, it would have been more pragmatic to increase the sample size, for instance, under an optional stopping design. Nonetheless, we performed a Bayes factor design analysis to estimate how many subjects would need to be sampled to reach a power of 90%, defined as a 90% chance of obtaining strong evidence for H1, given that H1 generated the data with an effect size of 0.5. Sample size estimation suggested 73 subjects would be required for this, and even more if we were to obtain a good chance of correctly inferring null effects. The sample size estimate is far from what was obtained in this study and may provide some perspective on why inconclusive results were observed. To test for report modality effects of this kind, we think it is important to be able to infer the null, and thus, we recommend an optional stopping design with an upper limit of at least 150 subjects. According to our simulations, this sampling policy would yield a 75% chance of correct inference on a true null.

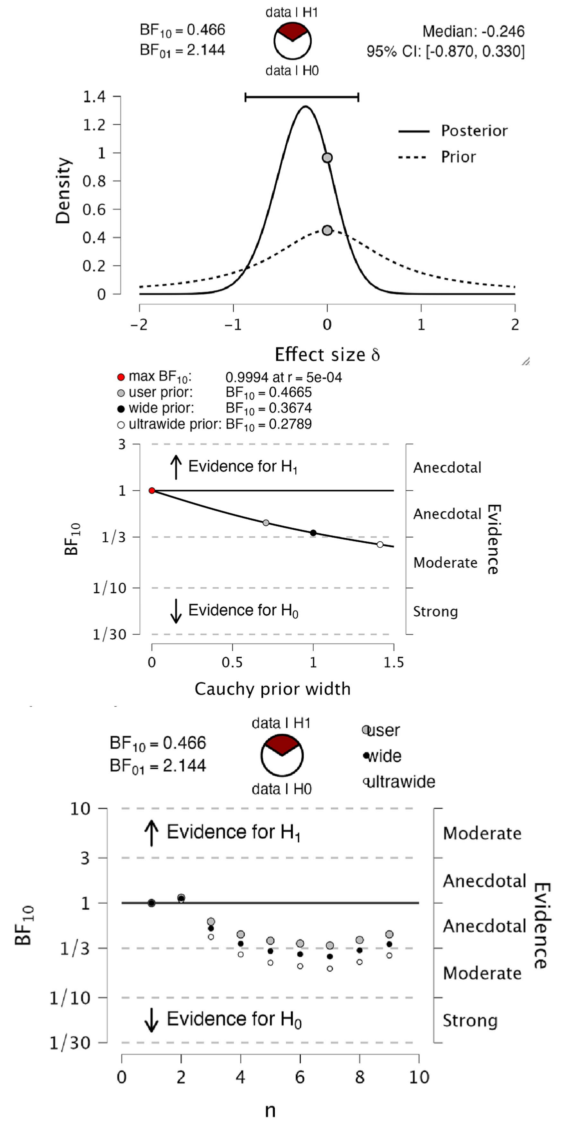

FIGURE 4
Bayesian paired *t*-test on d′. Upper panel, *t*-test information on prior and posterior values. The middle panel shows a Bayes factor robustness test for the same test: the lower panel sequential analysis of evidence levels over observer sample size.

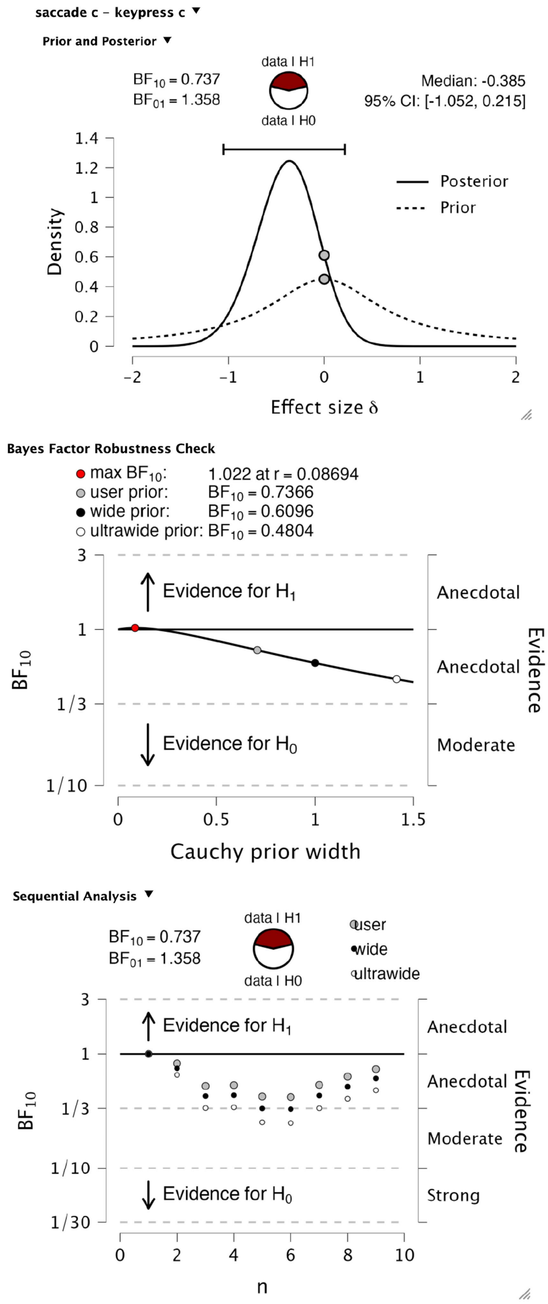

FIGURE 5
Bayesian *t*-test on criterion values. Upper panel, *t*-test information on prior and posterior values. The middle panel shows a Bayes factor robustness test for the same test. Lower panel, sequential analysis of evidence levels over observer recruitment.

TABLE 2 Summary statistics for both report modes during scanning sessions.

Descriptive statistics	Saccade d′	Saccade c	Keypress d′	Keypress c
Valid	9.000	9.000	9.000	9.000
Missing	0.000	0.000	0.000	0.000
Mean	3.219	0.141	3.482	0.228
Std. error of mean	0.339	0.107	0.253	0.074
Std. deviation	1.016	0.322	0.759	0.223
Variance	1.033	0.104	0.575	0.050
Minimum	1.840	−0.387	2.160	−0.086
Maximum	4.720	0.635	4.740	0.499

Ceiling effect

Another factor that may have contributed to inconclusive findings is that the hit rate of the main experiment was quite high for both the hand and eye report conditions (approximately 90%). While this leaves some room for performance improvements above 90% and more room for performance decreases downwards, we may have achieved greater sensitivity to any report effect if the task was harder.

TABLE 3 d-prime values reported in Overgaard and Sørensen (2004).

Participant#	hand_pre-cue_exp3	eye_pre-cue_exp3	delta-pre-cue_exp3	hand_post-cue_exp3	eye_post-cue_exp3	delta_post-cue_exp3	hand_pre-cue_exp4	eye_pre-cue_exp4	delta-pre-cue_exp4	hand_post-cue_exp4	eye_post-cue_exp4	delta_post-cue_exp4
1	2.850	2.600	−0.250	3.200	2.300	−0.900	0.600	2.350	1.75	2.6	2.7	0.1
2	1.900	0.000	−1.900	1.700	0.100	−1.600	2.600	0.000	−2.6	2.85	2.25	−0.6
3	0.900	2.600	1.700	3.200	2.800	−0.400	3.200	1.200	−2	2.15	2.55	0.4
4	1.650	2.450	0.800	2.450	2.200	−0.250	−0.600	2.500	3.1	1.5	1.75	0.25
5	0.150	−1.200	−1.350	0.700	0.300	−0.400	1.950	3.350	1.4	2.2	2	−0.2
6	3.000	3.000	0.000	2.450	2.800	0.350	0.500	2.650	2.15	3.2	2.65	−0.55
7	0.700	2.500	1.800	2.400	2.450	0.050	1.750	0.650	−1.1	2.2	2.35	0.15
8	−0.200	1.850	2.050	2.100	2.200	0.100	1.900	−0.100	−2	2.35	2.35	0
9	0.650	2.050	1.400	2.550	1.700	−0.850	1.800	2.400	0.6	2.35	2	−0.35
10	−1.950	1.400	3.350	2.250	1.700	−0.550	2.400	1.150	−1.25	2.4	2.2	−0.2

Experiments 3 and 4 resulted in d-prime values in pre- and post-cue report conditions. Eye blinks (eye) and button presses (hand).

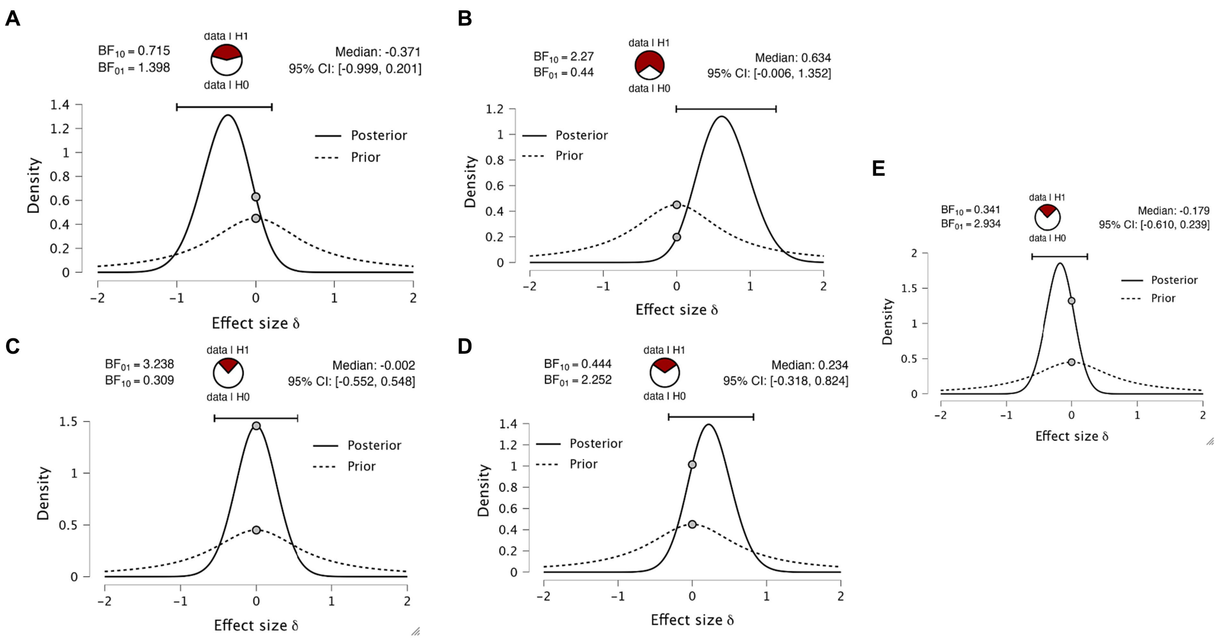

FIGURE 6
Bayesian paired *t*-tests on Overgaard and Sørensen (2004). Plots indicate the Bayes factors for **(A)** experiment 3 pre-cue, **(B)** experiment 3 post-cue, **(C)** experiment 4 pre-cue, **(D)** experiment 4 post and **(E)** data aggregated data averaging all four experiments and combining with the data obtained from the new dataset reported here.

Going forward, we recommend staircasing psychophysical performance to a lower level of 75% for one of the report modalities before performing the experiment.

Modeling of report-semantic mappings

Another minor limitation is that the mapping between present versus absent reports and the laterality of the action was counterbalanced across subjects. However, this was not modeled in the statistical analysis. Due to the provenance of the data, we do not have access to the counterbalancing information, so unfortunately, this could not be modeled. This should be included in any future testing.

Why this question remains important

Report modality is still not typically considered an important factor in the experimental design of perceptual studies. The potential demonstration of RDPP would elevate report modality as an important factor in designing perceptual experiments. This

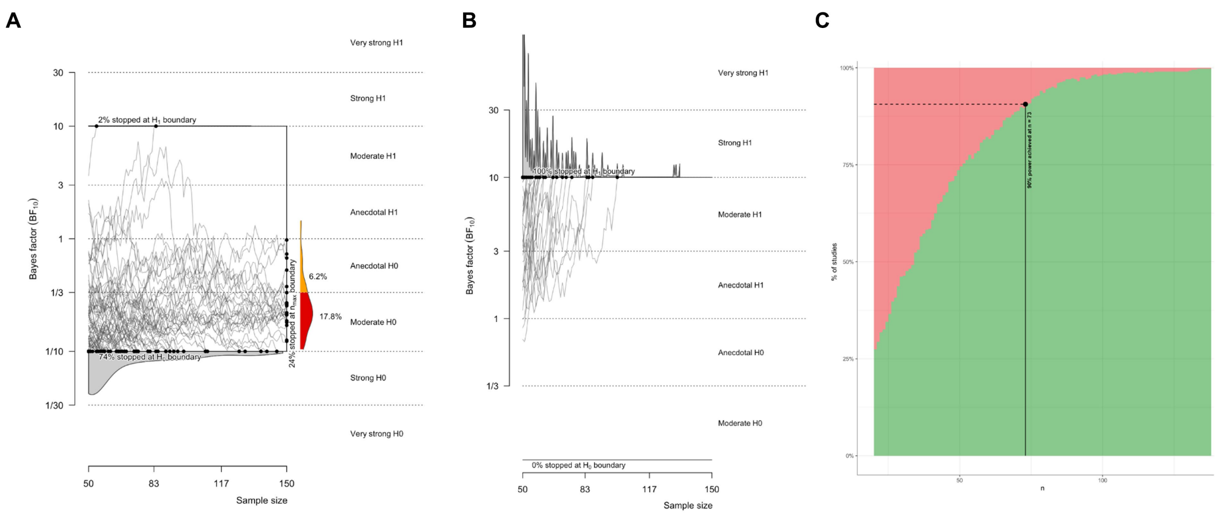

FIGURE 7
Bayes factor design analysis simulations for future studies. **(A)** Evidence trajectories for experiments where H0 is true. **(B)** Evidence trajectories for experiments where H1 is true. **(C)** Estimated sample size for a power of 90% for correctly detecting a medium effect size.

would challenge our current models of perception and provoke new research into the mechanisms underlying these effects. For this reason, we advocate for this experimental question to be empirically resolved.

Conclusion

The data presented here show no substantive evidence of whether report modalities influence sensory perception. Nevertheless, our attempt to answer this question exposes an overlooked question that remains necessary to answer.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the National Hospital for Neurology and Neurosurgery Ethics Committee, London, UK. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MO conceived of the idea and wrote the manuscript. OH conceived of the idea, ran the experiment, analyzed data, and wrote the manuscript. BR conceived of the idea, ran the experiment, and analyzed data. All authors contributed to the article and approved the submitted version.

Funding

A Novo Nordisk Foundation Exploratory Interdisciplinary Synergy Grant, ref. NNF20OC0064869, funded OH.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Andersen, L. M., Pedersen, M. N., Sandberg, K., and Overgaard, M. (2016). Occipital MEG activity in the early time range (<300 ms) predicts graded changes in perceptual consciousness. *Cereb. Cortex* 26, 2677–2688. doi: 10.1093/cercor/bhv108

Andersen, L. M., Vinding, M., Sandberg, K., and Overgaard, M. (2022). Task requirements affect the neural correlate of consciousness. *Eur. J. Neurosci.* 56, 5810–5822. doi: 10.1111/ejn.15820

Bisiach, E., Vallar, G., and Geminiani, G. (1989). Influence of response modality on perceptual awareness of contralesional visual stimuli. *Brain* 112, 1627–1636. doi: 10.1093/brain/112.6.1627

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., et al. (1995). Analysis of fMRI time-series revisited. *NeuroImage* 2, 45–53. doi: 10.1006/nimg.1995.1007

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15, 870–878. doi: 10.1006/nimg.2001.1037

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton, Mifflin and Company.

Gomi, H., Abekawa, N., and Shimojo, S. (2013). The hand sees visual periphery better than the eye: motor-dependent visual motion analyses. *J. Neurosci.* 33, 16502–16509. doi: 10.1523/JNEUROSCI.4741-12.2013

Hulme, O. J., Friston, K. F., and Zeki, S. (2009). Neural correlates of stimulus reportability. *J. Cogn. Neurosci.* 21, 1602–1610. doi: 10.1162/jocn.2009.21119

Joanette, Y., Brouchon, M., Gauthier, L., and Samson, M. (1986). Pointing with left vs right hand in left visual field neglect. *Neuropsychologia* 24, 391–396. doi: 10.1016/0028-3932(86)90024-2

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18. doi: 10.1016/S1364-6613(02)00013-X

Leopold, D. A., and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553. doi: 10.1038/379549a0

Marcel, A. J. (1993). "Slippage in the Unity of consciousness" in *Ciba foundation symposium 174- experimental and theoretical studies of consciousness* (Chichester, UK: John Wiley & Sons, Ltd.), 168–186.

Overgaard, M., and Grünbaum, T. (2011). Consciousness and modality: on the possible preserved visual consciousness in blindsight subjects. *Conscious. Cogn.* 20, 1855–1859. doi: 10.1016/j.concog.2011.08.016

Overgaard, M., and Mogensen, J. (2017). An integrative view on consciousness and introspection. *Rev. Philos. Psychol.* 8, 129–141. doi: 10.1007/s13164-016-0303-6

Overgaard, M., and Sandberg, K. (2012). Kinds of access: different methods for report reveal different kinds of metacognitive access. *Philos. Trans. Royal Society London Series B* 367, 1287–1296.

Overgaard, M., and Sandberg, K. (2021). The perceptual awareness scale – recent controversy and debate. *Neurosci. Conscious.* 2021, 1–8. doi: 10.1093/nc/niab044

Overgaard, M., and Sørensen, T. A. (2004). Introspection distinct from first-order experiences. *J. Conscious. Stud.* 11, 77–95.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/

Sandberg, K., and Overgaard, M. (2015). "Using the perceptual awareness scale (PAS)" in *Behavioural methods in consciousness research*. ed. M. Overgaard (Oxford: Oxford University Press)

Schönbrodt, F. D., and Stefan, A. M. (2019). BFDA: an R package for Bayes factor design analysis (version 0.5. 0).

Skewes, J., Frith, C. D., and Overgaard, M. (2021). Awareness and confidence in perceptual decision making. *Brain Multiphys.* 2:100030. doi: 10.1016/j.brain.2021.100030

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., and Wagenmakers, E. J. (2019). A tutorial on Bayes factor design analysis using an informed prior. *Behav. Res.* 51, 1042–1058. doi: 10.3758/s13428-018-01189-8

Swick, D., and Turken, A. U. (1999). Response selection in the human anterior cingulate cortex. *Nat. Neurosci.* 2, 920–924. doi: 10.1038/13224

Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. A. F. (2015). No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* 19, 757–770. doi: 10.1016/j.tics.2015.10.002

Witt, J. K. (2011). Action's effect on perception. *Curr. Dir. Psychol. Sci.* 20, 201–206. doi: 10.1177/0963721411408770

Witt, J., and Proffitt, D. (2005). See the ball, hit the ball: apparent ball size is correlated with batting average. *Psychol. Sci.* 16, 937–938. doi: 10.1111/j.1467-9280.2005.01640.x

TYPE Original Research
PUBLISHED 16 November 2023
DOI 10.3389/fpsyg.2023.1127221

OPEN ACCESS

EDITED BY
Antonino Raffone,
Sapienza University of Rome, Italy

REVIEWED BY
Parviz Azadfallah,
Tarbiat Modares University, Iran
Avinash Lingojirao Tandle,
SVKM's Narsee Monjee Institute of Management Studies, India

*CORRESPONDENCE
Andrés Gómez-Emilsson
algekalipso@gmail.com

RECEIVED 23 January 2023
ACCEPTED 19 October 2023
PUBLISHED 16 November 2023

CITATION
Gómez-Emilsson A and Percy C (2023) The heavy-tailed valence hypothesis: the human capacity for vast variation in pleasure/pain and how to test it.
Front. Psychol. 14:1127221.
doi: 10.3389/fpsyg.2023.1127221

The heavy-tailed valence hypothesis: the human capacity for vast variation in pleasure/pain and how to test it

Andrés Gómez-Emilsson[1]* and Chris Percy[2]

[1]Qualia Research Institute, San Francisco, CA, United States, [2]University of Derby, Derby, United Kingdom

Introduction: Wellbeing policy analysis is often criticized for requiring a cardinal interpretation of measurement scales, such as ranking happiness on an integer scale from 0-10. The commonly-used scales also implicitly constrain the human capacity for experience, typically that our most intense experiences can only be at most ten times more intense than our mildest experiences. This paper presents the alternative "heavy-tailed valence" (HTV) hypothesis: the notion that the accessible human capacity for emotional experiences of pleasure and pain spans a minimum of two orders of magnitude.

Methods: We specify five testable predictions of the HTV hypothesis. A pilot survey of adults aged 21-64 ($n = 97$) then tested two predictions, asking respondents to comment on the most painful and most pleasurable experiences they can recall, alongside the second most painful and pleasurable experiences.

Results: The results find tentative support for the hypothesis. For instance, over half of respondents said their most intense experiences were at least twice as intense as the second most intense, implying a wide capacity overall. Simulations further demonstrate that survey responses are more consistent with underlying heavy-tailed distributions of experience than a "constrained valence" psychology.

Discussion: A synthesis of these results with prior findings suggests a "kinked" scale, such that a wide range of felt experience is compressed in reports at the high end of intensity scales, even if reports at lower intensities behave more cardinally. We present a discussion of three stylized facts that support HTV and six against, lessons for a future survey, practical guidelines for existing analyses, and implications for current policy. We argue for a dramatic increase in societal ambition. Even in high average income countries, the HTV hypothesis suggests we remain far further below our wellbeing potential than a surface reading of the data might suggest.

KEYWORDS
wellbeing and happiness, pain, valence, emotion, psychophysics, philosophy

1. Introduction

> *"Am I not the same being who once enjoyed an excess of happiness, who at every step saw paradise open before him, and whose heart was ever expanded towards the whole world? And this heart is now dead; no sentiment can revive it. My eyes are dry; and my senses, no more refreshed by the influence of soft tears, wither and consume my brain."* Johann Wolfgang von Goethe, The Sorrows of Young Werther (1774)

The intensity of felt experience has long attracted attention, both academic and lay alike. The quote from Goethe above illustrates how the notion of extreme highs and lows of emotion has captured the imagination of novelists. Sufficiently so in this case that Goethe's depiction of depression and suicide reportedly led to copycat suicides and the decision to ban his book (Furedi, 2015).

In this paper, we propose a hypothesis that the human capacity for felt sensations and emotions encompasses an incredible range of highs and lows, focusing here on emotional experiences of pleasure and pain. We call this the "Heavy-Tailed Valence" (HTV) hypothesis, named after the early affective-circumplex model of emotions (Russell, 1980) and the feature of heavy-tailed distributions whereby more extreme experiences happen more frequently than casual observation suggests. The hypothesis holds that the most intense pleasures (or pains) are at least two orders of magnitude more intense to experience than the mildest – and that intense experiences are accessible at least in principle sufficiently often that there is policy relevance in considering them.

The HTV hypothesis contrasts against a "constrained valence" hypothesis, i.e., one in which the most intense experiences are no more than 10 times more intense than the mildest or in which any more intense experiences are so vanishingly rare that they can be discounted for practical purposes.

The constrained valence hypothesis is implicitly imposed in many policy interpretations of common measurement scales, such as the single order of magnitude spanned in an integer 11-point scale from 0 to 10. Such scales are common in wellbeing economics, wellbeing policy, and philanthropy (e.g., OECD, 2013; What Works Wellbeing, 2017; Plant, 2019, 2020), despite academic criticism and a call for ordinal-only interpretations (e.g., Bishop and Herron, 2015; Kero and Lee, 2016; Wodak, 2019). The debate on the adequacy of such scales remains contested as of 2023 (e.g., Larroulet Philippi, 2023; Samuelsson et al., 2023). However, the focus of debate is typically around comparability, linearity, and neutrality, rather than the implied human capacity of underlying experience. This paper contributes to the debate by introducing the distinction between constrained and heavy-tailed valence as a related but underexamined issue, by presenting initial evidence for the latter, and by describing practical implications for improved measurement, policy analysis, and policy ambitions implied by an HTV psychology.

The structure of the paper is as follows. The literature review in section two defines our terms within the context of academic work on emotions (§2.1), introduces the debate between cardinal and ordinal interpretation of measurement scales (§2.2), and sets out the conditions under which five empirical predictions can be derived to differentiate an HTV psychology from a constrained valence psychology (§2.3). The method section explains how our pilot survey approach tests two of these five predictions, by asking respondents to compare the most painful and most pleasurable experiences they can recall, alongside the second most painful and second most pleasurable such experiences.

The results are presented in section four, finding cautious support for the hypothesis. For instance, over half of respondents said their most intense experiences was at least two times as intense as the second most intense, which suggests only little extrapolation needed for the full range from the mildest experiences to span at least two orders of magnitude. Simulations also demonstrate a better fit to a heavy-tailed underlying distribution. Our results also raise doubts about the suitability of 0–10 integer scales, at least at the high end, with 81% of users opting for additional granularity when it is available. Over 85% of individuals also describe their most extreme experiences as more intense relative to their second such experiences than would be implied by the scores they place on a 0–10 scale.

In the Discussion section, we first summarize the key findings and explain how they might be interpreted in the context of other research that points toward the sufficiency of a 0–10 integer scale (§5.1). Specifically, we identify the possibility of a "kink" in self-report habits, such that the approximate cardinality of most of the scale, up to around 8 perhaps, may be sustained alongside a compression in reported experiences at the top end of the scale. Secondly, we present two practical implications for measurement techniques and two practical implications for analysts and research funders that would improve policy making in HTV settings (§5.2). We then turn to addressing potential theoretical criticisms of the HTV hypothesis. One concern is that the capacity for experience could arbitrarily be mapped to different scales without any implications for subjective experience. We refute this arbitrariness claim by exploiting the phenomenon of just-noticeable differences in a novel thought experiment: "the integer experience test" (§5.3). Brief accounts are then presented against six stylized facts that run counter to the HTV hypothesis, with three stylized facts presented in its favor (§5.4). Finally, we discuss the limitations and lessons learned from the pilot study, to lay the groundwork for larger scale empirical testing of the hypothesis (§5.5).

The conclusion summarizes the paper and contextualizes it within current data on wellbeing in high average income countries. Unlike the constrained valence hypothesis, our HTV hypothesis leads to a dramatically different interpretation of the current data. Rather than the complacency or incremental improvement potential revealed in the former, we would argue for far greater policy ambition. The future need not be slightly better than the present – it could be almost unimaginably better.

2. Literature review

This section first defines our terms in the context of academic work on emotions (§2.1), then introduces the debate between cardinal and ordinal interpretation of measurement scales (§2.2), and finally sets out the conditions under which five empirical predictions can be derived to differentiate an HTV psychology from a constrained valence psychology (§2.3).

2.1. Measurements of pleasure and pain

The measurement and taxonomies of emotion remain contested today, with no shortage of alternative theories (Mauss and Robinson, 2009; Ekkekakis, 2013; Keltner, 2019). Nonetheless, in most dimensional models of emotion, there is an axis for positivity or close variants on that theme: pleasurable-unpleasurable (Wundt, 1897), pleasantness-unpleasantness (Schlosberg, 1954), high-low valence (Russell, 1980), and so on.

Emotions at the negative end of these axes have typically not explicitly named pain, perhaps considered more of a sensation than other emotions named on the scales, such as stressed, anxious, fearful, or hostile. However, this may be an omission given developments in the understanding of pain. In a recent cross-disciplinary paper of medical, psychological, and psychiatric experts, Gilam et al. (2020) emphasize that pain is defined as an unpleasant subjective experience with a sensory and an emotional component, although they acknowledge (and regret that) pain has traditionally been researched and clinically treated separately from emotion. Pleasure is similarly not a "pure sensation" (Berridge and Kringelbach, 2008), although externally-stimulated sensations may often be a key input.

In the taxonomy of Dolan and Kudrna (2016), our research focuses primarily on experienced pleasure and pain sensations while acknowledging that some evaluative component remains present, particularly in survey instruments that rely on recollections of past experiences. Such experiential happiness is not easy to capture, but several high-quality methods are available for it. For instance, the Experience Sampling Method (ESM) asks people at random times of the day how they are feeling in the moment. The Day Reconstruction Method (DRM) asks people to recall how they felt during various activities over a 24-h period, which covers a longer period of time but at the cost of overlaying additional memory and evaluative processing of the experiences. Both are burdensome techniques in normal application and if such exercises were to capture rare, peak events, they would need to be asked over a long period of time and most likely in a wide range of circumstances. Such circumstances are unlikely always to be conducive to survey completion, especially considering the complex relationship between emotional intensity and self-awareness (Silvia, 2002), such as may be prompted by taking surveys about your emotions. What we pay attention to in a moment, itself an influenceable phenomenon, is also likely to be important for wellbeing and may differ from judgments about our preferences made in hindsight (Dolan et al., 2021).

In practice, many survey instruments only have space for fewer questions and less frequent surveying, often favoring 11-point integer scales from 0 to 10 for experienced happiness and more evaluative measures of happiness as a result (OECD, 2013; Plant, 2019). Indeed, measurement of pleasure/pain most commonly takes place on short self-report scales using common-sense language (e.g., Haefeli and Elfering, 2005; Dolan and Kudrna, 2016). A 2019 review of Outcome Measures by the Faculty of Pain Medicine (FPM, 2019) presents three pain quantity measures out of 16 instruments related to the topic. It recommends the NPRS, a 0–10 integer scale anchored by 0 "no pain" and 10 "extreme pain/worst possible pain," over the five-unit verbal rating scale and marking a horizontal line in the Visual Analog Scale.

The discussion of pain in the context of pleasure illustrates the ambiguity in sensations vs. emotions. Some sensations can be mostly separated from emotional content, e.g., experiences of heat, proprioception, or the color yellow may evoke none or several different emotions depending on the context. However, the sensation of pain is almost definitionally valent, whether based on external sensations (nociception), damaged nerves (neuropathy), or system hypervigilancy (nociplasticity). If there is no felt unpleasantness or discomfort associated with a potential pain experience, arguably no pain is being felt.

Our research focus on the capacity to experience adds an additional complication. It may be hard for an individual to understand their personal capacity to experience emotions until having tested those limits or had various uncommon, extreme experiences - or at least witnessed them at close enough quarters to empathize with the participant. We also note reason to believe that the capacity for emotional experience varies from person to person, given psychological instruments to measure such variation at a trait level (e.g., Larsen and Diener, 1987; Bachorowski and Braaten, 1994) and analysis of reported pain sensations in response to the same clinical stimulus (e.g., Wiech, 2016; Fillingim, 2017; Gilam et al., 2020).

2.2. Cardinal and ordinal scale interpretation

Survey scales may typically adopt a fairly constrained integer scale, such as from 1–5 or 0–10, but that does not mean the scale has to be interpreted in a cardinal setting, i.e., where an 8 is not just "high" and "much higher than 4" but specifically twice the value of 4 and the gap between 9 and 10 is the same as between 6 and 7. Indeed, many researchers have criticized such interpretation of Likert-style scales, arguing instead for an ordinal-only interpretation (e.g., Bishop and Herron, 2015; Kero and Lee, 2016; Wodak, 2019).

In clinical settings, pain scales often can be productively used with only an ordinal assumption, tracking self-reported progress over time and informing decisions on managing pain severity (e.g., medication or activity restrictions). The latter requires some common usage of language but this can be managed pragmatically by calibrating within an individual patient's experience over time. Indeed, there is widespread clinical acknowledgement of the limits of inter-personal comparisons given apparent individual variation in pain experience/reporting (e.g., Fillingim, 2017; Gilam et al., 2020).

Unfortunately, in a policy setting, particularly for wellbeing economics, average empirical insights drawn from scales like these are often used in a more strictly cardinal sense (e.g., What Works Wellbeing, 2017; HM Treasury, 2021). The cardinal interpretation of such scales as a reporting function in the context of human communication is commonly applied in practice as a necessary default.

The Happier Lives Institute (HLI) provides a rare theoretical defense of this cardinal interpretation (Plant, 2020). Their account predominantly centers on evaluative wellbeing data (with metrics like life satisfaction). However, there is an aspiration to expand its scope to incorporate additional measures, such as hedonic wellbeing, the focus in our particular scenario. We are unsure that evaluative and hedonic wellbeing are necessarily experienced with the same range of potential capacity, noting differences discussed by Dolan and Kudrna (2016). In particular, evaluative wellbeing may be more constrained at the top end by cognitive and meta-cognitive considerations, such as concerns about future implications or repeatability and what satisfaction is being measured relative to, i.e., what someone might expect or feel they deserved, relative to their personal past experience, identity narratives, and social norms within different communities. However, similar statistical methods as Plant (2020) could be applied to hedonic wellbeing data collected in the future, such as the argument from homoskedasticity of errors.

The theoretical case from Plant (2020) can be applied more directly to hedonic wellbeing without the need for new data collection and analysis. He argues that respondents are likely to interpret a 0–10 scale linearly by default, pointing to analogies with linear scales elsewhere for known cardinal entities (such as distance or income where objective and subjective measurements line up tolerably well), the mathematical difficulty of working with non-linear scales, or the game theoretic consequences of using scales to support effective interpersonal communication (the "Grice-Schelling hypothesis"). However, such rational application may not mean it accurately reflects the range of feeling in the moment, even though we use it afterwards for ease of communication or claim with hindsight how the scale "should" be used. By contrast, the widespread presence of accurately equi-interval metrics elsewhere in society (e.g., measuring distance/weight) might mean we have a tendency to over-impute and rationalize linearity into situations for which it is inappropriate (consider the representational fallacy discussed in Wodak, 2019).

HLI is currently enhancing this theoretical account with empirical data, exploring surveys to test the cardinality of 0–10 scales. Pilot results, maintaining the focus on evaluative wellbeing only, are presented in Samuelsson et al. (2023). The results are tentatively supportive of cardinality but with significant variation and an acknowledgement that more research is needed. For instance, only a slim majority, 56% of participants reported that they used the scale linearly. Most participants said they interpreted the end points of a 0–10 scale as the most extreme possible, split between the most extreme possible for any human and for themselves, but with some inconsistency in answers. However, a substantial minority anchor the reference points in their personal previous experience. They also find (ibid, Figure 15) interesting within-persons variation between their responses on an 11-point and a 10,001-point scale, although the between-persons averages at each of the 11 points line up linearly.

A possible reconciliation of the HTV hypothesis with the HLI account is presented in §5.1, building on the observation that HLI data and principles apply most strongly to common wellbeing experiences, i.e., day-to-day experiences, whereas our current evidence applies only at the extremes.

2.3. Contingent empirical predictions of the HTV

The key difference between the HTV and a constrained valence hypothesis concerns the human capacity for experience, with implications both for our understanding of the underlying psychology and for how it is measured. If the most intense pleasures (or pains) are within a single order of magnitude of the mildest pleasures (or pains) or if anything beyond that is discountably rare, then we would describe this as a constrained valence psychology. If the capacity to experience spans a much wider range, say at least two orders of magnitude, and these intense experiences are accessible to us then HTV psychology applies.

The inaccessibility of private experiences means that additional assumptions are needed to differentiate the two hypotheses. Under different sets of assumptions, we can specify five empirical differences between the survey results you would expect if asking people to reflect on experiences sampled either from an HTV psychology or a constrained valence psychology.

2.3.1. Ratio of intense to mild experiences

Focusing first on direct measures of the span of experience capacity, provided individuals have had enough experiences and can call enough of them to mind (even if only in a general sense) to encompass some mild or neutral events and some more extreme events, we would also expect most of them to describe the differences between their least and most intense events as dramatic, whether using narrative or reflecting their intuition as best they can numerically. Where they feel able to use a numerical parallel, most people with diverse life experiences should describe this span as more than 100x, i.e., two or more orders of magnitude.

2.3.2. Ratio of intense to average and average to mild experiences

Similar principles can be applied using the average point instead of a neutral point. The ratio of intense to average should be higher than the average of average to mild. This approach is less sensitive to identifying the most extreme and most neutral experience and it may be easier for respondents to reflect on an average experience as a reference point. However, it is a less direct measure of the actual span of experience.

2.3.3. Ratio of most to second most extreme experiences

Provided individuals have not had so many experiences that the full spectrum of possibilities is filled in, we would also expect larger differences between their most extreme and second most extreme memories. If you sample only 10 experiences first from a 10-step scale and secondly from a million-step scale (where the scales here linearly reflect the true underlying range of experience), using any identical distribution that spans the full range of experiences, whether via a uniform, normal, or heavy-tailed distribution, the ratio of the largest to the second largest will, on average, be much larger if sampled from the million-step scale. Whereas if a million or more experiences were sampled, this ratio can arbitrarily approach one on any continuous distribution. Provided the number of experiences sampled and recalled for the purposes of the comparison remains well below such levels, we arrive at a meaningful ratio. The stated ratios can then be extended to identify the implied number of equivalently sized steps to span at least two orders of magnitude from the top to the bottom end. If this number of steps feels intuitively low compared to the range of experience that actually exists or is reported as such by participants, it is indirect evidence for the HTV.

The previous three methods test the pragmatic relevance of the upper ends of intensity indirectly, in that if survey respondents are remembering them then we infer that the events are likely occurring with sufficient frequency to be relevant. In other words, vanishingly rare events would also be vanishingly rare in our survey data. As an aside, this implies that rejections of the HTV using these tests cannot differentiate between rejections based on span size as opposed to extrema accessibility. It is also possible to test distribution of experience more directly.

2.3.4. Distributional fit

All else being equal, a heavier tailed distribution will, by definition, have more extreme events more frequently than a narrow-tailed distribution. Indeed, a narrow-tailed distribution is one of the phenomena that could mean a high capacity to experience nonetheless translates in practice to a constrained valence psychology. A narrow-tailed distribution is likely unable to model the range of experiences in the majority of moderate, quotidian experiences while still preserving enough likelihood for outlier events two or more orders of magnitude out.[1] Distributional fit could be tested directly on valence data across individual experiences or indirectly on which underlying distributions would produce other metrics, such as the ratios specified in the first three methods. A limitation of this method, unlike the first three, is that the hypothesis is not yet explicit on how accessible extreme events need to be for relevance and so it is unclear how heavy-tailed a distribution needs to be. However, canonical distributions that are commonly approximated in nature could be used as initial reference points, such as the normal distribution (generated, e.g., via additive input processes by central limit theorem mechanisms) compared against the lognormal distribution (generated, e.g., via multiplicative input processes).

From an HTV perspective, recollection-based survey approaches should underestimate the true capacity to experience, as it is highly likely that most people's best and worst actual experiences by a particular date are not the best or worst experiences that are accessible in principle - or even the best or worst that they will have experienced by the end of their lives.

2.3.5. Direct inquiry

Finally, for completeness, it is also possible to ask people directly about the most intense and mildest experiences they can imagine and to comment on the span between them and what it might take to access different states.

Across all these methods, where we are analyzing responses from a sample of individuals, we also need to assume that differences between individuals in terms of how they report emotions and differences in terms of the underlying capacity to experience are either not correlated or at least only modestly biased with respect to the range being analyzed.

3. Methodology

We conducted a pilot survey to test two of the five empirical predictions from §2.3. This section explains the survey design (§3.1), the analytical methods that allow the survey question data to be related to the empirical predictions (§3.2), and finally the survey implementation, data cleaning, and participant demographics (§3.3).

3.1. Survey design

The survey is designed to compare the most and second most extreme experiences recalled, rather than average or mild experiences, reasoning that the assumption on modest numbers of recallable peak experiences is more plausible than assumptions on the reliability of assessing mild or average experiences. Further, a setting in which someone is more easily able to recollect peak pleasurable experiences may report their average experience as more positive, setting up a confounding correlation in the analysis of interest. It may also be hard to identify an average experience as an abstract idea and it might be heavily affected by recent activity. Whereas by asking respondents to focus on specific events, we are more confident they have specific emotions in mind. For the same reason, we did not prioritize the direct inquiry method, noting also that it elides differences between the capacity to imagine and capacity to experience, and that the under-estimate approach of recollection surveys helps build in a conservative methodology that would, all else being equal, increase confidence in HTV-positive findings.

We asked respondents to state what their most, second most, and third most pleasurable experiences were, to explain what category they fell into, and to write a short account of the most intense experience, with identical questions for the most painful experiences. With respondents having these experiences in mind, we asked them to use a slider to rate each one on a scale from zero to ten, from no pleasure/pain to the highest possible pleasure/pain. Non-integer responses were allowed out to a single decimal point. Illustrative descriptions were placed along the scale to support interpretation, anchoring the null experience at "0" for none, followed, e.g., by "1. Slightly bad feeling," "3. The pain is bothering me but can be ignored," and "8. The pain is so intense it is hard to think of anything else." These scalar questions allow us to contrast against standard measurement scales used.

The main novel question that generates traction on the HTV is then asking for free text estimates of the ratio: "Relative to the 2nd most painful experience, how many times worse was the single most painful experience in #1?" This paper refers to this question as the "described ratio" question. Respondents may have been primed by first providing the scalar responses, such that a numerical intuition for cardinality and consistency over-rides a true reflection on the felt experience, e.g., calculating that the ratio should be 1.25x as they have already provided a 10 and an 8 for the scalar questions, even if 1.25x underplays their felt experience. However, this bias was tolerated on the same basis as the recollection approach: it errs toward a conservative methodology that disfavors the HTV.

In designing the analysis, we wanted to allow individuals to think in terms of intensity of experience, accepting that different angles on pleasure or pain may have different intensities for different people. Some may think more of blissful joys, others more of adrenaline-filled thrills; some may focus on heartbreak, others on physical pain. It is debatable how much pleasures and pains with different sources and inflections are directly comparable, but nonetheless it is often possible to comment on which is more intense and whether it is much more or only slightly more intense, suggesting that many of us have some internal mechanism for forging quantitatively-nuanced comparisons even if the underlying experiences are multidimensional. Such mechanisms are required and implied by methods such as willingness to pay, time trade-off, and standard gamble questionnaires used in health economics (e.g., Lipman et al., 2019).

We also asked about current age and age the experiences happened, gender, current feeling of pleasure/pain, and which of their most extreme pleasure or pain was more intense.

3.2. Analytical approach

In addressing our primary question, the likely presence of HTV psychology in the sample, we use two analyses based on the options

1 Simply increasing the standard deviation of a normal distribution does not achieve this same result. As with the standardized normal distribution, the relative distributional shape is robust to linear transformations. The standard deviation can be altered arbitrarily with unit scaling without changing the underlying distribution.

set out in §2.3. The first examines the intensity ratio between most and second most intense experience, as described directly by respondents, considering pain and pleasure responses separately. As well as examining the described ratios descriptively, we consider how many similarly sized steps would need to be present at minimum for felt experience capacity to span two orders of magnitude on its way to the mildest experiences.

Secondly, we conduct statistical simulations to examine whether the described ratio responses are a better fit to an underlying set of experience valences that are normally or lognormally distributed, as example canonical narrower and heavier tailed distributions.

As secondary analysis, the survey questions also permit some insight into how users might be engaging with traditional 0–10 integer scales, an important related measurement question. We report also therefore on how many respondents made use of decimal points and how many users gave ratio responses that were consistent with their scalar scores, as would by implied by cardinal use of a 0–10 score with 0 as the neutral point, as requested in the notes to the user.

3.3. Survey implementation, sample selection, and participants overview

We ran the survey on Mechanical Turk in 2019, with respondents receiving US$ 1.75 for completing the survey. In addition to the implicit inclusion criteria (English speakers with access to the platform), there were two explicit inclusion criteria for responding to the survey: a good track record of task completion on Mechanical Turk and a master's qualification, both designed to increase the chance they would engage with the questions. Funding approval and operational/ethical sign-off were provided by the Qualia Research Institute leadership team, noting that the survey was opt-in, open only to adults on an anonymous basis, and reimbursed with a modest contribution to thank them for their time. The request to reflect on emotionally challenging experiences was seen as balanced by the request to reflect on pleasurable experiences, and by the focus on highly-educated respondents more likely accustomed to exploring challenging topics via questions and essays.

The initial 110 responses were analyzed to remove 13 likely bots (or non-serious completions), based on non-sensical or off-topic responses to the mini essay questions. Further exclusions or adjustments were also applied for responses that had ambiguous interpretations or were not mathematically consistent. Given uncertainty in this process, we tested the analyses against two types of sample. First, against smaller but clean samples having removed any ambiguous or conflicting answers. Secondly, against the maximum possible sample based on interpreting ambiguous answers in the direction that is most disfavorable to the HTV, in line with our other conservative methodological choices. For instance, a reported "1x" difference between the most and second most intense experiences would conflict with non-identical scalar scores. The clean sample excludes such responses but the maximum possible sample applies the HTV-disfavorable interpretation that the 1x is a valid response (i.e., no difference), assuming that the respondent's differences on the scalar questions are felt to be of negligible importance to them.

The full detail of exclusions and interpretations is shown in Table 1. For the primary analyses focusing on the described ration questions, the maximum HTV-disfavorable sample is 91 for the topic of pain experiences and 95 for pleasure experiences, i.e., sample exclusion rates of 6% and 2%, respectively. The clean sample for described ratio analysis is 77 for both, i.e., a sample exclusion rate of 21%. Secondary analyses contrast against the scalar responses, i.e., the 0–10 intensity scores, for which the clean samples are 65 and 64, respectively.

TABLE 1 Exclusions to generate clean samples for analysis.

Sequential exclusion steps for described ratio analyses		Sample size	
#	Exclusion detail	Pain responses	Pleasure responses
1	Full initial sample	110	110
2	Likely bots or non-serious completions, based on non-sensical or off-topic responses to the free text questions	97 (−13)	97 (−13)
3	Those who provided described ratio text that could not be interpreted as a number, e.g., "much worse"	96 (−1)	96 (−1)
4	Those reporting very high ratios of 100x or higher that perhaps have a narrative interpretation of "much much more intense" but risk being misleading if interpreted mathematically at face value (the HTV-disfavorable interpretation is to exclude such outliers)	91 (−5)	95 (−1)
5	Those who reported a "zero times" ratio difference or equivalent (interpretable disfavorably as the experiences being the same, i.e., 0x would be coded as 1x)	88 (−3)	93 (−2)
6	Those who reported a ratio difference between zero and one (e.g., 0.5), which mathematically implies that the second most extreme was more extreme than the most (only interpretable as the given ratio being on top of the original, e.g., 0.5x would be coded as 1.5x the original)	85 (−3)	86 (−7)
7	Those reporting "one times" difference or equivalent but whose reported scalar scores were not identical or within 0.4 points (interpretable disfavorably as 1x, although they might have meant one times better/worse on top of the original experience)	77 (−8)	77 (−9)
8	For separate analyses comparing the descriptive ratio responses against the scalar 0–10 question responses, we further exclude those whose scalar responses reveal a mathematical misunderstanding, i.e., the second most intense experience is scored as more intense than the most	65 (−12)	64 (−13)

TABLE 2 Sample demographics.

Demographic aspect	*N*	% Male[1]	Age range	Age mean (st. dev.)
Maximum HTV-disfavorable sample (pain)	91	49%	21–64	37 (10)
Maximum HTV-disfavorable sample (pleasure)	95	48%	21–64	37 (10)
Clean described ratio sample (pain)	77	47%	21–64	37 (10)
Clean described ratio sample (pleasure)	77	48%	21–64	37 (10)
Clean comparative sample (pain)	65	51%	21–60	37 (10)
Clean comparative sample (pleasure)	64	48%	22–64	38 (10)

[1]Note that all respondents self-described as either Male or Female (free text box).

Table 2 reports the sample demographics, identifying a near even gender balance and a wide age range, broadly consistent across the different analytical samples.

4. Results

4.1. Capacity estimation given described intensity ratios

Table 3 reports descriptive statistics for the described ratios between the most intense and second most intense experiences, for both the maximum HTV-disfavorable sample and the clean described sample as defined in Table 1, 2. The results suggest a wide range of described intensity ratios, from effectively no difference between most and second most intense experiences (i.e., ratios of 1x) to those suggesting far more dramatic differences (e.g., ratios of 5x+).

The median is 2x for both samples and both pleasure and pain. In other words, 50% of respondents describe their most intense experience as 2 or more times as intense as the second most intense. It would require around six equivalently sized ratio steps between the mildest and most intense possible experiences to support the two orders of magnitude in the HTV. Within the clean sample 75% or more respondents identify an intensity ratio of 1.5x or higher, which would translate into 10 equivalently sized steps. Even ratios of only 1.1x-1.2x would only require some 30 equivalent steps.

We note also that the sample exclusions in steps 3 and 4 of Table 1 correspond to responses that would strongly endorse an HTV psychology. All of these responses (6 for pleasure; 2 for pain) described dramatic differences between the most and second most intense experiences, such that it would require very few further steps (none, in some cases) to span two orders of magnitude. Even if these numbers are best interpreted qualitatively rather than quantitatively, they point to strongly felt differences in the respondents.

4.2. Statistical simulation of underlying valence distributions

Subsequently, we evaluate the described ratio responses by comparing them to simulations based on an assumed underlying valence distribution of experiences, choosing either a normal or log-normal latent distribution as examples with differing kurtosis. As discussed earlier, extrema ratios will approach one as the simulated individuals are assumed to isolate individual experiences across a

TABLE 3 Described ratios of most intense and second most intense experiences.

Descriptive statistics	Maximum HTV-disfavorable sample	Clean described ratio sample
Pain responses		
N	91	77
Range*	1.0–50.0	1.0–50.0
Mean (standard deviation)	4.9 (9.1)	5.6 (9.7)
Median (interquartile range)	2.0 (1.2–5.0)	2.0 (1.5–5.0)
Pleasure responses		
N	95	77
Range*	1.0 - 99.0	1.0–99.0
Mean (standard deviation)	5.1 (13.8)	6.0 (15.2)
Median (interquartile range)	2.0 (1.1–3.0)	2.0 (1.5–3.5)

*i.e., 1x would mean the most intense and second most intense experiences have effectively the same intensity for that respondent; 1.1x would correspond to 10% greater intensity, e.g., a 9.9 on a truly cardinal ratio scale compared to a 9.

larger number of recalled experiences. We simulate three scenarios reflecting different such numbers of recalled experiences (10, 100, and 1,000 experiences per individual), with 1,000 simulated individuals in each scenario. This part of the analysis is presented only for the described ratios for the pleasure questions ($n=77$), because the simulations lead to the same conclusions for the pain questions, as expected given the high similarity between the two distributions (Table 3).

The described ratio distribution we are analyzing is the ratio of two extrema - the largest and second largest an individual can recall - rather than the underlying distribution of individual valence experiences. As a result, direct measures of tail heaviness (such as kurtosis) or standard measures of normality are not applicable. Instead, the extrema statistics end up individually distributed in the limit according to the GEV (generalized extreme value) distribution, regardless of the underlying distribution (provided regularity conditions are met; see Haan and Ferreira, 2007). A convenient outcome of this process is that the standard deviation parameter chosen for the normal distribution comparison does not affect the quality of fit for the emergent extrema ratio distribution. Any positive standard deviation parameter chosen for the underlying normal distribution of experiences translates into the same distribution of extrema ratio (all normal distributions have kurtosis of 3), so we can choose a standard deviation of one without loss of generality and without a need for an optimization process to identify the best fit.

For the lognormal distribution, larger standard deviation parameter choices result in larger ratios in the subsequent extrema ratio distributions. The plotted data therefore compare a standard deviation of one in both cases as sufficient to demonstrate the better fit of the lognormal distribution, noting that a distribution with higher variance, skew, and kurtosis can easily be generated in the lognormal case to improve the fit yet further if desired. A mean of 5 is entered into both normal and lognormal distributions, to ensure that no extrema are negative in the former case. Our simulations suggest the distributional shapes are not sensitive to different choices of sufficiently positive means (only the y-axis range varies), provided the extrema ratio are always based on analyzing two positive numbers.

Finally, we normalize each distribution so it maps to a 0–1 scale, noting the arbitrariness of any maximum unit selection for the same underlying feature of reality, so that all three distributions can be plotted on the same rank-ordered chart for ease of comparison. The two simulated distributions are downsampled to n = 77 to match the applicable survey sample size for the same purpose, using equally spaced multiples of 13 from the size-ordered distribution of 1,000 simulated individuals plus the top extremum. This normalization retains all important parts of the distribution for our purposes because the actual numerical start and end points of the scale are arbitrary in any distribution. Post-normalization, rank ordering and relative centered size of individual data points are the same; skewness, and kurtosis features are unchanged; standard deviation changes only by the scaling parameter.

Figure 1 shows the three resulting distributions plotted on one chart for each scenario, reflecting different numbers of underlying experiences being sampled by individuals. In all three scenarios, the lognormal-derived distribution (green) is a closer fit to the survey respondents (red) at virtually all points in the distribution than the normal-derived distribution (blue). In the case where 1,000 experiences are sampled, the lognormal-derived distribution is a particularly close fit, suggesting that this particular parametrization may be worth investigating in future work.

4.3. Secondary analysis: assessment of 0–10 scale

Table 4 provides descriptive data on the scalar 0–10 intensity scores provided by respondents in the clean comparative sample. The majority of respondents appear to be anchoring their most intense historical experiences as fairly close to the most intense possible, with 75% identifying their most intense pain at 8.8 or higher and their most intense pleasure at 8.9 or higher. While a few people provide low end scores, the vast majority of our data is around 8–10 on the self-report 0–10 scale.

Unlike common 0–10 scales, which permit only integer responses, our sliding scale allowed respondents to select intensity scores to one decimal point. Out of the 97 non-bot responses, 81% of respondents used a decimal value in at least one of the four scalar responses (two each for pleasure and pain). 57% of all such scalar responses did use a decimal value. Indirect insight on the importance of such additional gradations comes from how frequently such gradations are used since they permit in theory the cardinal mapping of the 0–10 scale to two orders of magnitude rather than one. This insight is indirect only as it does not guarantee that users are engaging with the scale on such a basis. A safer conclusion is that the majority of respondents prefer to use the additional gradations where they are present and that an integer scale is therefore potentially missing or eliding information about the respondent's felt experience.

A more important assessment of 0–10 scale cardinality comes from comparing the described ratios with the inferred ratios. If the 0–10 scale is intended approximately cardinally by users with a zero neutral point, then dividing the most intense experience score by the second most intense should give a similar answer to the description they give when asked to do so directly. Figures 2, 3 plot each respondent for pain and pleasure responses, using the clean comparative samples from Table 1. The described ratios are strictly larger than the inferred ratios for 88% of respondents on the pleasure experiences ($n = 64$) and 86% on the pain experiences ($n = 65$). They are 1.5x higher or more for 59% of pleasure experiences and 65% of pain experiences.

5. Discussion

This section first summarizes the key findings and explains how they might be interpreted in the context of other research arguing for the sufficiency of a 0–10 integer scale (§5.1). Secondly, we present practical implications for measurement techniques, analysts, and research funders (§5.2). We then turn to addressing potential theoretical criticisms of the HTV hypothesis, a concern that capacity for experience could be arbitrarily mapped to difference scales (§5.3) and stylized facts that run counter to the HTV hypothesis (§5.4). Finally, we discuss the limitations and lessons learned from the pilot study, to lay the groundwork for larger scale future testing of the hypothesis (§5.5).

5.1. Summary of findings and literature synthesis

The two primary analyses in the pilot survey cautiously support the HTV claim that our capacity to experience pain and pleasure spans at least two orders of magnitude. In the first analysis, 50% of respondents describe their most intense experience as 2 or more times as intense as the second most intense. As adults presumed able to draw on a range of recalled life experiences, it is hard to maintain such a difference in intensity within a constrained valence psychology. It would only require around six equivalently sized ratio steps between the mildest and most intense possible experiences to support the two orders of magnitude in the HTV. Within the clean sample 75% or more respondents identify an intensity ratio of 1.5x or higher, which would translate into 10 equivalently sized steps.

In the second analysis, simulations demonstrate that a canonical heavy-tailed distribution of underlying experiences fit these described intensity ratio data far better than a narrow-tailed distribution. In other words, intense experiences, such as those whose descriptions imply a broad capacity to experience, are more frequent and more accessible in lived experience than would be likely under narrow-tailed distributions. This suggests that such experiences are not so vanishingly rare that they can be pragmatically discounted. Moreover,

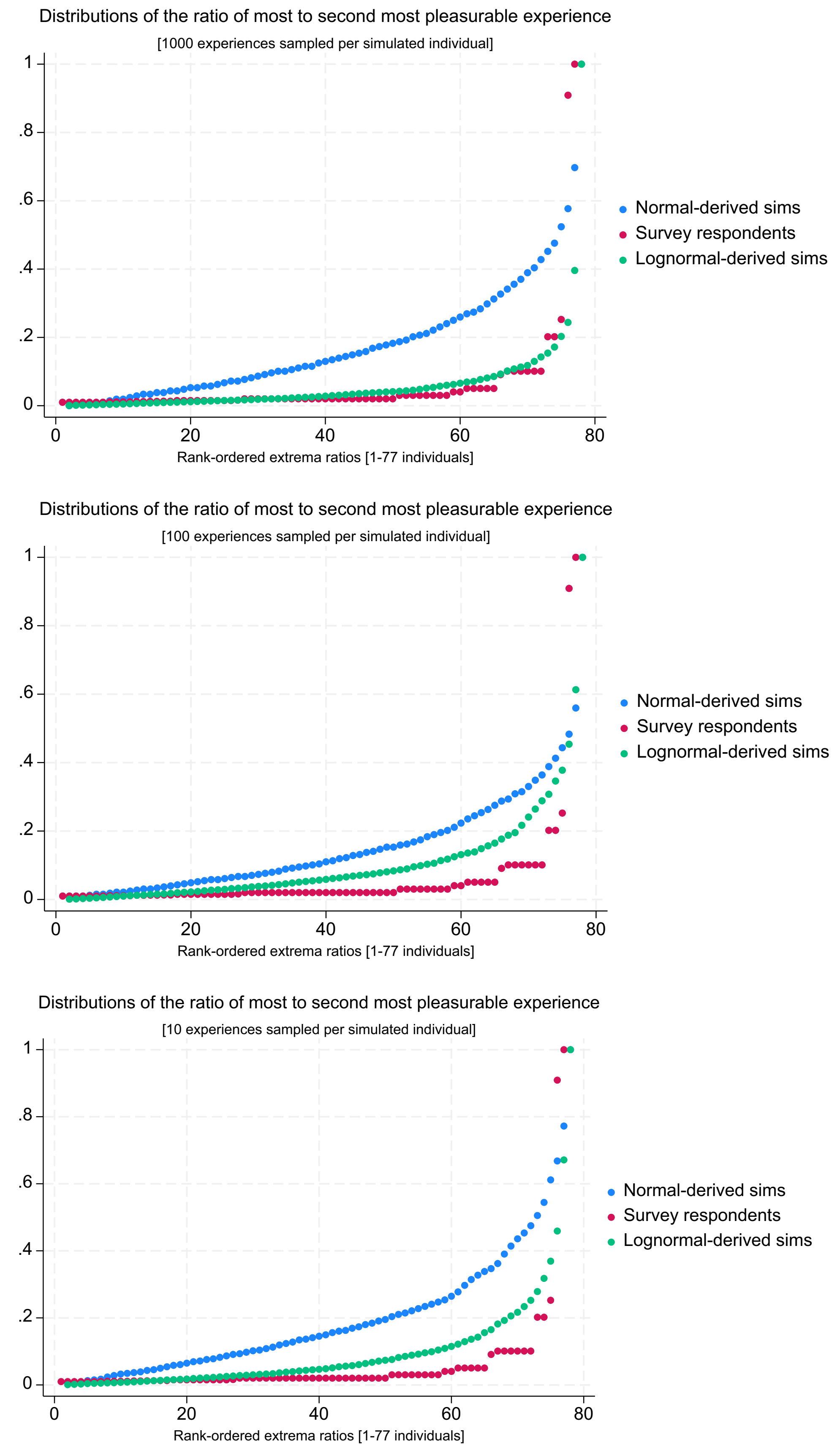

FIGURE 1
Survey respondents vs. thin/heavy tailed simulations – three scenarios.

our method is expected to underestimate the full range of the capacity to experience since respondents are asked to reflect on their actual remembered experiences. Many respondents, particularly younger ones, will on average have had a narrower range of actual experiences than are physically possible. We hope, for instance, that none of our respondents have experienced illegal torture.

Our data also question the suitability of a 0–10 integer scale for understanding intense experiences. Over 85% of respondents describe their most extreme experiences as more intense than their second most extreme experiences than implied by a cardinal interpretation of the 0–10 scale responses. 81% of respondents made use of decimal values in at least one of the four applicable questions, suggesting that they appreciated the additional granularity beyond the 11-points on an integer scale and that there might be valuable information to gain from such granularity.

TABLE 4 Intensity scores from the clean comparative sample.

Descriptive statistics	Most intense scores	Second most intense scores
Pain responses		
N	65	65
Range	3.0–10.0	1.0–10.0
Mean (standard deviation)	9.1 (1.3)	7.7 (1.8)
Median (interquartile range)	9.6 (8.8–10.0)	8.0 (7.0–9.2)
Pleasure responses		
N	64	64
Range	7.5–10.0	4.6–10.0
Mean (standard deviation)	9.3 (0.7)	8.4 (1.3)
Median (interquartile range)	9.6 (8.9–10.0)	8.5 (7.5–9.4)

Other research, discussed in section two, has argued however that an integer 0–10 scale can be interpreted cardinally in most practical circumstances, i.e., a single order of magnitude is sufficient (Plant, 2020; Samuelsson et al., 2023). Rather than take our evidence and the HTV as necessarily rejecting these claims, we suggest a reconciliation of the evidence with actionable implications for policy analysis.

Our synthesis suggests that non-linear interpretation of the 0–10 scale may only happen at the top end. Given that our survey focused on extreme experiences, our evidence by definition is restricted to the top end of the intensity scale for both pleasure and pain, with mean scores of around 8 or above.

People may broadly apply a linear interpretation (as required by Plant's Grice-Schelling hypothesis for effective communication) for most of an 11-point reporting scale, with 0 being a near neutral experience and 10 being the most intense they imagine is physically possible, up until perhaps 8 or 9. Between 9 and 10 there may be an additional order of magnitude or more of compressed experience, but most people may not have yet experienced the peaks and may even be unaware how much capacity is there. For everyday communicative purposes, all we really need to know is that 9 is already remarkably intense and that 10 is "even higher" since there is nothing above the 10. Indeed, for many policies governments are concerned with today, it may be adequate to consider insights up to 8 or 9 on the pleasure scale. For the pain scale, however, this is less comforting. This compression at the top end of the scale can be described as a "kink" in the scale.

A kink effect could reflect a general phenomenon in which reported outlier experiences are typically compressed when placed on

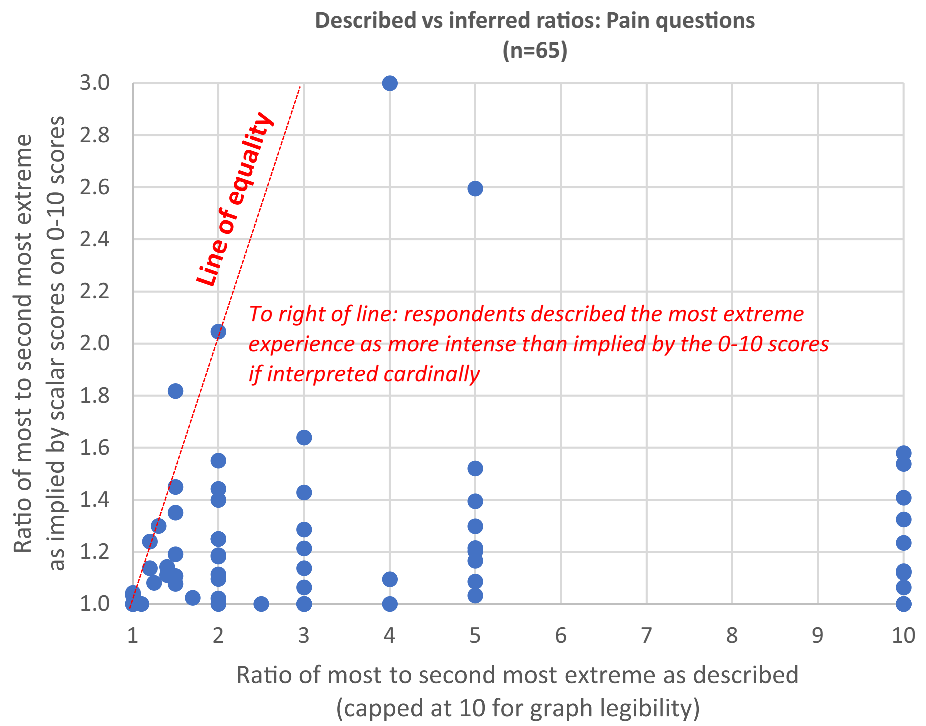

FIGURE 2
Described vs. inferred ratios: pain questions.

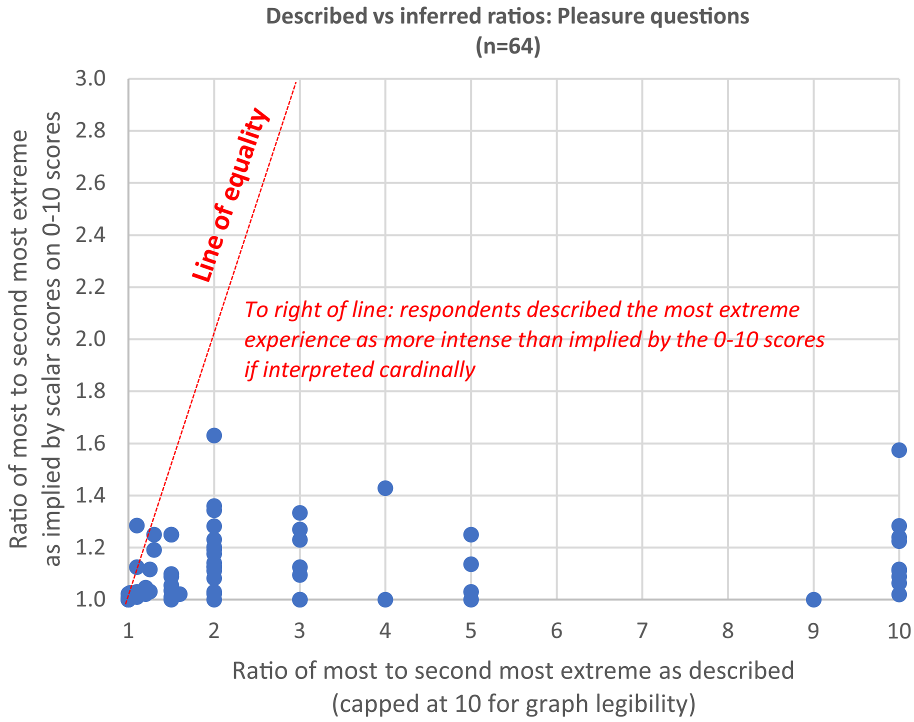

FIGURE 3
Described vs. inferred ratios: pleasure questions.

a scale without significant prior thought and analysis, consistent with either the log or arc-tan proposals from Ng (2008). One possible mechanism to generate this is as follows. Someone may have already used large parts of the scale to convey the fact that moderately intense experiences felt much more dramatic than mild ones, hence needing much larger scores on the integer scale for the numbers to align with felt intuition. As they turn their attention to more extreme experiences or the most extreme they can imagine, perhaps with prompting to help appreciate the full range of possible scenarios, they may find there is too little space left on the scale to capture the full range. To do so as best they can, without changing prior answers anchored from other discussions or first attempts to calibrate the scale to quotidian experiences, they have to compress the distance in each point reported as they approach the extremes. This is similar to a child counting from zero to ten for a task to be ready and realizing they need to add nine and a half, nine and three quarters, and so on to create extra time toward the end.

5.2. Implications for policy-focused analysis under the HTV

Our key claim is that felt experience intensity spans at least two orders of magnitude. This matters for measurement because many scales either assume a single order of magnitude or permit only ordinal interpretations. Under our hypothesis, there are four implications for improved policy-focused analysis, two focused on measurement and two on the analysis environment.

Measurement techniques should favor pleasure/pain scales with at least 100 and preferably 1,000+ gradations. For instance, if the currently common 0–10 scales are used, it should be possible to report a 7.8 or 9.3 rather than just 8 or 9 and this flexibility should be conveyed explicitly to users. Alternatively, a 0–1,000 scale might be presented directly. If visual analogs or "sliders" are used, they should also be sensitive to such levels of granularity.

Secondly, users providing measurements on scales should be advised about the location of a specific neutral, non-negative point, i.e., what number corresponds to an absence of any pleasure or pain sensations (perhaps 0). Similar to our pilot survey, several reference points should be provided along the scale with vivid corresponding adjectives and types of experiences. Current scales often state only the "best [or worst] possible," which conveys little about the scalar variation or the intensity of extreme experience, especially given the bland mindset that might prevail during administrative tasks like filling out surveys.

For analysts working with the integer data scales around a single order of magnitude that are common today, ordinal analysis techniques should be preferred. Where such techniques are insufficient, e.g., for comparative intervention trade-offs or cost–benefit analyses for policy/funding reasons, analysts should test the sensitivity of their conclusions to "kinks" in the measurement scale. For instance, testing whether results would still be valid if 8–10 corresponded to as large a change in felt experience as 0–8 on a 0–10 scale. Many policy interventions are analyzed based on arithmetic averaging of self-report data. To be valid in a kinked scale setting, the data need to be first weighted (i.e., with higher weights at the top end) before averaging and entering into subsequent analyses. For instance, to the extent that policies are based on the arithmetic average of pain scales, we can expect an increase in the priority of especially painful conditions once HTV is taken into account.

Finally, researchers and research funders/policy-makers should consider research that might explicitly test for kinks or non-linearities in 0–10 integer scales. For instance, trade-off survey techniques on medical conditions used to define QALYs and DALYs could also be applied to wellbeing questions (e.g., Lipman et al., 2019). Such "willingness to trade" surveys could test, for instance, how many reasonably pleasant days (5–7/10) someone might be willing to trade down to bland days (2–3/10) in exchange for one more "best day you have ever had/could imagine" (9–10/10). Insights on time consistency and intensity/duration trade-offs are needed to interpret such data (see, e.g., McAuliffe and Shriver, 2022) but building a body of such evidence nonetheless helps put parameters around the potential non-linearities that might be present in common survey reporting.

5.3. A counter to the arbitrariness critique

The HTV is contrasted against constrained valence in a phrasing that suggests there is a real difference in human experience. It suggests that we can imagine two different worlds, one where it is true and one where it is false, and it *feels* different to exist in either of those two worlds. It is not just that the hedonic rulers have different markings. One critique is to accept HTV but to restrict its relevance just to the design of measurement techniques: the number of gradations on a scale and how the scale is interpreted by users, rather than anything real about human experience. This critique might point to issues like the subjectivity of emotion (e.g., the unknowability and incommunicability of how your "worst pain ever" compares to mine) or the arbitrariness of unit selection (e.g., objects measured to the nearest millimeter span more orders of magnitude than measured to the nearest meter, but the real lengths are unchanged).

The "integer experience test" thought experiment refutes the arbitrariness critique by imagining an environment where one could toggle the HTV hypothesis on and off for a single experience, e.g., for a few minutes or hours engaged in a single activity, showing that the test participant both feels and reports genuine differences depending on whether the HTV setting is activated.

The thought experiment requires two core assumptions. The first assumption is that there is something to remark on. There are at least some experiences that vary in the pleasure or pain intensity evoked in an individual. This does not remove the possibility of multiple other dimensions that affect the emotional response, nor does it deny the possibility of complex, mixed-valence experiences (Gómez-Emilsson, 2022), and nor does it require the reaction to be the same across individuals or over time. The second assumption is that we do not have infinite sensitivity, i.e., that current human systems cannot perfectly identify differences at the smallest imaginable level of tweaking pain or pleasure. For instance, our nervous system is not perfectly sensitive. We can also imagine tiny numerical differences in scenarios that would not meaningfully alter the joy we experience from them (e.g., a life extended by a single femtosecond or a large lottery win extended by a single cent).

In this hypothetical test, a participant begins in a neutral state with no/negligible pain (or pleasure) being experienced. We expect most readers have experienced or can imagine such a neutral state.[2] Trivially small increases in pain are added into the experience, gradually increasing until the participant consciously reports noticing a difference, however small. That marks a single integer step or score, drawing on the principle of just-noticeable or least perceptible differences in psychophysics as commonly applied to direct sensory dimensions such as light brightness or sound intensity/pitch (e.g., Kollmeier et al., 2008). The process continues, continually marking integer steps with every consciously reported difference by the participant. While there would likely be very many such steps in any scenario, in an HTV there would be many more such steps – a higher score - before the human system loses consciousness or ceases to notice any difference no matter how much pain is added.

The same process can be done separately for experiences of pleasure, e.g., perfect brain simulations of different scenarios and sensations that elicit pleasure responses, noting many possible contributing factors in such scenarios (e.g., Gilam et al., 2020). If, at some point, additional pleasure experience translates (for any reason) into discomfort or unhappy emotions that the experimenter cannot correct for in the scenario, then we would conclude the pleasure capacity has been capped at the prior number of integer steps. Again, there would be more steps reported in the HTV scenario than a constrained valence scenario. Both the felt experiences and the reports of that experience differ between the scenarios.

The "integer experience test" is independent of any particular measurement scale or a participant's emotional sensitivity. It is independent because the test is intra-individual, i.e., the participant compares their experience in one hypothetical world vs. their experience in another. It cannot be implemented as a physical experiment, but as a thought experiment it can be presented against other theory-only critiques such as the arbitrariness critique.

5.4. Counters to stylized facts against the HTV

Our research and discussions have identified six stylized facts that might be levied against the HTV: (i) action potential phases; (ii) decreased sensitivity at extremes; (iii) hedonic adjustment; (iv) diminishing returns to scale in economic analysis; (v) behavioral change predictions; and (vi) an argument from evolutionary efficiency. Each is presented briefly below with a brief counter-argument.

(i) Many cells that are central to human conscious experience have action potential phases, notably neuron cells in the brain as well as the plasma membranes of most cells (e.g., Purves et al., 2012). The course of the action potential has several phases: the rising phase, the peak phase, the falling phase, the undershoot phase, and the refractory period. The last phase is relevant for this argument and corresponds to the period when subsequent action potential is very difficult or impossible to initiate. In other words, a forced pause after excitation. This places a biological limit on how frequently neural patterns can fire in a period of time. However, while this phenomenon may place a limit on conscious experience insofar as it is mediated by neural

2 As an aside, we do not necessarily associate such a neutral state with having "zero" value in any ethical sense, reserving this topic for discussion elsewhere.

patterns, it says nothing about where this limit may be – and how vast a range of experience it might demarcate.

(ii) The decibel scale is usefully logarithmic because human perception of sound intensity more closely responds to the logarithm of intensity, instead of its linear value. Effectively, the human ear reduces its sensitivity as the sensory input increases. If a similar principle applies to valence, it might suggest modest intervals of increased experience even as the drivers of that experience increase exponentially. However, it is also possible that we may only be able to definitely tell that a pain has got worse when increased by a constant percentage, but that does not mean that increased pain units below that threshold are not still unpleasant, it is just that we are so overwhelmed by the volume of pain it is hard to be sure. Sensory data is additionally only one of several inputs to emotional intensity (Gilam et al., 2020).

(iii) The hedonic treadmill is the claim that humans return to previous and relatively stable levels of happiness ("the happiness set point") following major experiences (Brickman and Campbell, 1971; Perez-Truglia, 2012). Various mechanisms have been proposed, e.g., assessing value against our past memories, evolutionary motivation to set new baselines and keep driving for improvement, neurochemical desensitization, or "abundance denial" given pressures of personal or social identity (e.g., Solomon and Corbit, 1974; Ahmed, 1998; Easterbrook, 2004; Rivat and Ballantyne, 2016). However, the empirical evidence on the hedonic treadmill is contested (Diener et al., 2006; Gardner and Oswald, 2007) and competing mechanisms can exist, such as when increased exposure can increase the joys experienced (e.g., connoisseurs of food or wine; Frederick and Loewenstein, 1999). Nonetheless, even if true, these arguments refer to adaptation or desensitization over time, rather than critiquing the possibility of dramatic experience in the present moment. This may have consequences for the political or personal implications of HTV but not for its truth as a description of our psychological capacities.

(iv) A more direct claim for adaptation within a single experience may come from the diminishing returns to scale applied in many economic models of human preferences and behavior. For instance, for most consumption goods, economists have long typically observed that gaining each additional unit reduces the utility we expect from the next unit and our corresponding willingness to pay (e.g., Gossen, 1854). If it becomes harder and harder to increase valence as individuals either suffer more or experience more pleasure, there may be practical limits to such experiences – although these limits could well be far off from everyday levels today.

(v) Indirect evidence to support the naïve interpretation of linear scales can be found in Kaiser and Oswald (2022), who show that self-reported dissatisfaction with various aspects of life is approximately linearly correlated with the probability of trying something new in that aspect of life. However, there is no compelling reason to believe that satisfaction with someone's job, house, or partner would be experienced with the same capacity range, or adequately reported on the same sort of scale, as emotions of pleasure and pain, nor that the probability of action should be linearly correlated to underlying emotions. Indeed, as emotions become felt more extremely, it is plausible that other parts of the mental machinery may attempt to dampen down the urgency to act in the present, so that the pros and cons can be weighed up in a more cautious, future-oriented frame of mind.

(vi) Feelings of pleasure and pain play an important function in improving our chances of surviving, reflected in brain structure and functionality (Brehm, 1999; Berridge and Kringelbach, 2008). Plant (2020) notes that processing and experiencing sensations is costly in terms of energy, reflected also in points around brain structure frugality for pleasure sensations. If we assume more intense sensations are more costly, then there is an evolutionary incentive to make our sensations and subsequent emotions only just intense enough to drive us toward action, with enough bandwidth to weigh up an appropriately broad range of options (noting that "wanting to repeat something" and "liking the experience" are related but not identical constructs). While it is unclear how much bandwidth would be needed to reflect the high dimensionality of options that the human system might face, this is an argument that urges toward more tightly bounded capacities. A counter argument would note that in the ranges of normal circumstances and behavior – presumably the ranges that evolution primarily incentivizes for - there may be many individual factors that need to trigger a positive/negative shift, which need to be combined in some way to generate the overall emotional input into decision making. It may be rare for many of these factors to co-occur, so the range of emotional experience is typically well bounded and energy efficient, but in order to account for all possible factors, the capacity for feeling should they all happen to co-occur needs to be vastly higher.

Further to the counter-arguments above, we have also identified three stylized facts that suggest the HTV is likely to apply in human context: (a) empirical observations about neurological function, (b) the accounts of those who developed and apply certain pain/discomfort scales, and (c) the presence of extreme events that might prompt dramatic emotional responses.

(a) Certain empirical observations about neurological function identify patterns that are characterized by heavy tailed distributions. If these heavy-tailed neurological features extend to the neurological components of valence experience, then the HTV hypothesis is more likely. In one example, Klaus et al. (2011) found that neuronal avalanches in macaque monkeys are characterized by heavy-tailed power law distributions. It is possible that more intense experiences sometimes correspond to more intense cascades of bursts of activity in particular neuronal networks. Power laws have also been reported in spike counts (Teich et al., 1997) and ion channel fluctuations (Toib et al., 1998), potentially due to information transmission optimization features (Beggs and Plenz, 2003; albeit contested, e.g., Bedard et al., 2006; Dehghani et al., 2012). Heavy-tailed distributions of neurological activity may also translate into heavy-tailed accounts of pain experiences, such as in cluster headache frequency data (Gómez-Emilsson, 2019).

(b) Vivid accounts of the range of possible experiences can be seen in the testimony of those who created certain sensory pain scales. The KIP scale for pain intensity is recorded on a 0 to 10 scale, with the explicit instruction in the context of cluster headaches to interpret the data logarithmically: a KIP 10 is not twice as bad as a KIP 5 but 10 times as intense (Cluster Busters, 2020). Schmidt's Sting Index reports the pain of insect skins based on personal experience on a scale that he suggests be interpreted logarithmically: "Each number is like 10 equivalent of the number before. So 10 honey bee stings are equal to 1 harvester ant sting, and 10 harvester ant stings would equal one bullet ant sting" (Peterson, 2018).

(c) Even if our own lives have been characterized by a relatively modest range of painful and pleasant events giving rise to a modest range of emotional responses, the possibility of far more dramatic events may allow us to infer proportionately more intense responses. For instance, the majority of the population who have never been physically tortured or taken heroin may nonetheless expect a truly intense experience if that were to happen.

In this paper, we suggest that the balance of stylized facts and counter-arguments point toward the HTV being true. However, there remains scope for dispute within these interpretations, meaning that empirical testing is required to establish the case either way with confidence. Given limitations in our pilot study, we would also recommend a larger and refined survey drawing on the lessons learned from this exercise.

5.5. Limitations and ideas for future studies

One foundational critique of our approach is whether it is reasonable to ask respondents to translate their felt experience into numerical scores. While the majority of respondents provided mathematically consistent answers, some did not, affirming the difficulty of this exercise. Self-reported measures of happiness and subjective wellbeing are widely used, but with frequent discussion of the possible limitations (see, e.g., Diener et al., 2018). Based on this experience, we suggest that useful insights can be gained from such data, even if best considered as numerical intuitions with considerable measurement error, rather than precise data. However, future surveys could take steps to help respondents engage with the method.

In this case, concerns may be exacerbated by the focus on extreme events, which may be harder to recall and analyze than evaluative wellbeing in general. We might worry that (some) respondents are hyperbolic in their responses or unable to quantify such feelings more generally, noting concerns about numeracy (e.g., Bruine de Bruin and Slovic, 2021). Acknowledging that such concerns cannot be fully alleviated, additional questions may help assess how worried we should be.

Questions could assess a propensity to hyperbole, perhaps through direct self-report, asking about how friends might describe the respondent, or asking questions that might elicit more easily tested exaggeration. Questions could similarly be designed to test someone's ability to quantify in general and reason about ratios in other contexts. Providing examples for the ratio question may help people feel more able to give very small increases without diminishing the difference. Definitions of experience duration and emotions vs. moods may also improve consistency. Similarly, we could ask directly whether the most and second most extreme experiences are "about the same" in intensity, even if different scores are given. Alternatively, we could move away from mathematical self-report to visual reasoning, e.g., ask people to draw or select a homunculus depiction that reflects the different intensities they feel.

The comparison of most to second most intense experiences is sensitive to a number of factors, as discussed in §2.3, including respondents who may be able to recall many thousands of experiences and the possibility that they had two unusual but very similar outlier experiences. A larger survey helps increase the confidence that a small proportion of such outlier individuals would not skew the sample. Now that this pilot survey has identified initial parameters for parts of the valence distribution, future surveys could also ask respondents more directly about the likely number of equally sized steps from mildest to most intense experiences, anchored on their extrema ratios. Diagrams could be used to illustrate this abstract request, as well as interactive applets to demonstrate how draft answers play out in practice. We could also randomize whether respondents are asked to describe the ratio or the scalar scores first, in case a deliberate aim for internal consistency alters their responses, as well as randomizing question ordering more generally. A larger sample would permit investigation of potential asymmetries between pleasure and pain, the location of non-linearities in the scale, relationships with age, and types of experience.

More work could also be done to test the other three empirical predictions of HTV detailed in this paper. For instance, we can ask people about their mildest and average recalled experiences and about the most intense and most mild experiences they can imagine. Providing scenarios may help anchor the extreme events that are imaginable. We might also gain insight into the capacity question by asking respondents how they think the extreme experiences they have reported compare to how much better or worse it could get, how their experiences might compare to average human experience in their region, how their friends might have described a similar experience, and whether they feel their definition of what a "10" experience would feel like has changed over time and why.

Other foundational concerns include biases around memory validity, placing some caveats on the precision of findings while still permitting an initial directional assessment. For instance, fading affect bias suggests that negative memories tend to be forgotten more quickly than positive ones (Skowronski et al., 2014), but would still permit separate analysis of each construct. Another possibility is not remembering enough experiences to compare them, for instance being able to identify the most intense but being unsure about the second, such that it ends up being selected almost at random from the next dozen or so recalled experiences. Focusing individuals on a particular type of pleasure (such as loving relationships) or pain (such as physical damage) may make it easier to recall the salient experiences at the cost of narrower scope.

More generally, where our brains continually reconstruct past experiences to generate present-day narratives and attempt to support present-day planning, it is possible the actual valence of past experiences might either be exaggerated or downplayed over time to better suit those goals. For discussion see, e.g., Fredrickson (2000), reconstructive memory theory (Hemmer and Steyvers, 2009), and cultural influences on memory (Wang, 2020). Diary methods could be used (e.g., the ESM and DRM methods discussed earlier) to capture experiences nearer to the time, but may need to span years to have a chance of capturing peak experiences for many individuals.

Thinking more ambitiously, we would also welcome alternative methodologies for investigating this question that do not rely as strongly on self-report of recollections. For instance, correlations between neural activity and both experienced and recalled intensity, qualitative longitudinal research, self-report relative to induced peak experiences, revealed preferences, trade-off surveys, and various non-self-report measures (for discussion see Lucas et al., 2003; Mauss and Robinson, 2009; Goto and Schaefer, 2020).

6. Conclusion

This paper has argued, cautiously, for the Heavy-Tailed Valence (HTV) hypothesis: that the accessible human capacity for emotional experiences of pleasure and pain spans a minimum of two orders of magnitude.

Where the hypothesis applies, we have provided actionable advice to the research community. For practical measurement scale design, we suggest allowing non-integer responses or 100+ gradations and providing vivid reference points for users. For researchers, we recommend testing robustness to kinks in felt experience at the top end of the scale and conducting trade-off surveys to calibrate scale interpretation, similar to those used in QALY/DALY estimation in public health.

In quantitative support of this hypothesis, we report results from a pilot survey in which over half of respondents said that their most intense experiences were at least two times more intense than the second most intense. As such, it would only take six steps of the same magnitude between the most mild and the most extreme experiences to identify a range of capacity spanning two orders of magnitude. The evidence is only indirect, but with enough room for error that the Heavy-Tailed Valence hypothesis has some base credibility, sufficient to motivate more robust testing, especially as methodological choices were made to disfavor the HTV. Additional indirect evidence is found in simulations demonstrating that the reported data fit better to underlying heavy-tailed distributions of experience valence rather than narrow-tailed distributions. In qualitative support of this hypothesis, we have discussed three stylized facts in its favor and identified counter-arguments to six stylized facts against it.

Assessment of survey evidence and stylized facts is important for analyzing hypotheses about the capacity to experience because personal introspection may reveal orders of magnitude variation in experience for some individuals but not others. The former may believe the hypothesis, but the latter have little reason to do so and may feel doubtful about accounts from the former. For instance, a study by Holz et al. (2021) suggests that while we generally trust others' accounts of their emotions, intense vocalizations of peak emotions are often distrusted. This paper suggests that the balance of evidence weighs in favor of the HTV, but recognizes the limited evidence so far and the importance of further research, drawing on the lessons learned from this pilot study.

In addition to measurement and analysis implications, the integer experience test demonstrates that the capacity for felt experience is not just an arbitrary or subjective choice of units. As a result, a prevailing HTV psychology also has important implications for personal and societal wellbeing ambitions. We close by briefly reflecting on these ambitions.

Recent evidence (e.g., Helliwell et al., 2023; Zimmer et al., 2023) shows most residents in high average income countries have fairly low pain prevalence (e.g., 25%) of mostly mild pain and seemingly high reported happiness (e.g., 7–8 out of 10). Organizations like the World Happiness Report and the What Works Wellbeing Centre explicitly interpret these as cardinal scales, implying there is room only for incremental improvements in the pain/pleasure components of wellbeing for the majority of residents. The numerical instinct that 8 is quite close to 10 might implicitly constrain societal ambition, focusing policy attention into different areas. However, combining our hypothesis and results with the work of Plant (2020), we identify a likely "kink" in the 0–10 measurement scale at the top end. As a result, we would make the explicit case that there is at least an additional order of magnitude of potential gains between around 8 and 10 in how the happiness scales are commonly used today. Such a "kink" leads to the opposite conclusion to mainstream think tanks; there is scope for much greater ambition than at present.

The capacity to experience likely also varies from person to person (cites in §2.1) and is likely amenable to alteration and training, as suggested by the effect of pain relief to dull emotional and even empathetic responses (Durso et al., 2015; Mischkowski et al., 2019), emotional blunting in SSRI treatments for depression (e.g., Goodwin et al., 2017), and therapeutic services both to reduce and increase the intensity of emotional experience (e.g., Engelhard et al., 2011; clinical trials on anhedonia, Phillips et al., 2022). Such context-dependency does not refute HTV. By contrast, it may make it more relevant. If our ability to train ourselves to increase the range of joy that can be experienced is much higher than we thought before, then there may be more value in investing in such effort.

A symmetrical implication applies to pain reduction. Heavy-tailed distributions of experience suggest that a large proportion of suffering might exist in the extremes. In such cases, some ethical frameworks might shift resource toward the most extreme cases of suffering. Even if it might take considerable effort to reduce someone from a suicidal 10 to a survivable 8 on a pain scale, this may outweigh improving many individuals from an annoying 6 to an ignorable 3. However, such implications are not guaranteed; they depend also on duration trade-offs, the tractability of the problem, probability of success, productivity gains for the beneficiary, and the costs of the interventions identifiable for further research.

Finally, we might briefly consider implications for individuals. As well as greater caution around avoiding extreme pains, we might consider the spectrum of pleasure. If peak events are hard to repeat more than a few times (such as parenthood), hard for individuals to influence (e.g., lottery or world cup wins), or draw on extensive build up for their emotional intensity (e.g., gaining a hard-fought promotion), there may be little we can do about it, even knowing such peak experiences exist. However, we suggest a more optimistic view, drawing inspiration from personal accounts and brain scans of activities like jhana meditation (e.g., Hagerty et al., 2013). That some experiences are hard to design for does not rule out the possibility for all experiences. We wonder what could be achieved with a more widespread attitude that peak experiences are possible, that an environment can be made more conducive to their occurrence, and that we can get better at noticing and appreciating them.

Reflecting on the link between hedonic and evaluative wellbeing, we wonder about the importance of integrating peak moments constructively into a personal life narrative. A single religious experience, with a 10/10 positive valence, may become a foundational memory for an individual, inspiring greater acts and happiness for years to come. By contrast, a single ill-timed drug experience may also have a 10/10 valence at the time, but be felt by the individual as a moment of shame and confusion in later years that they fear to experience again in case it leads to addiction; a memory to be pushed away rather than drawn upon. Could the latter be made more like the former? If life could be punctuated with many more and more incredible experiences than commonly believed, how differently would we live it for ourselves? How can we structure our collective institutions to promote, support, and leverage peak experiences?

Data availability statement

The datasets presented in this study can be found in online repositories at: https://osf.io/udmxn/. Details: Open Science Framework Data Repository under project name Heavy-Tailed Valence Hypothesis (HTV).

Ethics statement

The studies involving humans were approved by the Qualia Research Institute. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AG-E came up with the HTV hypothesis, identified intuitions for its validity, and ran the initial survey. CP developed the integer experiment test to demonstrate the hypothesis could not be collapsed into arbitrary unit mapping, identified evidence against the hypothesis, considered its validity, reviewed the literature to ascertain the originality of the argument, relationship to pre-existing research, and potential implications. AG-E and CP worked together to analyze the data and contributed to writing the manuscript and testing and refining the arguments. All authors contributed to the article and approved the submitted version.

Acknowledgments

We would like to thank the Qualia Research Institute for funding and encouraging this work, the survey respondents for sharing their intense emotional experiences, David Pearce for inspiring us to tackle the topic of suffering by illuminating its sources and magnitudes, Elisabetta Fato for research assistance, anonymous and named colleagues for their feedback during this process, including Aveek Bhattacharya, Ben Davies, Nancy Hey, Marcin Kowrygo, Michael Plant, Eric Schneider, and Asher Soryl, commenters on the Effective Altruism Forum, and the Frontiers editorial team for their support and constructive challenge throughout. Any errors of course remain ours alone and no endorsements of specific positions are implied by these acknowledgments.

Funding

The Qualia Research Institute and its leadership team provided the funding to reimburse the pilot survey respondents.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmed, S. H. (1998). Transition from moderate to excessive drug intake: change in hedonic set point. *Science* 282, 298–300. doi: 10.1126/science.282.5387.298

Bachorowski, J.-A., and Braaten, E. B. (1994). Emotional intensity: measurement and theoretical implications. *Personal. Individ. Differ.* 17, 191–199. doi: 10.1016/0191-8869(94)90025-6

Bedard, C., Kroger, H., and Destexhe, A. (2006). Does the 1/f frequency scaling of brain signals reflect self-organized critical states? *Physical Rev. Lett.* 97:118102. doi: 10.1103/PhysRevLett.97.118102

Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177. doi: 10.1523/jneurosci.23-35-11167.2003

Berridge, K. C., and Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology* 199, 457–480. doi: 10.1007/s00213-008-1099-6

Bishop, P. A., and Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *Int. J. Exerc. Sci.* 8, 297–302.

Brehm, J. W. (1999). The intensity of emotion. *Personal. Soc. Psychol. Rev.* 3, 2–22. doi: 10.1207/s15327957pspr0301_1

Brickman, P., and Campbell, D. (1971). "Hedonic relativism and planning the good society" in *Adaptation level theory: a symposium*. ed. M. H. Apley (New York: Academic Press), 287–302.

Bruine de Bruin, W., and Slovic, P. (2021). Low numeracy is associated with poor financial well-being around the world. *PLoS One* 16:e0260378. doi: 10.1371/journal.pone.0260378

Cluster Busters. (2020). Busting protocol — keeping track. Available at: https://clusterbusters.org/treatment-options-busting/keeping-track/ or https://web.archive.org/web/20230607175604/https://clusterbusters.org/resource/keeping-track/

Dehghani, N., Hatsopoulos, N. G., Haga, Z. D., Parker, R. A., Greger, B., Halgren, E., et al. (2012). Avalanche analysis from multi-electrode ensemble recordings in cat, monkey and human cerebral cortex during wakefulness and sleep. *Front. Physiol.* 3:302. doi: 10.3389/fphys.2012.00302

Diener, E., Lucas, R. E., and Oishi, S. (2018). Advances and open questions in the science of subjective well-being. *Collabra Psychology* 4:15. doi: 10.1525/collabra.115

Diener, E., Lucas, R. E., and Scollon, C. N. (2006). Beyond the hedonic treadmill: revising the adaptation theory of well-being. *Am. Psychol.* 61, 305–314. doi: 10.1037/0003-066X.61.4.305

Dolan, P., and Kudrna, L. (2016). "Sentimental hedonism: pleasure, purpose, and public policy" in *Handbook of Eudaimonic well-being*. ed. J. Vittersø (Cham: Springer), 437–452.

Dolan, P., Laffan, K., and Kudrna, L. (2021). The Welleye: a conceptual framework for understanding and promoting wellbeing. *Front. Psychol.* 12. doi: 10.3389/fpsyg.2021.716572

Durso, G. R. O., Luttrell, A., and Way, B. M. (2015). Over-the-counter relief from pains and pleasures alike: acetaminophen blunts evaluation sensitivity to both negative and positive stimuli. *Psychol. Sci.* 26, 750–758. doi: 10.1177/0956797615570366

Easterbrook, G. (2004). *The progress paradox: How life gets better while people feel worse*. New York, US: Random House Trade Paperbacks.

Ekkekakis, P. (2013). *The measurement of affect, mood, and emotion: A guide for health-behavioral research*. Cambridge, UK: Cambridge University Press.

Engelhard, I. M., van den Hout, M. A., Dek, E. C. P., Giele, C. L., van der Wielen, J.-W., Reijnen, M. J., et al. (2011). Reducing vividness and emotional intensity of recurrent "flashforwards" by taxing working memory: an analogue study. *J. Anxiety Disord.* 25, 599–603. doi: 10.1016/j.janxdis.2011.01.009

Fillingim, R. B. (2017). Individual differences in pain. *Pain* 158, S11–S18. doi: 10.1097/j.pain.0000000000000775

FPM. (2019). *Outcome measures*. London: The Faculty of Pain Medicine of the Royal College of Anaesthetists.

Frederick, S., and Loewenstein, G. (1999). "Hedonic adaptation" in *Well-being: Foundations of hedonic psychology*. eds. D. Kahneman, E. Diener and N. Schwarz (New York, NY: Russell Sage Foundation), 302–329.

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: the importance of peaks, ends, and specific emotions. *Cognit. Emot.* 14, 577–606. doi: 10.1080/026999300402808

Furedi, F. (2015). The Media's first moral panic. History Today 65. London, UK: History Today Ltd.

Gardner, J., and Oswald, A. J. (2007). Money and mental wellbeing: a longitudinal study of medium-sized lottery wins. *J. Health Econ.* 26, 49–60. doi: 10.1016/j.jhealeco.2006.08.004

Gilam, G., Gross, J. J., Wager, T. D., Keefe, F. J., and Mackey, S. C. (2020). What is the relationship between pain and emotion? Bridging Constructs and Communities. *Neuron* 107, 17–21. doi: 10.1016/j.neuron.2020.05.024

Gómez-Emilsson, A. (2019). Cluster headache frequency follows a long-tail distribution. Available at: Forum.effectivealtruism.org and https://forum.effectivealtruism.org/posts/XLrzi6TrCqhTMtf8d/cluster-headache-frequency-follows-a-long-tail-distribution

Gómez-Emilsson, A. (2022). Just look at the thing! – how the science of consciousness informs ethics. Available at: Forum.effectivealtruism.org and https://forum.effectivealtruism.org/posts/bvtAXefTDQgHxc9BR/just-look-at-the-thing-how-the-science-of-consciousness

Goodwin, G. M., Price, J., De Bodinat, C., and Laredo, J. (2017). Emotional blunting with antidepressant treatments: a survey among depressed patients. *J. Affect. Disord.* 221, 31–35. doi: 10.1016/j.jad.2017.05.048

Gossen, H. (1854). *Die Entwicklung der Gesetze des menschlichen Verkehrs und der daraus fließenden Regeln für menschliches Handeln*. Berlin, DE: Friedrich Vieweg und Sohn.

Goto, N., and Schaefer, A. (2020). "Emotional intensity" in *Encyclopedia of personality and individual differences*. eds. V. Zeigler-Hill, and T. Shackelford (Switzerland: Springer), 1311–1319.

Haan, L., and Ferreira, A. (2007). *Extreme value theory: An introduction*. New York, US: Springer.

Haefeli, M., and Elfering, A. (2005). Pain assessment. *Eur. Spine J.* 15, S17–S24. doi: 10.1007/s00586-005-1044-x

Hagerty, M., Isaacs, J., Brasington, L., Shupe, L., Fetz, E., and Cramer, S. (2013). Case study of ecstatic meditation: fMRI and EEG evidence of self-stimulating a reward system. *Neural Plast.* 2013, 1–12. doi: 10.1155/2013/653572

Helliwell, J. F., Layard, R., Sachs, J. D., Aknin, L. B., De Neve, J.-E., and Wang, S. (2023). *World happiness report 2023. 11th* Edn. Paris, France: Sustainable Development Solutions Network.

Hemmer, P., and Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Top. Cogn. Sci.* 1, 189–202. doi: 10.1111/j.1756-8765.2008.01010.x

HM Treasury. (2021). Wellbeing guidance for appraisal: supplementary green book guidance. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005388/Wellbeing_guidance_for_appraisal_-_supplementary_Green_Book_guidance.pdf

Holz, N., Larrouy-Maestri, P., and Poeppel, D. (2021). The paradoxical role of emotional intensity in the perception of vocal affect. *Sci. Rep.* 11:9663. doi: 10.1038/s41598-021-88431-0

Kaiser, C., and Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proc. Natl. Acad. Sci.* 119:e2210412119. doi: 10.1073/pnas.2210412119

Keltner, D. (2019). Toward a consensual taxonomy of emotions. *Cognit. Emot.* 33, 14–19. doi: 10.1080/02699931.2019.1574397

Kero, P., and Lee, D. (2016). Likert is pronounced "LICK-urt" not "LIE-kurt" and the data are ordinal not interval. *J. Appl. Meas.* 17, 502–509.

Klaus, A., Yu, S., and Plenz, D. (2011). Statistical analyses support power law distributions found in neuronal avalanches. *PLoS One* 6:e19779. doi: 10.1371/journal.pone.0019779

Kollmeier, B., Brand, T., and Meyer, B. (2008). "Perception of speech and sound" in *Springer handbook of speech processing*. eds. J. Benesty, M. Mohan Sondhi and Y. Huang (Heidelberg, DE: Springer).

Larroulet Philippi, C. (2023). Well-being measurements and the linearity assumption: a response to Wodak. *Australas. J. Philos.*, 1–7. doi: 10.1080/00048402.2023.2227642

Larsen, R. J., and Diener, E. (1987). Affect intensity as an individual difference characteristic: a review. *J. Res. Pers.* 21, 1–39. doi: 10.1016/0092-6566(87)90023-7

Lipman, S. A., Brouwer, W. B. F., and Attema, A. E. (2019). QALYs without bias? Nonparametric correction of time trade-off and standard gamble weights based on prospect theory. *Health Econ.* 28, 843–854. doi: 10.1002/hec.3895

Lucas, R. E., Diener, E., and Larsen, R. J. (2003). Measuring positive emotions. In S. J. Lopez and C. R. Snyder (Eds.), *Positive psychological assessment: a handbook of models and measures*. Washington, DC: American Psychological Association 201–218

Mauss, I. B., and Robinson, M. D. (2009). Measures of emotion: a review. *Cognit. Emot.* 23, 209–237. doi: 10.1080/02699930802204677

McAuliffe, W., and Shriver, A. (2022). The relative importance of the severity and duration of pain. Rethink priorities. US: San Francisco. Available at: https://rethinkpriorities.org/publications/the-relative-importance-of-the-severity-and-duration-of-pain

Mischkowski, D., Crocker, J., and Way, B. M. (2019). A social analgesic? Acetaminophen (paracetamol) reduces positive empathy. *Front. Psychol.* 10. doi: 10.3389/fpsyg.2019.00538

Ng, Y. K. (2008). Happiness studies: ways to improve comparability and some public policy implications. *Economic Record* 84, 253–266. doi: 10.1111/j.1475-4932.2008.00466.x

OECD (2013). OECD guidelines on measuring subjective well-being. *OECD*. doi: 10.1787/9789264191655-en

Perez-Truglia, R. (2012). On the causes and consequences of hedonic adaptation. *J. Econ. Psychol.* 33, 1182–1192. doi: 10.1016/j.joep.2012.08.004

Peterson, C. (2018). Stung by a warrier wasp. Available at: https://www.youtube.com/watch?v=unk6n3_QdlM

Phillips, R., Walsh, E., Jensen, T., Nagy, G., Kinard, J., Cernasov, P., et al. (2022). Longitudinal associations between perceived stress and anhedonia during psychotherapy. *MedRxiv.* doi: 10.1101/2022.03.09.22272139

Plant, M. (2019). The measurement of wellbeing. Happier Lives Institute. Available at: https://www.happierlivesinstitute.org/report/the-measurement-of-wellbeing/

Plant, M. (2020). A happy possibility about happiness (and other subjective) scales: an investigation and tentative defence of the cardinality thesis. PhilPapers. Available at: https://philpapers.org/rec/PLAAHP

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., Lamantia, A.-S., Mcnamara, J. O., et al. (2012). *Neuroscience*. Sunderland, MA: Sinauer Associates.

Rivat, C., and Ballantyne, J. (2016). The dark side of opioids in pain management. *PAIN Rep* 1:e570. doi: 10.1097/PR9.0000000000000570

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Samuelsson, C., Dupret, S., Plant, M., and Kaiser, C. (2023) *Can we trust wellbeing surveys? A pilot study of comparability, linearity, and neutrality*. Oxford: Happier Lives Institute.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychol. Rev.* 61, 81–88. doi: 10.1037/h0054570

Silvia, P. J. (2002). Self-awareness and emotional intensity. *Cognit. Emot.* 16, 195–216. doi: 10.1080/02699930143000310

Skowronski, J., Walker, W., Henderson, D., and Bond, D. (2014). "Chapter three - the fading affect Bias: its history, its implications, and its future" in *Advances in experimental social psychology*. eds. J. Olson and M. Zanna, vol. *49* (New York, US: Academic Press), 163–218.

Solomon, R. L., and Corbit, J. D. (1974). An opponent-process theory of motivation: I Temporal dynamics of affect. *Psychol. Rev.* 81, 119–145. doi: 10.1037/h0036128

Teich, M. C., Heneghan, C., Lowen, S. B., Ozaki, T., and Kaplan, E. (1997). Fractal character of the neural spike train in the visual system of the cat. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 14, 529–546. doi: 10.1364/JOSAA.14.000529

Toib, A., Lyakhov, V., and Marom, S. (1998). Interaction between duration of activity and time course of recovery from slow inactivation in mammalian brain Na+ channels. *J. Neurosci. Off. J. Soc. Neurosci.* 18, 1893–1903. doi: 10.1523/JNEUROSCI.18-05-01893.1998

Wang, Q. (2020). The cultural Foundation of Human Memory. *Annu. Rev. Psychol.* 72, 151–179. doi: 10.1146/annurev-psych-070920-023638

What Works Wellbeing. (2017). *A guide to wellbeing economic evaluation*. London: What Works Wellbeing

Wiech, K. (2016). Deconstructing the sensation of pain: the influence of cognitive processes on pain perception. *Science* 354, 584–587. doi: 10.1126/science.aaf8934

Wodak, D. (2019). What if well-being measurements are non-linear? *Australas. J. Philos.* 97, 29–45. doi: 10.1080/00048402.2018.1454483

Wundt, W. (1897). Outlines of psychology. Classics in the history of psychology. The Wayback Machine, York University 2010, Toronto. Available at: http://psychclassics.asu.edu/index.htm

Zimmer, Z., Zajacova, A., Fraser, K., Powers, D., and Grol-Prokopczyk, H. (2023). A global comparative study of wealth-pain gradients: investigating individual-and country-level associations. *Dialogues Health* 2:100122. doi: 10.1016/j.dialog.2023.100122

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 27 November 2023
DOI 10.3389/fpsyg.2023.1306023

Check for updates

OPEN ACCESS

EDITED BY
Luca Simione,
UNINT - Università degli studi Internazionali di Roma, Italy

REVIEWED BY
Larissa Albantakis,
University of Wisconsin-Madison, United States
Robert French,
Oakland Community College, United States

*CORRESPONDENCE
Marius Usher
✉ marius@tauex.tau.ac.il
Niccolò Negro
✉ niccolo.negro.research@gmail.com

RECEIVED 03 October 2023
ACCEPTED 08 November 2023
PUBLISHED 27 November 2023

CITATION
Usher M, Negro N, Jacobson H and Tsuchiya N (2023) When philosophical nuance matters: safeguarding consciousness research from restrictive assumptions.
Front. Psychol. 14:1306023.
doi: 10.3389/fpsyg.2023.1306023

When philosophical nuance matters: safeguarding consciousness research from restrictive assumptions

Marius Usher[1,2]*, Niccolò Negro[1,3]*, Hilla Jacobson[4,5] and Naotsugu Tsuchiya[6,7]

[1]School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel, [2]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel, [3]Monash Centre for Consciousness and Contemplative Studies, Monash University, Melbourne, VIC, Australia, [4]Department of Cognitive and Brain Sciences, Hebrew University of Jerusalem, Jerusalem, Israel, [5]Department of Philosophy, Hebrew University of Jerusalem, Israel, [6]Faculty of Medicine, Nursing, and Health Sciences, School of Psychological Sciences, Monash University, Melbourne, VIC, Australia, [7]Laboratory of Qualia Structure, ATR Computational Neuroscience Laboratories, Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan

In this paper, we revisit the debate surrounding the Unfolding Argument (UA) against causal structure theories of consciousness (as well as the hard-criteria research program it prescribes), using it as a platform for discussing theoretical and methodological issues in consciousness research. Causal structure theories assert that consciousness depends on a particular causal structure of the brain. Our claim is that some of the assumptions fueling the UA are not warranted, and therefore we should reject the methodology for consciousness science that the UA prescribes. First, we briefly survey the most popular philosophical positions in consciousness science, namely physicalism and functionalism. We discuss the relations between these positions and the behaviorist methodology that the UA assumptions express, despite the contrary claim of its proponents. Second, we argue that the same reasoning that the UA applies against causal structure theories can be applied to functionalist approaches, thus proving too much and deeming as unscientific a whole range of (non-causal structure) theories. Since this is overly restrictive and fits poorly with common practice in cognitive neuroscience, we suggest that the reasoning of the UA must be flawed. Third, we assess its philosophical assumptions, which express a restrictive methodology, and conclude that there are reasons to reject them. Finally, we propose a more inclusive methodology for consciousness science, that includes neural, behavioral, and phenomenological evidence (provided by the first-person perspective) without which consciousness science could not even start. Then, we extend this discussion to the scope of consciousness science, and conclude that theories of consciousness should be tested and evaluated on humans, and not on systems considerably different from us. Rather than restricting the methodology of consciousness science, we should, at this point, restrict the range of systems upon which it is supposed to be built.

KEYWORDS

consciousness, unfolding argument, functionalism, recurrency, causal structure, IIT, Blockhead

1 Introduction

Understanding how consciousness relates to the structure and activity of the brain is one of the most challenging tasks of scientific endeavor. Whereas a few decades ago, the subject of consciousness was exclusively philosophical, it has become a major subject of research in neuroscience in the last few decades. What started as a search for neural correlates of consciousness (NCC) (Crick and Koch, 1990) - for a discussion, see (Chalmers, 2000) - has now matured to the development of a multitude of theories that aim to answer the more difficult question of how consciousness can be explained by the organization of brain processes (Dehaene et al., 1998; Lamme, 2006; Tononi et al., 2016; Solms and Friston, 2018; Solms, 2019; Gidon et al., 2022; Seth and Bayne, 2022). These theories are subject to an intensive debate that involves experimental research (Crick and Koch, 1998; Zeki and Bartels, 1998; Landman et al., 2003; Sligte et al., 2008; Aru et al., 2012; de Graaf et al., 2012; Liu et al., 2012; Bronfman et al., 2014; King and Dehaene, 2014; Mudrik et al., 2014; Noy et al., 2015; Josselyn and Tonegawa, 2020; He, 2023), philosophical analysis (Block, 1995, 2011; Chalmers, 1995, 1996; Phillips, 2011, 2016; Cohen et al., 2016; Usher et al., 2018; Bronfman et al., 2019; Ellia et al., 2021; Ellia and Chis-Ciure, 2022; Michel, 2023), as well as clinical/neuropsychological research (Owen et al., 2006; Monti, 2015). These aspects of consciousness research are necessarily intertwined, because all the consciousness theories have specific philosophical starting points and implications.

This interplay between abstract theoretical considerations and experimental research is illustrated by a recent philosophical argument – the *unfolding argument* (UA) – which has been proposed with the aim of prescribing which types of consciousness theories are scientifically valid, prior to empirical testing (Doerig et al., 2019). According to the UA, causal-structure theories, such as the *Integrated Information Theory* (IIT) (Tononi et al., 2016) and the *Recurrent Processing Theory* (RPT) (Lamme, 2006), are either false or unscientific. This reduction of the theory space could be beneficial for consciousness research, as there is currently a proliferation of consciousness theories (for discussions, see Aru et al., 2020; McFadden, 2020; Del Pin et al., 2021; Doerig et al., 2021; Signorelli et al., 2021; Seth and Bayne, 2022). However, the UA also has its own philosophical assumptions, which have come under severe criticism (Kleiner, 2020; Negro, 2020; Tsuchiya et al., 2020; Albantakis, 2020a; Kent and Wittmann, 2021; Kleiner and Hoel, 2021; Mallatt, 2021; Usher, 2021). The authors of the UA have responded to this criticism (Herzog et al., 2022). More recently, Doerig et al. (2021) have expanded the UA into a research program that is meant to set up a set of "hard criteria for empirical theories of consciousness," which not only restricts the type of admissible consciousness theories, but also explicitly prescribes a restrictive methodology for consciousness research.

The aim of this paper is to argue that the UA research program is too restrictive, making explicit the specific points of disagreement, and more generally to show how implicit assumptions can influence our methodological choices in consciousness science. By bringing to the fore a variety of implicit assumptions and reconsidering the relations of extant scientific theories to traditional philosophical positions and arguments regarding the nature of consciousness and the feasibility of its scientific investigation, we hope to make the debate more informative. This is important, given the high price of prematurely abandoning promising classes of theories without empirical testing (see (Melloni et al., 2021, 2023) for promising attempts to test consciousness theories). In doing so, we aim to broaden the discussion to several central issues that are critical to consciousness research, such as (i) the grounding of consciousness theories in functionalism, mind-brain (MB) type-identity, or behaviorism; and (ii) the reality of phenomenal experience and the role of first-person and neural evidence in consciousness science. Finally, we aim to propose an alternative research program for consciousness research, which is bold in its methodology, but somewhat more restrictive in its scope (to account for human consciousness, first). We start with a brief recap of the broad philosophical positions on consciousness and a summary of the UA before we critically examine its soundness.

2 Philosophical positions in consciousness research

Given that many concepts employed in the contemporary science of consciousness derive from philosophy, it is important to have a clear understanding of some influential philosophical frameworks on consciousness. Historically, in Western philosophy the traditional, theory on the relation between mind and matter was *dualism*, the view that mind and matter are fundamentally distinct (Descartes, 1641/1996). There are various versions of dualism, but they are not popular in contemporary consciousness research.[1] Scholars in this field mostly follow *physicalism* (Francken et al., 2022; Bourget and Chalmers, 2023) – the view that consciousness supervenes with metaphysical necessity on the physical.

There are many ways to make the mind–body relation more precise, under a physicalist framework. A first option is behaviorism, which considers the mind as a set of behavioral dispositions. Historically, behaviorism has been presented either as logical behaviorism, the view that mental terms can be conceptually reduced, via *a priori* analysis, to behavioral terms (Ryle, 1949), or as methodological behaviorism, which was motivated by the drive to base psychology on firm scientific grounds by focusing on purely outer and publicly observable phenomena (Watson, 1913); reprinted in Watson (1994). Despite being quite influential in the past, behaviorism is now widely accepted to be deficient as an account of mental states and processes [see textbook discussions in (Braddon-Mitchell and Jackson, 2007) and (Bayne, 2021)]. We thus focus on two more influential physicalist theories, which both play a central role in the debate surrounding the UA against causal-structure theories: (i) the *mind-brain (MB) type identity theory* (Place, 1956; Smart, 1959); and (ii) *functionalism* (Putnam, 1967).

1. **MB-type-identity.** According to type-identity theorists, types of mental states are types of physical (brain) states, in the same

1 Most present-day dualists adhere to property (rather than substance) dualism, according to which mental properties exist "over and above" physical properties. An important distinction is between interactionist dualism and epiphenomenalism. *Interactionist dualism* (Descartes, 1641/1996) comes under conflict with the entrenched principle of the *causal closure of the physical laws*. *Epiphenomenalism,* on the other hand, is unattractive because, among other things, it does not allow for an evolutionary account of consciousness (Braddon-Mitchell and Jackson, 2007).

way as *water is* H_2O (or *clouds are vapor*; Place, 1956; Smart, 1959). Conscious states such as pain are thought to be identical to specific types of brain states (e.g., a particular type of *cortical-thalamic neural oscillation*). According to the type-identity theory, the identity of conscious mental states is determined by their physical constitution.

2. **Functionalism**. Originally, functionalism has been developed in opposition to both MB-identity theory and behaviorism, and was motivated by the *multiple realizability* argument (Putnam, 1967), which asserted that it is unlikely that all mental states (or processes) of the same type (e.g., pain or the desire to drink water) are always realized, and moreover *must* be realized, by the same type of brain states (or processes). According to *functionalism,* it is not the material constitution of mental states that determines their identity; instead, it is the role they play in the cognitive system of which they are a part. This independence of functional roles from their substrate can be expressed in different ways. A distinction, which will play a role in our later discussion, can be drawn between the material properties of the substrate (e.g., whether it is made of carbon or silicon) and the structural properties of the substrate (e.g., its *network* connectivity). A theory can be substrate-independent with respect to the material properties of the substrate without being independent with respect to its structural/network properties. More specifically, according to functionalism, the identity of a mental state, such as pain, is determined by its causal relations to sensory inputs, behavioral outputs, and, importantly, to *other mental states*. This focus on the relations among internal mental states marks the main difference between *functionalism* and behaviorism, which conceives of mental properties as behavioral dispositions (Fodor, 1981; Braddon-Mitchell and Jackson, 2007) independent of transitions among internal states that mediate between stimulus and responses. This difference between functionalism and behaviorism is eloquently illustrated in a review by Fodor (1981), portraying the main difference as follows: "According to logical behaviorism,[2] it is a necessary truth that any system that has our stimulus–response contingencies also has our headaches" (Fodor, 1981, p. 118). For functionalism, on the other hand, mental states (e.g., headaches) are determined by their place in the cognitive algorithm that generates the stimulus–response contingencies. This can be applied to phenomenal experiences: for a functionalist, the phenomenal character of a mental state depends on the cognitive algorithm in which that state plays a role; that is, consciousness is reducible to the functional and relational profile a mental state bears to stimuli, behavior, and other mental states. Although there are many versions of functionalism (see (Braddon-Mitchell and Jackson, 2007) for textbook discussion), this introduction will be enough for the present purposes.

Consciousness is considered to pose a problem for all physicalist theories (see, e.g., Jackson, 1982), but there are also special difficulties that it is thought to pose for functionalism. In particular, while functionalism has been seen as a very successful approach in cognitive research[3] (Block and Fodor, 1972), it has been attacked as an account to consciousness by two lines of argument, which stem from the idea that phenomenal properties are *intrinsic* properties, and cannot be fully captured by the *relational,* functional, properties functionalism focuses on. A prominent example of this line of challenge is provided by the inverted-spectrum (or qualia) arguments (Block and Fodor, 1972; Shoemaker, 1975, 1982; Palmer, 1999), which aim to show that two systems can have mental networks with the same functional profile, while having different (more precisely, inverted) experiences, or the converse (Block, 1990). A second anti-functionalist objection is *the absent qualia argument* (Block, 1978). Perhaps one of the sharpest attacks on functionalism as a theory of consciousness (belonging to the second class of anti-functionalist objections) is Ned Block's *China-Brain* (Block, 1978), which asks us to consider a simulation of a human brain, in which all neurons in that brain are replaced by a large set of people (he offers the population of China to the task), with all communication between the neurons replaced by telephone communication between the people in this population. Supposing the China population is linked to sensory and motor organs in the same way as the original person's brain, this (China-brain) simulation will produce the same behavior as a normal person (perhaps in slow motion).

The intuition that Block appeals to is that while we readily accept that the original person has, e.g., gustatory phenomenal experiences, when she consumes a chocolate ice cream, we feel reluctant to accept the same for the China population. While this argument is not conclusive – it has not persuaded most functionalists (see (Braddon-Mitchell and Jackson, 2007) for a textbook discussion), who can insist that the simulation (i.e., the population of China in this example) has the same experiences as the simulated person – we wish to mark this as a central argument, since, as we will see, it has much in common with the UA, to which we turn next.

3 The unfolding argument

The unfolding argument (UA) was proposed to refute a large set of consciousness theories called *causal-structure* theories[4], namely theories that hold that consciousness depends on the causal structure of the brain. Two prominent theories of this sort are the integrated

2 Although this passage focuses on logical behaviorism, both logical and methodological variants of behaviorism are at odds with functionalism, with respect to how mental states are determined.

3 One may consider cognitive psychology as a very successful functionalist project, which transcended methodological behaviorism and folk-psychology, by relying on experimental manipulations of sensory inputs and observing behaviors (including reaction-time and eye-movements) to infer internal states and processes, such as memory, goals and attention processes, or cognitive inferences, which mediate complex behaviors.

4 Note that while Doerig et al. (2019) focused particularly on IIT, the UA targets all *causal-structure* theories and not only IIT. We also wish to clarify that we are agnostic on whether IIT and RPT are successful theories or not. We hold, however, that they are both valid hypotheses that should not be dismissed without empirical tests, on purely theoretical grounds.

information theory (*IIT*) (Oizumi et al., 2014; Tononi and Koch, 2015; Albantakis et al., 2023) and the recurrent processing theory (*RPT*) (Lamme, 2006), which both assume that consciousness depends on the presence of recurrent brain connectivity. Categorizing IIT and RPT in relation to traditional philosophical positions is not straightforward (Tononi and Koch, 2015; Tononi, 2017; Grasso, 2019; Cea, 2020; Negro, 2022; Tononi et al., 2022). According to both theories, consciousness depends on the causal structure of the brain, but both theories hold that consciousness may be multiply realizable, and allow that it can be realized in non-biological systems, as long as those systems have the *abstract* (i.e., independent of the specific and fine-grain biological details) network-structure that the theories associate with consciousness. This alignment with multiple realizability may suggest that IIT and RPT are compatible with functionalism[5]. Interestingly, despite the UA-proponents arguing to endorse a functionalist approach, they still consider these models as false or outside the range of science. This seems to be because UA proponents appear to consider functional characteristics at a lower level of resolution (i.e., less sensitive to specific properties of the system), namely at the behavioral rather than at a network level (which is already relatively abstract). For example, a functional characteristic, such as network-recurrency, which IIT and RPT assume necessary for consciousness, is considered by the UA-proponents as an *implementation* detail that can be multiply realized by a system without network-recurrency (i.e., a feedforward network), as long as the latter is behaviorally equivalent to the original (recurrent network) in consciousness experiments. We thus come to a somewhat paradoxical situation, in which an argument that is framed to support functionalism, rules out (as false or outside the range of science), on purely theoretical grounds, a class of models that could be compatible with functionalism under some interpretations. In the next section we shall examine whether the UA supports, or even coheres with functionalism in general.

In particular, the UA proponents claim that functional properties can be behaviorally measured from the third-person perspective without assuming that a particular brain architecture (e.g., recurrent vs. feedforward) determine consciousness in the first place, and therefore only a theory that associates consciousness with functional/behavioral properties can be confirmed or falsified through behavioral research. In particular, they assert that "consciousness must be described in terms of what it does, and not how it does it" (Doerig et al., 2019, p. 56). According to the UA, for any conscious (recurrent) brain that mediates behavior in a consciousness experiment, we can construct a brain that replaces the recurrent (RN) with feed-forward networks (FFN), that is behaviorally equivalent to the original brain. Therefore, as no behavioral experiment testing consciousness can distinguish between such brain variants, the UA concludes that all theories that assume consciousness to depend on a certain causal structure (e.g., recurrent vs. feed-forward) are either false or unfalsifiable.

The UA has the following form:

> "(P1): In science we rely on physical measurements (based on subjective reports about consciousness).
>
> (P2): For any recurrent system with a given input–output function, there exist feedforward systems with the same input–output function (and vice-versa).
>
> (P3): Two systems that have identical input–output functions cannot be distinguished by any experiment that relies on a physical measurement (other than a measurement of brain activity itself or of other internal workings of the system).
>
> (P4): We cannot use measures of brain activity as a-priori indicators of consciousness, because the brain basis of consciousness is what we are trying to understand in the first place.
>
> (C): Therefore, EITHER causal structure theories are falsified (if they accept that unfolded, feedforward networks can be conscious), OR causal structure theories are outside the realm of scientific inquiry (if they maintain that unfolded feedforward networks are not conscious despite being empirically indistinguishable from functionally equivalent recurrent networks)" (Doerig et al., 2019, p. 53).

This conclusion rules out causal-structure theories from consciousness science, without the need to test them on their ability to account for data. A further UA-variant has been proposed, which replaces P2 (the behavioral equivalence between RN and FFN) with the behavioral equivalence between a physical system and its computer simulation (Herzog et al., 2022). Accordingly, instead of building a behaviorally equivalent FF-robot, we can create a robot controlled by a computer simulation of a real brain. Herzog et al. (2022) argue that, since we are typically running our computer simulations of RNs on serial computers, such a robot will be indistinguishable from the actual person, regarding any consciousness test. And therefore, we would not have any reason to deem one system as conscious and the other as non-conscious.

We believe this conclusion is premature and that both variants of the UA are unsound. Still, before we turn to our counterarguments, we wish to note that the simulation version is similar to the simulation created by Block's China-Brain argument in invoking two functionally identical systems in order to show, on *a priori* grounds, that a particular approach to consciousness is invalid. Yet, the two arguments suggest opposite conclusions – the UA was intended to cohere with functionalist approaches, whereas the China-Brain was intended to undermine functionalism. Even if the China-Brain argument is

5 This needs to be qualified in relation to the distinction between substrate and network invariance. Causal structure theories (like IIT and RPT) might be considered as functionalist from the perspective of substrate-invariance, but not from the perspective of network-invariance as they hold that consciousness depends on the network structure but not on its material composition (e.g., carbon vs. silicon). One potential interpretation of causal structure theories is that they require algorithmic rather than implementational recurrence (Butlin et al., 2023, p 21). However, this faces the problem that the algorithmic level is just not well defined, in general. The assumption that neurons count as the level of implementation is adhoc. More specifically, IIT requires network (rather than algorithmic) recurrence, which might still be seen as a functional property but at a higher resolution that just behavior (see next section).

inconclusive, we believe it would be surprising if this opposite argument were accepted as conclusive.

To better understand this complex dialectic, we must focus on the exact relation between the UA and functionalism. In the next section, we point out that the UA-rationale can lead to a stronger argument that rules out as invalid not only causal-structure scientific theories of consciousness, but also (and perhaps contrary to the original motivation) functionalist theories.

4 An UA-type argument against functionalism (and cognitive science)

In section two, we have surveyed the essential distinctions between functionalism and behaviorism. In their response to critics, Herzog et al. (2022) have clarified that their position is *functionalist* rather than *behaviorist*.[6] We accept that they are motivated to account for consciousness based on latent processes of the organism and that this is consistent with functionalism, which is a productive framework in cognitive science. However, we believe that some of the assumptions of the UA are not consistent with functionalism, but rather point in a direction closer to behaviorism. This is because (as we will shortly argue, in subsections 4.1 and 4.2) these assumptions together with the basic rationale of the argument and a plausible assumption regarding the indeterminacy of algorithm by behavior lead to the conclusion that theories according to which consciousness is determined by the system's algorithms, or information processing, are just as problematic as causal structure theories. According to behaviorism, mental properties are exclusively determined by behavioral dispositions, while according to functionalism, mental properties are determined by the algorithm that underlies behavior (Putnam, 1967; Fodor, 1981). Furthermore, functionalism regards the causal relations among the system's internal mental states as crucial to their identity, which is closer to the causal structural theories' claim. This dissociation between behavior and cognitive algorithm was clearly illustrated in the *Blockhead*-argument (Block, 1981), in which we are provided with a dissociation between behavior and algorithm.

The argument compares a person who acts as a result of information processing with a robot, called Blockhead, that acts the same as a person, as a result of inspecting a large look-up table that contains an extensive list of behaviors that ordinary people are likely to provide in response to possible questions (Block, 1981). Functionalists have accepted this dissociation, as an illustration of a functionalist thesis that Block calls *psychologism*, according to which mental properties such as intelligence depend on the character of the internal information processing that produces the relevant behavior and not on the input–output behavior alone. As formulated by Block, *psychologism* is the doctrine that:

> "Two systems could have actual and potential behavior typical of familiar intelligent beings, that the two systems could be exactly alike in their actual and potential behavior, and in their behavioral dispositions and capacities and counterfactual behavioral properties (*i.e.*,, what behaviors, behavioral dispositions, and behavioral capacities they would have exhibited had their stimuli differed) – the two systems could be alike in all these ways, yet there could be a difference in the information processing that mediates their stimuli and responses that determines that one is not at all intelligent while the other is fully intelligent" (Block, 1981, p. 5).

Critically, we will argue that insisting that the only available resource in trying to account for consciousness is input–output behavior (P3-P4) conflicts with this principle and seems more consistent with *behaviorism* than with *functionalism*. At a first pass, the position of UA proponents brings to mind *methodological* behaviorism and is silent about accounts of the metaphysics of consciousness, since it is possible to hold that consciousness is constituted by the functional profile of a physical system while maintaining that the functional profile itself can be detected by looking only at behavioral responses. However, we will attempt to show that (at least in conjunction with the other premises of the UA) the methodology the UA prescribes, leads to the conclusion that both MB type-identity theories and functionalist theories lie outside the realm of scientific inquiry. If this is the case, then in present context, the methodological thesis (P1, P3-4) implies that no metaphysical theory that attempts to uncover the internal underpinning (whether neural or functional) of external input–output patterns is within our reach. We argue that this way of addressing mental phenomena is an overly austere (and restrictive) methodology that ignores actual practice in the cognitive sciences.

A similar analysis of the UA argument has been recently presented by Kleiner (2020); see also (Kleiner and Hoel, 2021), who concluded that on the basis of the UA premises one could rule out (as false or unscientific) any functionalist theory of consciousness[7], including theories such as the Global-Workspace (Dehaene et al., 1998). This is because (following the UA rationale), one cannot distinguish (on the basis of input–output functions) between a system that is driven by a global-workspace and one driven by a lookup table of it. In their reply to Kleiner (2020), Herzog et al. (2022) argue that:

> "if the workspace is defined in functional terms, then the lookup table also realizes a global workspace. Contrary to causal structure, there is no mathematical theorem stating that the same i/o functions can be realized with and without a global workspace (see also Ganesh, 2020). In summary, we agree that Kleiner's argument applies to theories that identify consciousness with a certain non-functional process claimed to be necessary and sufficient as, indeed, many theories do (Doerig et al., 2021). However, theories may be cast in functional terms, or propose that consciousness should not simply be identified with a single

6 Herzog et al. (2022) state that "A behaviorist would claim that internal states are useless to understand the mind, if they would use the word "mind" at all [...]. In contrast, we take subjective states seriously and assume that we can learn about them through i/o observations. We cannot measure consciousness directly, but we can measure subjective reports (verbal or otherwise) and link them to brain activity" (p. 3).

7 More precisely, this conclusion applies to any model of consciousness that depends non-trivially on physical systems [see definition 2.8, and Lemma 2.9, in Kleiner, (2020)].

process, just as life is not identified with a single process (Machery, 2012)." (Herzog et al., 2022, p. 10).

We believe that there is an important ambiguity in this statement that is critical to the differences between behaviorism and functionalism. Indeed, functionalism allows for the possibility that mental properties (e.g., pains) do not uniquely determine the neural structure or processes. However, it still requires them to be uniquely associated with the functional algorithm, which generates behavior (Fodor, 1981). The question is in what sense the global workspace and its look-up table are functionally identical? If the difference between these two systems is to be found at the neural level, one may suggest (as we interpret Herzog et al. to argue) that we have multiple neural processes that implement the same global-workspace algorithm, all of which are associated with consciousness. However, this functionalist solution becomes out of reach if we have two different *cognitive algorithms* that mediate the same input–output contingency: a global workspace and its lookup table. It is not clear to us on what basis it would be the case that the Global-Workspace and its lookup-table are identical, *qua algorithm*. The problem is that if the global workspace and its lookup table are not identical, *qua* cognitive algorithms, then, at best, they can be equivalent only in terms of outer behavioral dispositions. And if this is the level at which consciousness should be investigated, then a version of behaviorism follows. This result, we believe, would exclude not only causal structure theories from consciousness science, but also functionalist theories. This result is possible only if a dissociation between behavior and the cognitive algorithm were possible. In the next section, we expand on why such an indeterminacy (similar to the one suggested by the UA between behavior and neural structure) is likely to manifest between behavior and cognitive algorithm, under the restrictive methodology advocated by the UA.

4.1 The behavior-algorithm indeterminacy

In the previous sub-section, we have highlighted that if multiple cognitive algorithms can determine the same behavior, then a similar rationale as that of the UA could also apply to functionalist theories of consciousness. In this sub-section, we unpack this claim by beginning with the traditional distinction among different levels of describing cognitive systems[8].

Level 1 ("behavioral level"): Describes the input–output function of the system – i.e., its actual and potential behavior and responses to any possible stimuli. In this context, the inputs and outputs are mathematical values (that can also represent properties in the physical world). The I-O function describes the "problem" to be computed.

Level 2 ("functional-algorithmic level"): Describes the algorithm by means of which the Input–Output function of level 1 is being computed. It describes the specific information processes by means of which the system solves the problem (of how to achieve the outer behavior).

Level 3 ("physical-implementation level"): Describes the physical structure that implements the algorithm of level 2.

Now, return to the UA. The argument crucially appeals to the assumption that the behavioral level does not determine the causal structure, namely the physical implementation level: the exact same I/O function can characterize systems with multiple causal structures. This is *P2* – "for any recurrent system with a given I/O function, there exist feedforward systems with the same input–output function" (Doerig et al., 2019, p. 53) – generalized to all causal structure theories (as it should, if it is to prove its conclusion regarding all causal-structure theories). The proponents of the UA argue that I/O functions provide the primary evidence for scientific theories of consciousness. Assuming this, they conclude that causal structures are not the right place to look for consciousness, if we are after a *scientific* explanation of consciousness. However, just as it can be argued that the behavioral level cannot determine causal structure, so it can be argued that, likewise, it cannot determine the functional-algorithmic level.[9] In terms of the triple distinction above, level 1 may not determine both *level 3 and level 2*. We thus formulate P2* in this way: each I/O function can be computed by many different algorithms, just as each algorithm can be realized by different physical structures.

The most straightforward illustration of the claim that different algorithms can compute any I/O function is provided by the above-mentioned "Blockhead thought experiment" (Block, 1981). The Blockhead's and a real person's algorithms are drastically different, even if they produce the same behaviors - so, they are Level 1-equivalent, but not Level 2-equivalent. The real person carries a variety of cognitive processes (such as mental inferences, goal directed memory search, value-estimations, etc.), while Blockhead only searches its lookup-table and selects the first possible response. The Blockhead system, however, can compute only functions that range over a finite number of input-arguments. Here, then, are other examples that extend this (finite) limitation. The first only illustrates the rationale underlying the (unlimited range) claim.

Take any algorithm that receives an input x and outputs f(x). Start by adding two steps before you start the original algorithm: first, add n to x (Step 1: $y = x + n$). Then subtract n from the result (Step 2: $z = y - n$). You are back at x. Step 3, continue with the original algorithm. Since this can be done with any n, we have infinitely many algorithms for the same I/O function. The second illustration concerns the highly instrumental sorting algorithms – ones that arrange elements of a list in a particular order (e.g., from highest to lowest).[10] Importantly, there

8 This tripartite distinction is inspired by Marr's three levels of analysis Marr (1982). Marr called level one "the computational level", but his characterization of this level makes it clear that it is concerned with the abstract specification of behavior. As Love (2015) puts it: "The nature of the computing device (i.e., implementation level) and how the computation is carried out (i.e., the algorithmic level) are irrelevant at this level of analysis. The sole concern of Marr's computational level is the abstract problem description, which consists of detailing the input–output relationships.

9 We believe that both of these statements need to be qualified. The indeterminacies above are likely to apply when one examines simple functions from one space to another, $Y = f(X)$ (say, like in a categorization task that requires subjects to name visual objects). It is more doubtful that they apply in the case in which the transformation applies to temporal entities, which impose restrictions on the temporal duration of the transformation.

10 To mention but two simple examples, the *Quick-Sort* algorithm applies a divide and conquer strategy to divide a list into sub-lists: pick an element, called a pivot, from the list. Reorder the list: locating all elements with values less than the pivot before the pivot and all elements with values greater than

is a mathematical proof that *any* computable function can be computed by different algorithms[11] (Miller, 2014). This brief discussion substantiates our claim that two systems that are indistinguishable with respect to Level 1 (I/O behavior) can be different with respect to Level 2 (algorithm): behavior is not sufficient for determining cognitive functions [see Albantakis (2020b) for a vivid illustration of this idea, showing that the same behavioral function can be computed by multiple algorithms, each involving a different number of conscious entities].

4.2 Upshot: the argument proves too much

The upshot of the reasoning presented above is as follows. If the only available data is behavioral (as characterized by inputs and outputs), then it may be impossible (under restrictive conditions, see below) to differentiate between different physical-implementation theories: there can be multiple theories that explain sensory inputs and behavioral outputs equally well. However, under similarly restrictive conditions, it may also be impossible to differentiate between theories concerning the algorithmic (cognitive) processes that generate the behavior. The indeterminacy of causal structure by input–output functions expressed by the original (P2) may plausibly apply to the level above it (level 2), as in both cases different theories – specifying different physical structures and different algorithms, respectively – can underlie the same input–output functions (P2*). Generalizing the rationale of the UA, not only causal structure theories, but also theories attempting to uncover the specific information processes, algorithms or computations that underlie the input–output functions that characterize familiar conscious systems, would then be deemed invalid or unscientific.[12] This conclusion is undoubtedly too strong and must be rejected, because it would drastically reduce the number of viable and productive approaches to study consciousness. Given that progress in the field has clearly been made, this is an unwarranted conclusion. Hence, the rationale that leads to it must be rejected. And since this rationale is similar to that employed by the original UA (in deducing that a group of theories is unscientific by appealing to an indeterminacy assumption), the UA itself should be rejected.

Let us be more explicit about why the relevant conclusion is untenable. It directly results (based on purely *a priori* grounds) in a refutation of functionalism as a valid approach to consciousness (see also Kleiner, 2020). The methodology suggested by the UA is thus closer to behaviorism, as grounded on the assumption that consciousness is whatever results in behavior obtained in consciousness experiments. This conclusion (ruling out functionalist theories as non-scientific), implied by the very rationale of the UA and reflected by the methodology its proponents suggest, contradicts the viewpoint of UA proponents themselves. Moreover, it appears that significant progress has been obtained in cognitive science, demonstrating that the indeterminacy between behavior and algorithm, can be resolved by relying on less restrictive methods and thus progressing beyond the "observable data" to infer the underlying entities and processes. Therefore, the UA – or its underlying rationale – proves too much and there must be something wrong with it. In what follows, we point out a few weaknesses in the UA reasoning.

5 Examining the philosophical premises of the UA (P1, P3, P4)

5.1 The scientific significance of the first-person perspective

We will examine the UA-premises to understand what must go wrong in the argument we just presented. If even one of these premises is false, the UA-conclusion that causal structure theories of consciousness are false or unfalsifiable would be undermined. Similarly, the conclusion that functionalist theories of consciousness are unscientific would be undermined by negating any of the P1, P2*, P3, P4 statements.

Here we focus on the more philosophical (P1, P3, P4) premises, for two reasons. First, while we also reject P2 – the robust behavioral equivalence of any RNN to an FFN; see (Usher, 2021), this premise can be replaced with a simulation version (which is somewhat less controversial), but leads to an equally puzzling conclusion (see China Brain). Second, as the debate about P2 is somewhat technical[13],

the pivot after it (equal values can go either way). The sub-arrays are then sorted recursively. The even simpler *Bubble-Sort* algorithm, in contrast, works by repeatedly stepping through the list, comparing each pair of adjacent items, and swapping them if they are in the wrong order, until no swaps are needed.

11 If one considers a program that computes x+n-n rather than x as a different algorithm, then we have a proof already. However, one may object that an optimizing compiler, which operates on this program, would replace the line x+n-n by just x. The question can thus be reformulated as: "Do we have for each function, a single algorithm that cannot be reduced by any optimizing compiler to the same program (i.e., which are not equivalent in any sense other that they represent the same function)?." This question can be expressed by Turing machines. There are two important results: (i) For every Turing machine there is an equivalent (but not identical) Turing machine; (ii) There is no program that can decide whether any two Turing machines are equivalent. Therefore, as a compiler is a program itself, it is not possible that the compiler can reduce one program of the function to another one, or recognize them as equivalent, for any two equivalent programs [for a review, see (Miller, 2014)].

12 This is because for any algorithm that underlies conscious behavior one may construct a behaviorally equivalent different algorithm (like in Blockhead), and then argue that attributing consciousness to one but not the other is either false or untestable. Contrary to what Herzog et al. (2022) claim, there are mathematical theorems showing there are multiple algorithms that can mediate the same input output function.

13 To our understanding the UA concludes that *since both FFN and RN are universal approximators they can approximate each other.* We believe that this conclusion is based on a logical fallacy. From the fact that FFN and RN are approximators of different entities (an FFN approximates functions, while an RN approximates dynamical systems), it does not follow that they approximate each other. First, the output of an RN cannot be characterized as a function of its input alone (which is what FFNs can achieve), since it also depends on the state of the network itself. This is a critical property of RNs, allowing them to exhibit dynamic properties such as self-sustained activation states without any input and to manifest the dynamical property of hysteresis. This is particularly important for our aims, because many conscious states (e.g.,

we focus here on the philosophical/methodological premises that are more central to the theme of this research topic.

A number of philosophical and methodological criticisms of the UA were also put forward (Albantakis, 2020a; Negro, 2020; Tsuchiya et al., 2020), raising the objection that the UA ignores the relevance of first-person experience in consciousness research (P1). After all, if we want to account for the relation between phenomenal experiences and brain processes, "how it feels to be a conscious subject" is the property of interest (Nagel, 1974). Though eventually interested in phenomenal consciousness, many researchers prefer not to tackle directly phenomenal consciousness, and focus instead on functional aspects, such as *access* [see (Block, 1995, 2011; Dehaene, 2014)]. This risks neglecting what many believe to be the real explanatory target (and the most challenging and fascinating aspect) of consciousness science (Chalmers, 1995; Block, 2002; Ellia et al., 2021).

First, the UA prescribes a methodology for the science of consciousness that excludes the use of first-person data (P1). Instead, Herzog et al. (2022) insist that publicly available data (i.e., objective measurements usually coming from experimental results of consciousness experiments) must be the only source of evidence for consciousness science. According to them, this methodology does not dismiss first-person phenomenological observations, but requires "transforming" them into public data via introspective reports. Only at that point can first-person data, transformed into behavioral evidence, be mapped onto neural evidence. Second, while the UA does not preclude brain measurements in consciousness research, it prescribes such measurements to be carried out only in a second stage (P4), once the conditions for presence of consciousness are established exclusively by behavioral reports (P3-P4; Doerig et al., 2019; Herzog et al., 2022). According to P4, relying on neural measurements to determine consciousness properties leads to *circularity.*

Here, we argue that this methodology, which excludes first-person data, or requires re-interpreting them as third-person data for consciousness science, and which defers neural data to a later stage, is overly *restrictive*, at least when we focus on human-consciousness (we defer to the Discussion section for a distinction between consciousness in humans and in general). Obviously, neither side of the debate denies the importance of behavioral data, nor the importance of neural data. The disagreement stems from how much evidential weight different scholars put on different types of data, and particularly on phenomenological (i.e., first-person) data. While the UA proponents claim that input–output (i.e., behavioral) data are the primary evidence for consciousness science and that purely phenomenological data, which are not translated into some public marker, are not scientific data at all, we maintain that behavioral data like introspective reports can only be valuable heuristics to be used in experimental settings, but they cannot be taken at face value. Furthermore, we highlight that the validity of third-person methods in consciousness science is grounded on first-person data to begin with (see also Ellia et al., 2021).

To illustrate this point, we can resort to the inferences we are licensed to draw in no-report paradigms[14] (Tsuchiya et al., 2015; Koch et al., 2016); see also (Overgaard and Fazekas, 2016) and (Block, 2019) for discussions. In their reply to Tsuchiya et al. (2020), Herzog et al. (2022) argued that even in these cases, we associate the presence of a conscious state with some sort of public measurement, such as optokinetic nystagmus or other physiological measurements. However, this is not always necessary. Suppose we present a non-masked, isolated, supra-threshold stimulus to an awake subject who attends to it. In that case, we can reasonably infer the subject will be conscious of the stimulus without relying on any sort of public measurement. The justification for this inference is provided by the fact that we are certain that *we* would be conscious of the stimulus in that condition, had we been in the subject's place. This inference is facilitated by the assumption that our brain is similar enough so that we should end up with similar visual experiences when viewing the same object under similar conditions. Thus, first-person data can guide and constrain our inferences about other people's experiences, and in this sense, they constitute an indispensable tool for consciousness research. This does not mean that first-person data must be the only evidence for consciousness science. In itself, the fact that an awake subject is conscious of an unmasked supra-threshold stimulus provides limited information on how consciousness relates to brain processes. But the no-report paradigms can be used to disentangle neural correlates of consciousness from that of reports and to eliminate confounding factors related to post-perceptual processes underpinning cognitive accessibility (Block, 2019; Malach, 2022). In fact, the reliance on phenomenal experience was one of the essential ingredients of the method of early psychophysicists (e.g., Fechner), who aimed to uncover laws that map the relation between the intensity of subjective experiences and objective aspects of the environment (see Ellia et al., 2021).

The role of first-person data for consciousness science can also be appreciated by focusing on the inferential reasoning that justifies our attribution of consciousness to subjects in standard experimental settings: we infer that a subject is conscious of the stimulus because we have a series of observations (e.g., the stimulus presentation, the subject looking at the screen, and so on) and some background knowledge that links those observations to consciousness. But crucially, if the information that *I* would be conscious of the stimulus if *I* were in the subject's position was not part of that background knowledge, I would not be justified in inferring that the subject is conscious of the stimulus. Thus, first-person experiences can provide part of the justificatory ground for attributing conscious states to other people. In experimental settings, they can be used to justify inferences about the conscious states of tested subjects.

dreams) depend on this sort of self-sustained activation states that are possible only for RNs, and thus appear (in the absence of input) outside the reach of FFNs. Second, there are independent reasons to reject P2: research within the field of neural computing (Siegelmann and Sontag, 1995; Cabessa and Siegelmann, 2012) indicates that FFN and RN differ drastically in terms of their computational power: while the former are far below Turing-computation, the latter can exceed it (see (Ruffini et al., 2022) for a recent and detailed discussion; see also Usher (2021) for an illustration showing that apparently equivalent RN and FFN, are not equivalent when tested with perturbations, and replies by Herzog et al., 2022).

14 In such paradigms the conscious status of the subject is inferred without requiring a verbal report (Tsuchiya et al., 2015; Koch et al., 2016). These are paradigms that seek to distil the confounds of neural correlates of reports from the true correlates of phenomenal consciousness (Aru et al., 2012). For example, a no-report design for binocular rivalry replaces the verbal report with the monitoring of the optokinetic nystagmus (that tracks the eyes' movement).

So far, we have focused on no-report paradigms to show that “first-person experience,” not transformed into public data/report, is relevant to consciousness science even when behavioral evidence is scarce or non-available. This means that first-person data do not necessarily need to be “transformed” into publicly available data to be of any use to consciousness science: even as phenomenological data, they are legitimate scientific tools[15]. We thus claim that verbal/behavioral reports (which are considered primary in the UA-rationale), only provide us with reasonable evidence of (first-person) experience via an inference to the best explanation, or via analogy, under “*appropriate-circumstances*.” We believe that minimal conditions of such “appropriate circumstances” should include the following epistemological conditions:

i. We have no grounds to suspect that the person is lying or concealing information.
ii. The mechanisms connecting experience to output systems must be reliable (i.e., not “damaged” - so that they can relay experience).
iii. We know that the system is similar to us along relevant dimensions, like brain structures and information processes.

Condition (i) is violated if we have grounds to suspect that a person conceals information. In such a case we are justified not to take her verbal introspection reports at face value. Condition (ii) is violated in certain disorders of consciousness, when a person can be conscious, but unable to express their conscious states in the form of outputs because the output pathways are damaged (Owen et al., 2006; Monti, 2015).

Here, we focus on condition (iii), which allows us to see an important shortcoming of the UA. The difference between the FF-robot and the RNN-robot is that the latter is similar to humans, along the dimension of the brain causal structure, whereas the former is not. Moreover, we know that human brain structures that implement FF-like computation, like the cerebellum, have no particular role in constituting human consciousness (Massimini and Tononi, 2018).

Similarity with humans is important because human beings are the only creatures whose consciousness we are certain of. Indeed, Albantakis (2020a) argues that based on inference to the best (or at least, good enough) explanation, we have little reason to maintain that the FF-robot is conscious. This is because scientific investigation on human consciousness provides evidence that recurrence is necessary for human consciousness (Lamme and Roelfsema, 2000; Pitts et al., 2014; Koch et al., 2016) while there is no evidence of FF structures being relevant to human consciousness. Accordingly, we are licensed to put lower credence in the hypothesis that the FF-robot is conscious because it coheres poorly with neuroscientific background knowledge, derived from studies of *human* consciousness (in the Discussion section we will focus on what this stance implies for what we believe to be the *scope* of consciousness research).

15 We acknowledge the existence of views according to which there is no reality to phenomenal experience, beyond what is expressed in overt judgments or in intentional content (Dennett, 1990). In this paper we address only views that accept the reality of conscious experience (see also Ellia et al., 2021).

To illustrate how this approach, which combines behavioral with phenomenological and neural data, is not circular (and also falsifiable), we apply it to testing the recurrence-hypothesis in human consciousness. This hypothesis, endorsed by IIT (Albantakis et al., 2023) and RP theorists (Lamme, 2006) and denied by the UA authors, states that recurrent processing is necessary for human consciousness. According to P4, making consciousness dependent on neural measures such as recurrent processes is circular. Here, we follow Kleiner (2020) and Kleiner and Hoel (2021) in arguing that using neural markers of consciousness does not require endorsing the theory we aim to test. For example, a theory of human consciousness may predict that a certain type of causal network in a particular state corresponds to a particular conscious experience. We can now carry out experiments to test such predictions by applying TMS (or some other intervention, say, optogenetics or pharmacology) that transiently stimulates/disrupts the recurrent processes in a specific brain area (see (Michel and Malach, 2022) for a similar idea). We do not need to rely on that theory again to interpret the predicted effect. Instead, we carry the experiments on ourselves (or on other volunteers) to determine if we *feel* a difference in experience [for the volunteers, we can rely on their introspective reports, provided that the above-mentioned appropriate circumstances occur; see also (Ruffini et al., 2022)]. Obviously, such experiments, ideally employing a multitude of methods, can provide converging evidence that either increases or decreases our degree of belief in the hypothesis under study (note that we qualify this to human consciousness), and we do not need to assume a theory to arrive at this result. In this scenario, we can either diminish or increase support for the claim that recurrence is necessary for human consciousness.[16] So, the charge of circularity does not apply.

Finally, we wish to suggest a more ecumenical methodology for consciousness science, which acknowledges the need to integrate different types of data instead of relying solely on input–output data as primary source of consciousness data. These can be neuroscientific data, psychological data, and phenomenological data. For example, Block (2007) suggests considering both psychological and neuroscientific evidence. Similarly, Bayne and Shea (2020) suggest a natural kind strategy that is aimed to identify consciousness through a set of markers that cluster together: the scientific study of phenomena like hepatitis has improved by starting from a cluster of symptoms (e.g., fever, jaundice, etc.) to then investigate the biological underpinnings of the cluster (e.g., the presence of some viruses). In the same way, according to Bayne and Shea, consciousness science could take advantage of a cluster of observed “consciousness symptoms” to investigate the cluster’s mechanistic underpinnings. Our proposed “integrative approach” to consciousness science suggests that consciousness could be associated with a cluster of

16 The test suggested above examines the necessity of neural recurrence to consciousness. A similar test can be suggested to test its sufficiency. In the latter, one starts from a no consciousness state (say, a dreamless sleep) and carries out an intervention that enhances the strength of recurrent connections. If the subject experiences an increase in consciousness (say transition to dream state) this would provide support for the idea the neural recurrence is sufficient for consciousness. It is not our aim here to argue for either the necessity or sufficiency of recurrent activity, but only to suggest that they are valid scientific hypotheses.

phenomenological, functional, and neural properties, and therefore evidence must be gathered from paradigms that are sensitive to all these properties [see also (Shea, 2012; Birch, 2022)]. This is not a very new idea, and is consistent with the fact that we often use brain measurements (e.g., polygraph) to validate the veridical status of verbal reports [see also (Koch and Tsuchiya, 2007; Block 2008) for a discussion of possible visual experiences in neglect patients]. Here, we stress the importance of including phenomenological data in this "integrative approach." More specifically, phenomenological data can constrain the inferences allowed by observing neural and behavioral data, in the sense that they can define the legitimacy of those data for consciousness studies. In other words, without first-person data, we would not be able to explain why neural and functional data should be data about consciousness. The methodology for consciousness science we propose is thus a methodology that requires an integration of different types of data, and, contrary to the methodology suggested by the UA authors, acknowledges the necessary role of first-person experience in theorizing about consciousness.

It could be argued that any methodology founded on first-person experience is founded on shaky grounds, since we are prone to introspective errors and we are often confused about our experiences (Dennett, 1990; Cohen and Dennett, 2011; Schwitzgebel, 2011) - for a discussion, see (Smithies, 2013). However, our methodology does not rely on the assumption that our beliefs of what we are conscious of must be accurate, but only that we have some sort of phenomenological awareness of the contents of consciousness. The awareness itself, and not what the awareness is about, is what constrains our theorizing about consciousness, and it is thus the foundational datum for the science of consciousness. As Searle puts it, "consciousness consists in the appearances themselves. *Where appearance is concerned we cannot make any appearance-reality distinction because the appearance is the reality*" (Searle, 2008, p. 76; italics in the original). Thus, we believe that a fundamental mistake implicit in P1, P3, P4 is the assumption that behavioral data is primary to consciousness studies [see also (Kleiner, 2020; Kleiner and Hoel, 2021)].

In addition, we also believe that assumptions P1, P3, and P4 are overly restrictive, in virtue of a verificationist approach to science, which considers as scientifically meaningful only those statements that can be empirically verified. Philosophers of science have pointed out that this approach implies a clear distinction between empirical and theoretical statements, which is often unwarranted (Quine, 1951; Hanson, 1965). Surpassing this restrictive verificationist stance can ensure that other important aspects of scientific theories, like parsimony (as measured by complexity measures), consistency with background knowledge, and elegance be included in the practice of consciousness science, and there is no need to fear statements that are not directly verifiable: the scientific status of theories does not depend on whether they are constituted uniquely by empirically verifiable statements, but depends instead on whether the type of research program they generate is progressive or not (Lakatos, 1970).[17]

To conclude, we believe that in consciousness research we need to start from our own phenomenal conscious experience as primary, and investigate its physical underpinning, to be searched in the neural data. Behavior, of course, should be used, but may not always be needed (when our conscious phenomenology is clear enough, for example).

6 Discussion

We have reviewed the UA argument against causal structure theories of consciousness. We argued that the argument rests on multiple assumptions that are either not formally proven or reflect a set of overly restrictive philosophical assumptions about the proper methodology of consciousness research. We have also argued that if the rationale of the UA argument is accepted, one can construct a similar argument that targets not only causal structure theories but also functionalist ones [see also (Kleiner and Hoel, 2021)]. We believe this is the outcome of the UA-assumptions, which, despite the authors' aspirations, make functionalist theories of consciousness scientifically illegitimate and leaves little logical space for theories of consciousness.

We have suggested that premises P1, P3, and P4 are too restrictive, methodologically speaking. Instead, we propose an integrative approach, in which consciousness can be studied in tandem, through phenomenal, behavioral and neural data (Bayne and Shea, 2020). In particular, we have suggested that similarities in brain processes and structures are crucial to determine the presence and types of conscious states. Below we highlight several implications that this approach to consciousness research has, and we discuss some potential objections.

6.1 Restrictive methodology vs. restriction on the scope of current consciousness research

According to the UA, consciousness science should be primarily based on behavioral data – purely first-personal observations and 'direct' brain-based evidence (unmediated by behavior) are excluded, and a large set of theories are false or lie outside the realm of scientific investigation. However, since for many of us phenomenality, as grasped from the first-person perspective, is the primary aspect of consciousness (i.e., phenomenal realism; see (Block, 2002)), this seems equivalent to proclaiming that there can be no science of consciousness. Here we propose a different kind of limitation on consciousness research – one on the range of systems upon which theories of consciousness should be tested and built. Specifically, we argue that, at least at this stage, theories

17 Contrary to what the strongest form of verificationism implies, there are abundant examples showing that directly unobservable entities (atoms, electrons or black holes) are taken as real in the physical sciences, once they enable us to account for a variety of data in a parsimonious way. This makes them valid scientific entities (for example, Boltzmann atomic theory was accepted despite strong opposition from Ernst Mach, based on positivist/verificationist commitments). The phenomenal experience of others is a similar aspect of reality we believe science needs to accept as valid, rather than restricting itself to verbal/behavioral protocol. We acknowledge that the UA proponents might consider phenomenal experience as a scientifically valid construal that is inferred from observations, rather than being directly observable. However, although that might be true for other systems, consciousness remains directly observable in us. If UA proponents do not want to endorse the strongest form of verificationism, they would need to accept the scientific legitimacy of the first-person perspective.

of consciousness should be tested and built upon the case of human consciousness.[18] But this restriction on the range of systems that can test theories of consciousness does not imply a methodological limitation: we can, and should, be bold concerning the methods we employ to study human consciousness, giving pride of place to first-person and brain-based evidence.

We argue that the distinction between these two types of restrictions is crucial to theoretical perspectives on consciousness research in general and to the UA in particular. In fact, our proposed limitation concerns the domain of theory-testing (i.e., how we test a theory against empirical data), whereas the "hard criteria" suggested by Doerig et al. (2021) concerns theory-building (i.e., how a scientific theory is constructed in the first place) (for specific criticisms of their criteria, see Fahrenfort and van Gaal, 2021; Haun and Tsuchiya, 2021; Seth and Hohwy, 2021). Thus, we propose that, as far as consciousness research in humans is concerned, the restrictions expressed by premises 1 and 4 of the UA, and hence its conclusion, should be rejected. Specifically, in this case, contra P1, consciousness science need not rely *only* on physical measurements (like behavioral data), as in the no-report paradigm and, contra P4, it can use direct measurements of brain activity (independently of behavior). This is because, first, we all know, from first-person experience, that there are conditions when we are conscious of a suprathreshold object on display without distraction. Second, the deep similarity to other humans makes the generalization possible (Sober, 2000).

Things are different concerning the extrapolation to non-human consciousness (in cases where there is no significant similarity with human 'hardware'). In this case, only behavioral measurements are available and direct measurements of 'brain' (or other 'hardware') activity are of little use to the scientist. But this is because (in this case) we lack the first-person perspective from the very outset. And in the complete absence of similarity to humans, extrapolation becomes more difficult.

The limitations on the investigation of non-human consciousness are reflected by various familiar philosophical lines of reasoning, such as Ned Block's *"harder problem of consciousness"* (Block, 2002), which argues that we lack rational ground for believing that systems that do not share our physical realization are or are not conscious. Consider Commander Data[19] – a robot whose functional organization is similar to that of a human but whose physical realization is quite different. *Prima facie*, the functional similarity seems to provide a reason for attributing consciousness to Data, yet, the physical dissimilarity seems to provide a reason for denying such attribution. On the one hand, upon interacting with Data, you will likely take it for granted that he is conscious. On the other hand, upon discovering that he is a robot with a different 'brain' realization, your intuition might be that he is non-conscious. Block argues that there is no rational ground for adjudicating between these intuitions. We have no conception of rational belief to the effect that Data *is* or *is not* conscious – Data's consciousness is meta-inaccessible: "Not only do we lack a ground of belief, but we lack a conception of any ground of belief" (Block, 2002, p. 405). According to Block, the deep root of this epistemic problem is that we lack the justificatory basis to generalize the science of consciousness to systems like Commander Data.

> "the example of a conscious creature on which the science of consciousness is inevitably based is us […] But how can science based on us generalize to creatures that do not share our physical properties? It would seem that a form of physicalism that could embrace other creatures would have to be based at least in part on them in the first place, but that cannot be done unless we already know whether they are conscious" (Block, 2002, p. 407).

The problem, then, is not that the first-person perspective (independently of behavior) is not crucial for the study of consciousness, but that we lack that perspective in the case of Commander Data and other differently realized creatures. The same rationale holds with respect to the FF-robot discussed by the UA.

The idea that we can learn about consciousness, in general, from what we know about human consciousness, specifically, is problematic. Since it is unclear whether we can directly use our knowledge of human consciousness as justificatory ground for the attribution of consciousness to entities significantly different from us along various dimensions, it is not clear that we have a justificatory basis to either exclude or include the FF-robot from the realm of conscious entities. This would remain true even if we ascertain that in the human case consciousness depends on some kind of causal structural properties. Such confirmation of particular causal structural properties may not be necessary for consciousness in other systems. Note, though, that neither can we know whether functionalist theories capture what is crucial for non-human consciousness, since a functional organization similar to our own may be neither necessary nor sufficient for non-human consciousness. Thus, the epistemic problem that concerns extrapolation to other (differently realized) minds afflicts not only theories of physical realization (and causal structure theories), but also theories of functional organization. Earlier, we argued that in the human case, first-person and ('direct') neurological data are available, so all levels of inquiry and all theories of consciousness are legitimate. Our present point is that in the case of alien consciousness, the relevant kinds of data are unavailable and the rationale guiding our (human) consciousness-theories is inapplicable.

Given this situation, we argue that there are two theoretical options. The first is to adopt a "humility principle": given the human-centered methodology for consciousness science we are advocating, at this stage, we should in fact remain silent on alien, or non-human, consciousness. If the humility principle is adopted, the results of consciousness science should not generalize beyond creatures that are relevantly and substantially similar to us (see (Carruthers, 2019) for arguments supporting a similar position). This does not mean that we should not care about non-human consciousness, or that non-human consciousness will forever remain beyond our reach. Rather, the humility principle warns us that the current knowledge we have about consciousness is highly context-dependent (i.e., based on human-consciousness), and therefore many inferences currently drawn about non-human systems [whether they are theory-driven or not, see (Birch, 2022) for a discussion and (Butlin et al., 2023) for a case-study] might be unwarranted.

The second option is to adopt a more ambitious stance, by either formulating theories in a context-independent way [this is what Kanai and Fujisawa (2023) call 'universality'] or by justifying extrapolations to

18 In fact, "human-consciousness" might be too restrictive: biological systems which are neurologically similar to humans may also fall under the explanandum of current consciousness research. The central point is that systems that are very different from humans in their internal structures, such as FF-robots, indeed fall outside the scope of current science.

19 Commander Data is a fictional character in the *Star Trek* franchise.

the non-human case through arguments based on analogical reasoning, abductive reasoning, or a combination of both (Melnyk, 1994). One could start, for example, from the structural similarities between the source domain (i.e., the domain for which the original hypothesis is formed, for example the domain of humans in the case of consciousness) and the target domain (i.e., the domain for which the hypothesis is supposed to hold, for example organisms radically different from humans and non-biological systems in the case of consciousness), and then claim that given that phenomenal properties correlate with specific properties in the source domain, the most parsimonious hypothesis is that a correlation between similar properties in the target and phenomenal properties occurs (Barron and Klein, 2016; Godfrey-Smith, 2017; Bayne and Shea, 2020; Tsuchiya and Saigo, 2021; Birch, 2022). In this case, inferences about the conscious state of target systems could be justified, thus solving Block's 'harder-problem' (Hohwy, 2004), by acknowledging how often in science unobservable entities are legitimately posited in the context of discovery (for example, the electron). These inferences can be conjectures driving further testing, and if they are based on brain (or at least, "hardware") similarity, they could potentially be tested by implanting specific structures, mimicking the relevant structures of target systems, into our own brains [see also (Shevlin, 2021) for a discussion].

These two options have contrasting strengths and weaknesses: the humility principle can ensure that our applications of consciousness theories are more grounded, at the cost of limiting the explanatory power of such theories. Adopting the more ambitious stance, instead, can ensure stronger explanatory power, at the cost of either requiring a further ampliative argument or formulating theories that might end up being too liberal in their ascription of consciousness [for discussions, see (Block, 1978; Block, 2002; Shevlin, 2021; Kanai and Fujisawa, 2023)].

Independently of which of these two stances toward extrapolative practices is favored, we argue that consciousness science should be firmly built on evidence gathered from humans (Negro, 2020), and such evidence should include not only behavioral evidence but also neural and phenomenological (first-person) evidence. This means that human-based theories of consciousness should not be dismissed because of what they predict about non-human consciousness (Albantakis, 2020a; Tsuchiya et al., 2020).

6.2 Objections

6.2.1 Non-circular testing of the recurrence hypothesis

In section 5.1, we have argued that it is possible to test, in a non-circular way, the hypothesis that recurrence is necessary for human consciousness. One objection that we anticipate to the TMS-test we propose is that one should not just interfere with recurrent processes, but replace them with some appropriately tuned FF-circuits. Let us assume that such a circuit exists, that allows an FF-robot to respond as a normal human in consciousness experiments (it responds that orange is more similar to red than yellow, that this pain is unpleasant or that this stimulus has not been seen). Obviously, there is no way to test this robot in our experiment which affects recurrent connections (as it does not have any). This could be a problem if we aim to establish a general theory of consciousness, that extends to non-human C, because if the FF-robot is in fact conscious, then the hypothesis that recurrence is necessary for consciousness *in general* would be disproved. So, if the goal were to build a theory of consciousness in general, we would require a further argument to show that the FF circuit is not consciousness-generating in general, and not only in humans. But if we adopt the humility principle we can restrict the scope of the TMS-experiment above-proposed, and safely conclude that it offers a critical test of human-C,[20] whereas if we maintain the possibility of extrapolation to non-human systems, we need an argument to show that the FF-circuit is not conscious. Thus, this objection could put pressure on the generality of the scope of consciousness science for those who do not subscribe to the humility principle. Still, it does not seem decisive against the idea that the TMS-test we proposed is valid for testing the recurrence hypothesis in humans.

7 Conclusion

To conclude, while the UA has opened a stimulating debate that contributed to clarifying a number of conflicting intuitions on the nature of consciousness[21], we believe that the (hard-criteria) UA program (Doerig et al., 2021) is too restrictive and that it hinders, rather than promotes, the scientific research of consciousness. We identified several problems, involving both philosophical and methodological viewpoints and proposed an alternative less restrictive approach that facilitates the convergence from phenomenology, theory, and neuroscience. As consciousness research is primarily based on phenomenology in humans, we cannot directly access non-human consciousness. In turn, this means that conjecture about (potential) non-human consciousness cannot be used to restrict a class of consciousness theories as the UA attempted.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

MU: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. NN: Conceptualization, Methodology, Writing – review & editing. HJ: Conceptualization, Methodology, Supervision, Writing – review & editing. NT: Conceptualization, Methodology, Supervision, Writing – review & editing.

20 We wish to clarify that the aim of this test is not to select between a functionalist and a brain identity theory based on a specific neural mechanism. We rather accept that the neural mechanism of consciousness has functional characteristics (such as enhancing perceptual discriminations and motor control), thus we fully expect that a disruption of the consciousness mechanism has functional effects. The test only asks if the neural mechanism has a causal (or constitutive) relation to consciousness in humans.

21 See Gidon et al. (2022) for a recent thought experiment involving replay of neural activity associated with consciousness, which further probes some striking predictions of causal-structure theories of consciousness. The replay is not only guaranteed to generate the same responses to the same inputs (as in UA), but also enforces the same spike trains in all neurons.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. MU was supported by a grant from the Israel Science Foundation (grant no. 1602/22). NT was supported by the National Health Medical Research Council (NT, APP1183280) and by Japan Society for the Promotion of Science, Grant-in-Aid for Transformative Research Areas (NT 20H05710, 23H04830). NN wishes to acknowledge the support provided by the Azrieli Foundation.

Acknowledgments

We wish to thank Michael Herrmann and Ariel Kernberg for a critical reading of the manuscript and for helpful discussions, and Aaron Kravitz for English editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Albantakis, L. (2020a). Unfolding the substitution argument. *Conscious (ness) realist*. Available at: https://www.consciousnessrealist.com/unfolding-argument-commentary/

Albantakis, L. (2020b). The Greek cave: why a little bit of causal structure is necessary… Even for functionalists. *Conscious (ness) realist*. Available at: https://www.consciousnessrealist.com/greek-cave/

Albantakis, L., Barbosa, L. S., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., et al. (2023). Integrated information theory (IIT) 4.0: formulating the properties of phenomenal existence in physical terms. *PLoS Comput. Biol.* 19:e1011465. doi: 10.1371/journal.pcbi.1011465

Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003

Aru, J., Suzuki, M., and Larkum, M. E. (2020). Cellular mechanisms of conscious processing. *Trends Cogn. Sci.* 24, 814–825. doi: 10.1016/j.tics.2020.07.006

Barron, A. B., and Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proc. Natl. Acad. Sci.* 113, 4900–4908. doi: 10.1073/pnas.1520084113

Bayne, T. (2021). *Philosophy of mind: an introduction*. London: Routledge, Taylor & Francis Group.

Bayne, T., and Shea, N. (2020). Consciousness, concepts and natural kinds. *Philos. Top.* 48, 65–83. doi: 10.5840/philtopics20204814

Birch, J. (2022). The search for invertebrate consciousness. *Noûs* 56, 133–153. doi: 10.1111/nous.12351

Block, N. (1978). Troubles with functionalism. *Minn. Stud. Philos. Sci.* 9, 261–325.

Block, N. (1981). Psychologism and behaviorism. *Phil. Rev.* 90, 5–43. doi: 10.2307/2184371

Block, N. (1990). Inverted earth. *Philos. Perspect.* 4, 53–79. doi: 10.2307/2214187

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188

Block, N. (2002). The harder problem of consciousness. *J. Philos.* 99, 391–425. doi: 10.2307/3655621

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–499. doi: 10.1017/S0140525X07002786

Block, N. (2008). Consciousness and cognitive access. *Proc. Aristot. Soc.* 108, 289–317. doi: 10.1111/j.1467-9264.2008.00247.x

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends Cogn. Sci.* 15, 567–575. doi: 10.1016/j.tics.2011.11.001

Block, N. (2019). What is wrong with the no-report paradigm and how to fix it. *Trends Cogn. Sci.* 23, 1003–1013. doi: 10.1016/j.tics.2019.10.001

Block, N., and Fodor, J. A. (1972). What psychological states are not. *Philos. Rev.* 81, 159–181. doi: 10.2307/2183991

Bourget, D., and Chalmers, D. J. (2023). Philosophers on philosophy: the 2020 Phil papers survey. *Philosophers' Imprint* 23:11. doi: 10.3998/phimp.2109

Braddon-Mitchell, D., and Jackson, F. (2007). *The philosophy of mind and cognition*. London: Blackwell.

Bronfman, Z. Z., Brezis, N., Jacobson, H., and Usher, M. (2014). We see more than we can report: "cost free" color phenomenality outside focal attention. *Psychol. Sci.* 25, 1394–1403. doi: 10.1177/0956797614532656

Bronfman, Z. Z., Jacobson, H., and Usher, M. (2019). Impoverished or rich consciousness outside attentional focus: recent data tip the balance for overflow. *Mind Lang.* 34, 423–444. doi: 10.1111/mila.12217

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., et al., (2023). *Consciousness in artificial intelligence: insights from the science of consciousness*.

Cabessa, J., and Siegelmann, H. T. (2012). The computational power of interactive recurrent neural networks. *Neural Comput.* 24, 996–1019. doi: 10.1162/NECO_a_00263

Carruthers, P. (2019). *Human and animal minds: the consciousness questions laid to rest*. Oxford: Oxford University Press.

Cea, I. (2020). Integrated information theory of consciousness is a functionalist emergentism. *Synthese* 199, 2199–2224. doi: 10.1007/s11229-020-02878-8

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219. doi: 10.1093/acprof:oso/9780195311105.003.0001

Chalmers, D. J. (1996). *The conscious mind: in search of a fundamental theory*. New-York, Oxford: Oxford University Press.

Chalmers, D. J. (2000). "What is a neural correlate of consciousness?" in *Neural correlates of consciousness: empirical and conceptual questions*. ed. T. Metzinger (The MIT Press), 17–39.

Cohen, M. A., and Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends Cogn. Sci.* 15, 358–364. doi: 10.1016/j.tics.2011.06.008

Cohen, M. A., Dennett, D. C., and Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends Cogn. Sci.* 20, 324–335. doi: 10.1016/j.tics.2016.03.006

Crick, F., and Koch, C. (1990). Toward a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275.

Crick, F., and Koch, C. (1998). Consciousness and neuroscience. *Cereb. Cortex* 8, 97–107. doi: 10.1093/cercor/8.2.97

de Graaf, T. A., Hsieh, P. J., and Sack, A. T. (2012). The 'correlates' in neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 191–197. doi: 10.1016/j.neubiorev.2011.05.012

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. London: Viking.

Dehaene, S., Kerszberg, M., and Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14529–14534. doi: 10.1073/pnas.95.24.14529

Del Pin, S. H., Skóra, Z., Sandberg, K., Overgaard, M., and Wierzchoń, M. (2021). Comparing theories of consciousness: why it matters and how to do it. *Neurosci Cons.* 2021:niab019. doi: 10.1093/nc/niab019

Dennett, D. (1990). "Quining Qualia" in *Mind and cognition*. ed. W. Lycan (Oxford: Oxford: Blackwell), 519–548.

Descartes, R. (1641/1996) in *Descartes: meditations on first philosophy: with selections from the objections and replies*. eds. J. Cottingham and B. Williams (Oxford: Cambridge University Press).

Doerig, A., Schurger, A., and Herzog, M. H. (2021). Hard criteria for empirical theories of consciousness. *Cogn. Neurosci.* 12, 41–62. doi: 10.1080/17588928.2020.1772214

Doerig, A., Schurger, A., Hess, K., and Herzog, M. H. (2019). The unfolding argument: why IIT and other causal structure theories cannot explain consciousness. *Conscious. Cogn.* 72, 49–59. doi: 10.1016/j.concog.2019.04.002

Ellia, F., and Chis-Ciure, R. (2022). Consciousness and complexity: neurobiological naturalism and integrated information theory. *Conscious. Cogn.* 100:103281. doi: 10.1016/j.concog.2022.103281

Ellia, F., Hendren, J., Grasso, M., Kozma, C., Mindt, G. P, et al. (2021). Consciousness and the fallacy of misplaced objectivity. *Neurosci. Cons.* 2021. doi: 10.1093/nc/niab032

Fahrenfort, J. J., and van Gaal, S. (2021). Criteria for empirical theories of consciousness should focus on the explanatory power of mechanisms, not on functional equivalence. *Cogn. Neurosci.* 12, 93–94. doi: 10.1080/17588928.2020.1838470

Fodor, J. A. (1981). The mind-body problem. *Sci. Am.* 244, 114–123. doi: 10.1038/scientificamerican0181-114

Francken, J. C., Beerendonk, L., Molenaar, D., Fahrenfort, J. J., Kiverstein, J. D., Seth, A. K., et al. (2022). An academic survey on theoretical foundations, common assumptions and the current state of consciousness science. *Neurosci. Cons.* 2022. doi: 10.1093/nc/niac011

Ganesh, N. (2020). *No substitute for functionalism -- a reply to 'Falsification & Consciousness'.*

Gidon, A., Aru, J., and Larkum, M. E. (2022). Does brain activity cause consciousness? A thought experiment. *PLoS Biol.* 20:e3001651. doi: 10.1371/journal.pbio.3001651

Godfrey-Smith, P. (2017). *Other minds: The Octopus and the evolution of intelligent life*. London: William Collins.

Grasso, M. (2019). IIT vs. Russellian monism: a metaphysical showdown on the content of experience. *J. Conscious. Stud.* 26, 48–75.

Hanson, N. R. (1965). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge, UK: Cambridge University Press.

Haun, A., and Tsuchiya, N. (2021). Reasonable criteria for functionalists; scarce criteria from phenomenological perspective. *Cogn. Neurosci.* 12, 95–96. doi: 10.1080/17588928.2020.1838473

He, B. J. (2023). Towards a pluralistic neurobiological understanding of consciousness. *Trends Cogn. Sci.* 27, 420–432. doi: 10.1016/j.tics.2023.02.001

Herzog, M. H., Schurger, A., and Doerig, A. (2022). First-person experience cannot rescue causal structure theories from the unfolding argument. *Conscious. Cogn.* 98:103261. doi: 10.1016/j.concog.2021.103261

Hohwy, J. (2004). Evidence, explanation, and experience: on the harder problem of consciousness. *J. Philos.* 101, 242–254. doi: 10.5840/jphil2004101525

Jackson, F. (1982). Epiphenomenal Qualia. *Philos. Q.* 32, 127–136. doi: 10.2307/2960077

Josselyn, S. A., and Tonegawa, S. (2020). Memory engrams: recalling the past and imagining the future. *Science* 367:eaaw 4325. doi: 10.1126/science.aaw4325

Kanai, R., and Fujisawa, I. (2023). Towards a universal theory of consciousness. doi: 10.31234/osf.io/r5t2n

Kent, L., and Wittmann, M. (2021). Time consciousness: the missing link in theories of consciousness. *Neurosci. Cons.* 2021. doi: 10.1093/nc/niab011

King, J. R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* 18, 203–210. doi: 10.1016/j.tics.2014.01.002

Kleiner, J. (2020). Brain states matter. A reply to the unfolding argument. *Conscious. Cogn.* 85:102981. doi: 10.1016/j.concog.2020.102981

Kleiner, J., and Hoel, E. (2021). Falsification and consciousness. *Neurosci. Cons.* 2021. doi: 10.1093/nc/niab001

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22

Koch, C., and Tsuchiya, N. (2007). Phenomenology without conscious access is a form of consciousness without top-down attention. *Behav. Brain Sci.* 30, 509–510. doi: 10.1017/S0140525X07002907

Lakatos, I. (1970). "Falsification and the methodology of scientific research Programmes" in *Criticism and the growth of knowledge*. ed. A. Musgrave (Cambridge: Cambridge University Press)

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. doi: 10.1016/j.tics.2006.09.001

Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X

Landman, R., Spekreijse, H., and Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vis. Res.* 43, 149–164. doi: 10.1016/S0042-6989(02)00402-9

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., et al. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* 484, 381–385. doi: 10.1038/nature11028

Love, B. C. (2015). The algorithmic level is the bridge between computation and brain. *Top. Cogn. Sci.* 7, 230–242. doi: 10.1111/tops.12131

Machery, E. (2012). Why I stopped worrying about the definition of life … and why you should as well. *Synthese* 185, 145–164. doi: 10.1007/s11229-011-9880-1

Malach, R. (2022). The role of the prefrontal cortex in conscious perception: the localist perspective. *J. Conscious. Stud.* 29, 93–114. doi: 10.53765/20512201.29.7.093

Mallatt, J. (2021). A traditional scientific perspective on the integrated information theory of consciousness. *Entropy* 23:650. doi: 10.3390/e23060650

Marr, D. (1982). *Vision* W. H. Freeman. The MIT Press.

Massimini, M., and Tononi, G. (2018) in *Sizing up consciousness. Towards an objective measure of the capacity for experience*. ed. F. Andersen (Oxford: Oxford University Press).

McFadden, J. (2020). Integrating information in the brain's EM field: the cemi field theory of consciousness. *Neurosci. Cons.* 2020:niaa016. doi: 10.1093/nc/niaa016

Melloni, L., Mudrik, L., Pitts, M., Bendtz, K., Ferrante, O., Gorska, U., et al. (2023). An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. *PLoS One* 18:e0268577. doi: 10.1371/journal.pone.0268577

Melloni, L., Mudrik, L., Pitts, M., and Koch, C. (2021). Making the hard problem of consciousness easier. *Science* 372, 911–912. doi: 10.1126/science.abj3259

Melnyk, A. (1994). Inference to the best explanation and other minds. *Australas. J. Philos.* 72, 482–491. doi: 10.1080/00048409412346281

Michel, M. (2023). Calibration in consciousness science. *Erkenntnis* 88, 829–850. doi: 10.1007/s10670-021-00383-z

Michel, M., and Malach, R. (2022). Making progress on the prefrontal debate. *J. Conscious. Stud.* 29, 158–164. doi: 10.53765/20512201.29.7.158

Miller, C. F. (2014). "Turing machines to word problems" in *Turing's legacy: developments from Turing's ideas in logic*, Miller: Cambridge University Press; 329–385.

Monti, M. M. (2015). "Disorders of consciousness" in *Emerging trends in the social and behavioral sciences*, 1–13. doi: 10.1002/9781118900772

Mudrik, L., Faivre, N., and Koch, C. (2014). Information integration without awareness. *Trends Cogn. Sci.* 18, 488–496. doi: 10.1016/j.tics.2014.04.009

Nagel, T. (1974). What is it like to be a bat? *Phil. Rev.* 83, 435–450. doi: 10.2307/2183914

Negro, N. (2020). Phenomenology-first versus third-person approaches in the science of consciousness: the case of the integrated information theory and the unfolding argument. *Phenomenol. Cogn. Sci.* 19, 979–996. doi: 10.1007/s11097-020-09681-3

Negro, N. (2022). Emergentist integrated information theory. *Erkenntnis*, 1–23. doi: 10.1007/s10670-022-00612-z

Noy, N., Bickel, S., Zion-Golumbic, E., Harel, M., Golan, T., Davidesco, I., et al. (2015). Ignition's glow: ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. *Conscious. Cogn.* 35, 206–224. doi: 10.1016/j.concog.2015.03.006

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588

Overgaard, M., and Fazekas, P. (2016). Can no-report paradigms extract true correlates of consciousness? *Trends Cogn. Sci.* 20, 241–242. doi: 10.1016/j.tics.2016.01.004

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., and Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science* 313:1402. doi: 10.1126/science.1130197

Palmer, S. E. (1999). Color, consciousness, and the isomorphism constraint. *Behav. Brain Sci.* 22, 923–943. doi: 10.1017/s0140525x99002216

Phillips, I. (2011). Perception and iconic memory: what Sperling Doesn't show. *Mind Lang.* 26, 381–411. doi: 10.1111/j.1468-0017.2011.01422.x

Phillips, I. (2016). No watershed for overflow: recent work on the richness of consciousness. *Philos. Psychol.* 29, 236–249. doi: 10.1080/09515089.2015.1079604

Pitts, M. A., Metzler, S., and Hillyard, S. A. (2014). Isolating neural correlates of conscious perception from neural correlates of reporting one's perception. *Front. Psychol.* 5:1078. doi: 10.3389/fpsyg.2014.01078

Place, U. T. (1956). Is consciousness a brain process? *Br. J. Psychol.* 47, 44–50. doi: 10.1111/j.2044-8295.1956.tb00560.x

Putnam, H. (1967). "The nature of mental states" in *Art, mind, and religion*. eds. W. H. Capitan and D. D. Merrill (Pittsburgh, PA: Pittsburgh University Press), 1–223.

Quine, W. V. O. (1951). Main trends in recent philosophy: two dogmas of empiricism. *Philos. Rev.* 60, 20–43. doi: 10.2307/2181906

Ruffini, G., Lopez-Sola, E., and Vohryzek, J. (2022). *Algorithmic structure of experience and the unfolding argument*. doi: 10.31234/osf.io/7nbsw

Ryle, G. (1949). *The concept of mind* London and New York: Routledge Taylor & Francis Group.

Schwitzgebel, E. (2011). *Perplexities of consciousness*, Cambridge, MA, MIT Press.

Searle, J. R. (2008). "Reductionism and the irreducibility of consciousness" in *Emergence: Contemporary readings in philosophy and science*. eds. M. A. Bedau and P. Humphreys (MIT Press), 69–80.

Seth, A. K., and Bayne, T. (2022). Theories of consciousness. *Nat. Rev. Neurosci.* 23, 439–452. doi: 10.1038/s41583-022-00587-4

Seth, A. K., and Hohwy, J. (2021). Predictive processing as an empirical theory for consciousness science. *Cogn. Neurosci.* 12, 89–90. doi: 10.1080/17588928.2020.1838467

Shea, N. (2012). Methodological encounters with the phenomenal kind. *Philos. Phenomenol. Res.* 84, 307–344. doi: 10.1111/j.1933-1592.2010.00483.x

Shevlin, H. (2021). Non-human consciousness and the specificity problem: a modest theoretical proposal. *Mind Lang.* 36, 297–314. doi: 10.1111/mila.12338

Shoemaker, S. (1975). Functionalism and qualia. *Philosophical studies: an international journal for philosophy in the analytic. Traditio* 27, 291–315. doi: 10.1007/BF01225748

Shoemaker, S. (1982). The inverted spectrum. *J. Philos.* 79, 357–381. doi: 10.2307/2026213

Siegelmann, H. T., and Sontag, E. D. (1995). On the computational power of neural nets. *J. Comput. Syst. Sci.* 50, 132–150. doi: 10.1006/jcss.1995.1013

Signorelli, C. M., Szczotka, J., and Prentner, R. (2021). Explanatory profiles of models of consciousness - towards a systematic classification. *Neurosci. Cons.* 2021. doi: 10.1093/nc/niab021

Sligte, I. G., Scholte, H. S., and Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS One* 3:e1699. doi: 10.1371/journal.pone.0001699

Smart, J. J. C. (1959). Sensations and brain processes. *Phil. Rev.* 68, 141–156. doi: 10.2307/2182164

Smithies, D. (2013). On the unreliability of introspection. *Philos. Stud.* 165, 1177–1186. doi: 10.1007/s11098-013-0150-6

Sober, E. (2000). Evolution and the problem of other minds. *J. Philos.* 97:365. doi: 10.2307/2678410

Solms, M. (2019). The hard problem of consciousness and the free energy principle [hypothesis and theory]. *Front. Psychol.* 9:2714. doi: 10.3389/fpsyg.2018.02714

Solms, M., and Friston, K. (2018). How and why consciousness arises: some considerations from physics and physiology. *J. Conscious. Stud.* 25, 202–238.

Tononi, G. (2017). "Integrated information theory of consciousness" in *The Blackwell companion to consciousness*, eds. Susan Schneider and Max Velmans. John Wiley & Sons Ltd. 621–633.

Tononi, G., Albantakis, L., Boly, M., Cirelli, C., and Koch, C. (2022). *Only what exists can cause: an intrinsic view of free will*.

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44

Tononi, G., and Koch, C. (2015). Consciousness: here, there and everywhere? *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 370:20140167. doi: 10.1098/rstb.2014.0167

Tsuchiya, N., Andrillon, T., and Haun, A. M. (2020). A reply to "the unfolding argument": beyond functionalism/behaviorism and towards a science of causal structure theories of consciousness. *Conscious. Cogn.* 79:102877. doi: 10.1016/j.concog.2020.102877

Tsuchiya, N., and Saigo, H. (2021). A relational approach to consciousness: categories of level and contents of consciousness. *Neurosci. Cons.* 2021. doi: 10.1093/nc/niab034

Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. A. F. (2015). No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* 19, 757–770. doi: 10.1016/j.tics.2015.10.002

Usher, M. (2021). Refuting the unfolding-argument on the irrelevance of causal structure to consciousness. *Conscious. Cogn.* 95:103212. doi: 10.1016/j.concog.2021.103212

Usher, M., Bronfman, Z. Z., Talmor, S., Jacobson, H., and Eitam, B. (2018). Consciousness without report: insights from summary statistics and inattention 'blindness'. *Phil. Transact. Royal Society B* 373:20170354. doi: 10.1098/rstb.2017.0354

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychol. Rev.* 20, 158–177. doi: 10.1037/h0074428

Watson, J. B. (1994). Psychology as the behaviorist views it. *Psychol. Rev.* 101, 248–253. doi: 10.1037/0033-295X.101.2.248

Zeki, S., and Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 353, 1911–1914. doi: 10.1098/rstb.1998.0343

frontiers | Frontiers in Neuroscience

TYPE Study Protocol
PUBLISHED 13 February 2024
DOI 10.3389/fnins.2024.1306344

OPEN ACCESS

EDITED BY
Luca Simione,
UNINT—Università degli studi Internazionali di Roma, Italy

REVIEWED BY
Zirui Huang,
University of Michigan, United States
Yu Hayashi,
University of Tsukuba, Japan
Bruno Van Swinderen,
The University of Queensland, Australia

*CORRESPONDENCE
Jitka Annen
✉ Jitka.Annen@uliege.be
Benedetta Cecconi
✉ Benedetta.cecconi@uliege.be

†These authors have contributed equally to this work

RECEIVED 06 October 2023
ACCEPTED 29 January 2024
PUBLISHED 13 February 2024

CITATION
Cecconi B, Montupil J, Mortaheb S, Panda R, Sanders RD, Phillips C, Alnagger N, Remacle E, Defresne A, Boly M, Bahri MA, Lamalle L, Laureys S, Gosseries O, Bonhomme V and Annen J (2024) Study protocol: Cerebral characterization of sensory gating in disconnected dreaming states during propofol anesthesia using fMRI. *Front. Neurosci.* 18:1306344. doi: 10.3389/fnins.2024.1306344

Study protocol: Cerebral characterization of sensory gating in disconnected dreaming states during propofol anesthesia using fMRI

Benedetta Cecconi[1,2]*, Javier Montupil[3,4], Sepehr Mortaheb[5], Rajanikant Panda[1,2], Robert D. Sanders[6,7], Christophe Phillips[8], Naji Alnagger[1,2], Emma Remacle[1], Aline Defresne[3,4,9], Melanie Boly[10], Mohamed Ali Bahri[11], Laurent Lamalle[11], Steven Laureys[1,2,12,13], Olivia Gosseries[1,2], Vincent Bonhomme[3,9†] and Jitka Annen[1,2,14*†]

[1]Coma Science Group, GIGA-Consciousness, GIGA Institute, University of Liège, Liège, Belgium, [2]Centre du Cerveau[2], University Hospital of Liège, Liège, Belgium, [3]Anesthesia and Perioperative Neuroscience Laboratory, GIGA-Consciousness, GIGA Institute, University of Liège, Liège, Belgium, [4]University Department of Anesthesia and Intensive Care Medicine, Centre Hospitalier Régional de la Citadelle (CHR Citadelle), Liège, Belgium, [5]Physiology of Cognition Research Lab, GIGA-Consciousness, GIGA Institute, University of Liège, Liege, Belgium, [6]Central Clinical School, Sydney Medical School & NHMRC Clinical Trials Centre, University of Sydney, Camperdown, NSW, Australia, [7]Department of Anaesthetics & Institute of Academic Surgery, Royal Prince Alfred Hospital, Camperdown, NSW, Australia, [8]GIGA-CRC—In vivo Imaging—Neuroimaging, Data Acquisition and Processing, GIGA Institute, University of Liège, Liège, Belgium, [9]Department of Anesthesia and Intensive Care Medicine, Liège University Hospital, Liège, Belgium, [10]Department of Psychiatry, Wisconsin Institute for Sleep and Consciousness, University of Wisconsin, Madison, WI, United States, [11]GIGA-CRC—In vivo Imaging—Aging & Memory, GIGA Institute, University of Liège, Liège, Belgium, [12]Cervo Brain Research Centre, University Institute in Mental Health of Quebec, Québec, QC, Canada, [13]Consciousness Science Institute, Hangzhou Normal University, Hangzhou, China, [14]Department of Data Analysis, University of Ghent, Ghent, Belgium

Background: Disconnected consciousness describes a state in which subjective experience (i.e., consciousness) becomes isolated from the external world. It appears frequently during sleep or sedation, when subjective experiences remain vivid but are unaffected by external stimuli. Traditional methods of differentiating connected and disconnected consciousness, such as relying on behavioral responsiveness or on post-anesthesia reports, have demonstrated limited accuracy: unresponsiveness has been shown to not necessarily equate to unconsciousness and amnesic effects of anesthesia and sleep can impair explicit recollection of events occurred during sleep/sedation. Due to these methodological challenges, our understanding of the neural mechanisms underlying sensory disconnection remains limited.

Methods: To overcome these methodological challenges, we employ a distinctive strategy by combining a serial awakening paradigm with auditory stimulation during mild propofol sedation. While under sedation, participants are systematically exposed to auditory stimuli and questioned about their subjective experience (to assess consciousness) and their awareness of the sounds (to evaluate connectedness/disconnectedness from the environment). The data collected through interviews are used to categorize participants into connected and disconnected consciousness states. This method circumvents the requirement for responsiveness in assessing consciousness and mitigates amnesic effects of anesthesia as participants are questioned while still under sedation. Functional MRI data are concurrently collected to investigate cerebral activity patterns during connected and disconnected states, to elucidate sensory disconnection neural gating mechanisms. We examine whether this gating mechanism resides at the thalamic level or results from disruptions in information propagation to higher cortices. Furthermore, we explore the potential role of slow-wave activity (SWA) in inducing disconnected consciousness by quantifying high-frequency BOLD oscillations, a known correlate of slow-wave activity.

Discussion: This study represents a notable advancement in the investigation of sensory disconnection. The serial awakening paradigm effectively mitigates amnesic effects by collecting reports immediately after regaining responsiveness, while still under sedation. Ultimately, this research holds the potential to understand how sensory gating is achieved at the neural level. These biomarkers might be relevant for the development of sensitive anesthesia monitoring to avoid intraoperative connected consciousness and for the assessment of patients suffering from pathologically reduced consciousness.

Clinical trial registration: European Union Drug Regulating Authorities Clinical Trials Database (EudraCT), identifier 2020-003524-17.

KEYWORDS

disconnected consciousness, serial awakening paradigm, propofol sedation, slow wave activity, sensory gating

1 Introduction

During wakefulness, under normal conditions, our subjective experience (i.e., consciousness) is usually strongly influenced by external, environmental stimuli, that is, our consciousness is connected to the physical world [i.e., connected consciousness (CC)]. When transitioning to dream states (such as physiological dreaming or anesthesia-induced dreaming), our subjective experience often continues to be remarkably rich. Highly vivid sensory experiences during dreaming are frequently reported, yet they are usually unaffected by external stimuli, i.e., our consciousness is disconnected from the physical world [i.e., disconnected consciousness (DC)] (Figure 1). Corticocortical connections are functionally preserved to generate dreaming experiences, yet, somewhere in the thalamocortical stream, stimuli from the external world are blocked from conscious processing. Currently, a major obstacle to identify the cerebral gating mechanisms underlying sensory disconnection is the lack of behavioral differentiation between disconnected and connected conscious states in the neuroimaging literature. Previous research conducted on anesthetized participants has usually assumed a binary context, comparing brain activity acquired during wakefulness with brain activity acquired during presumed unconsciousness, where unconsciousness was inferred from participants' unresponsiveness. Inferring the presence or absence of consciousness from (un) responsiveness[1] has now been widely shown to be inaccurate, as unresponsiveness does not always correspond to unconsciousness (Sanders et al., 2012). Studies using the isolated forearm technique (i.e., assessing responsiveness by preventing muscle relaxants to act on one of the forearms) revealed in fact that up to 37% (Sanders et al., 2012; Linassi et al., 2018) of anesthetized patients, despite looking deeply asleep, were conscious of external stimuli (Sanders et al., 2012; Linassi et al., 2018). However, more recent estimates from clinical practice suggest that 5–10% of patients in routine clinical care experience these episodes (Sanders et al., 2017; Lennertz et al., 2023). Episodes of intraoperative dreaming are more frequent and have been estimated to occur in 22–59% of anesthetized, unresponsive patients (Leslie et al., 2007; Errando et al., 2008; Noreika et al., 2011). This implies that the supposed neural signature of consciousness gathered from classical anesthesia studies might conflate disconnected, connected consciousness, unconsciousness or the alternation between these states. We here propose that sensory perception of external stimuli can fluctuate while consciousness remains constant and independently of arousal (e.g., during dreaming), resulting in disconnected and connected dream-like experiences.

To disentangle between these different states and investigate the neural basis of (dis)connected consciousness,[2] more recent studies have resorted to serial awakening paradigms during sedation (Radek et al., 2018; Scheinin et al., 2021; Casey et al., 2022a; Valli et al., 2023). With serial awakenings, participants are directly questioned about their mental activity that was ongoing before being awakened, minimizing the lack of explicit recall which is common due to amnesic effects of anesthesia or sleep (Siclari et al., 2013, 2017). However,

1 For brevity, the presence or lack of behavioral responses are written as (un) responsiveness.

2 For brevity, to refer to both connected or disconnected consciousness we use (dis)connected consciousness.

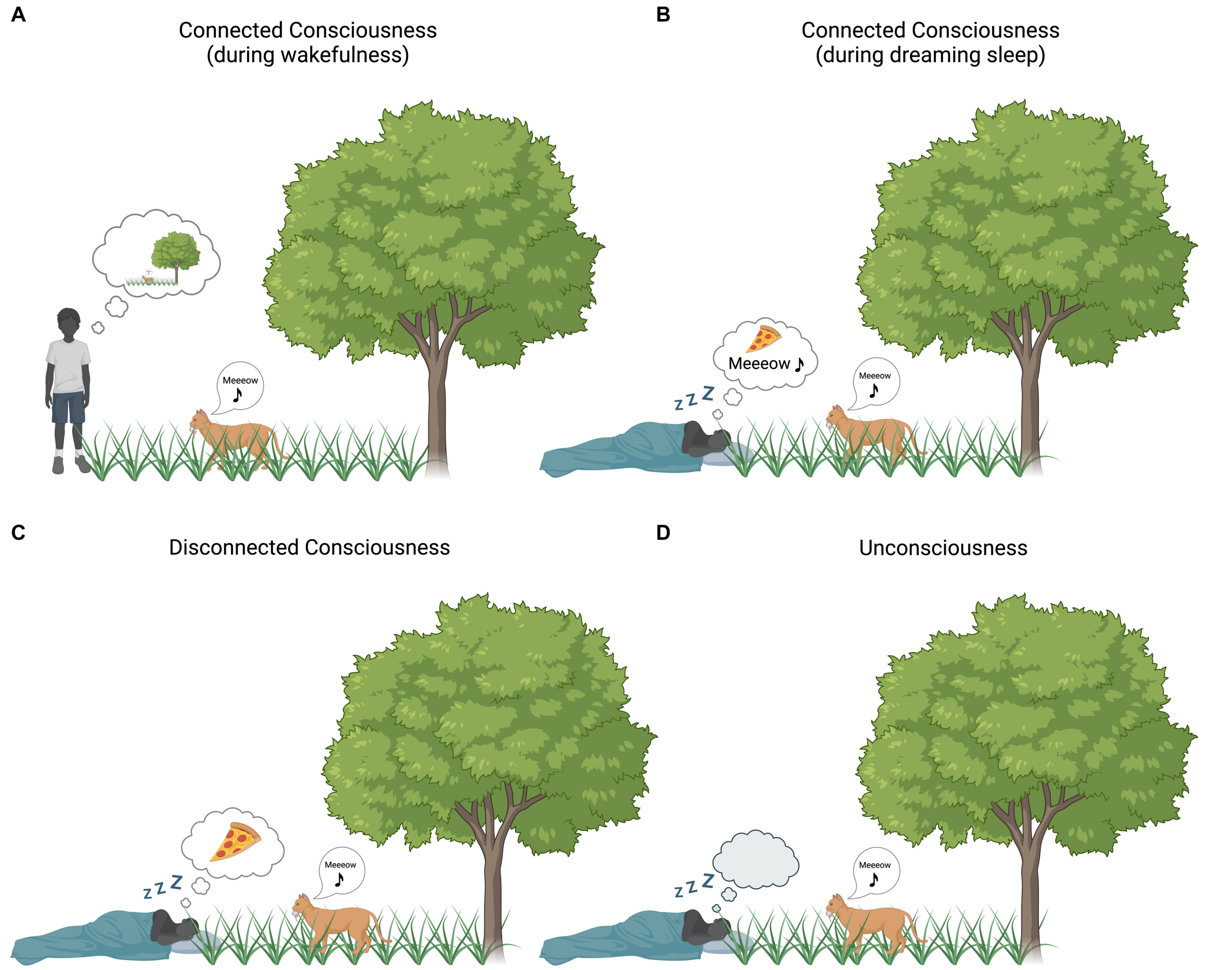

FIGURE 1
Environmental (dis)connection and consciousness. Bubble clouds represent subjective experience [i.e., (un)consciousness]. **(A)** Connected consciousness during wakefulness: the boy's subjective experience is strongly influenced by the surrounding environment. **(B)** Connected consciousness during dreaming: the subjective experience of the sleeping boy is partially influenced by the surrounding environment (the boy is dreaming of eating pizza) and, at the same time, sounds from the surrounding environment (in this case the cat's sound) are incorporated into his experience. **(C)** Disconnected consciousness during dreaming: the boy's subjective experience is not influenced by the surrounding environment (the boy is only dreaming about eating pizza, no environmental stimuli are incorporated into the ongoing experience). **(D)** Unconsciousness: the boy is devoid of any experience, whether originating internally or externally. Created with BioRender.com.

contrasting episodes of (unresponsive) disconnected consciousness with other behavioral states [e.g., responsive wakefulness (Scheinin et al., 2021; Valli et al., 2023), self-reported wakefulness (Casey et al., 2022a)] impedes the segregation of the neural correlate of disconnected consciousness. Indeed, instead of reflecting the neural correlates of sensory (dis)connection, such contrasts could reflect differences in responsiveness or arousal system. A common assumption, also shared in these studies, is that connected consciousness is wakefulness, but this is not necessarily always the case, not least due to the complexity of defining wakefulness. There is evidence that during anesthesia patients often incorporate auditory and somatosensory stimuli into their dreams or perceive such stimuli in a dream-like state (Leslie et al., 2007, 2009; Leslie, 2010; Noreika et al., 2011; Radek et al., 2018). These episodes have been referred to as "near-miss awareness" (Leslie, 2010). Evidence that environmental stimuli can be perceived during dreaming without necessarily triggering wakefulness is also documented in sleep studies (Nielsen, 1993; Leslie and Ogilvie, 1996; Schredl et al., 2009). As the search for the neural correlates of consciousness (NCC) has been refined over the years by distilling the proper NCC from its prerequisites and consequences, the same should be attempted in the search for the NC of (dis)connected consciousness.

Although there is suggestive evidence indicating that a breakdown in cortical effective connectivity (Massimini et al., 2005) might underlie the loss of consciousness in anesthesia (Boly et al., 2012) and sleep (Esser et al., 2009), the challenge heightens when attempting to pinpoint the specific mechanism responsible for the loss of environmental connection during these states. While we know that sensory stimuli reach primary sensory regions during presumed unconscious and dream states, it is unknown how they are processed in disconnected states in both primary and secondary regions. In REM sleep (Funk et al., 2016) and mild sedation with propofol (Boly et al., 2012), deactivation of primary areas has been shown to coexist with the activation of secondary/associative regions, favoring top-down over bottom-up cortical signaling. This imbalance between top-down/bottom-up information flow may be one of the potential mechanism for the decoupling of consciousness from environmental connection (Murphy et al., 2011; Funk et al., 2016; Andrillon and Kouider, 2020; Casey et al., 2022b). Recent studies also suggest a role in disconnectedness for specific thalamic nuclei, depressing cortical function and compromising thalamocortical information flow (Liu et al., 2015; Feng et al., 2017). Finally, a mechanistic role for local slow wave activity (SWA)–δ band (1–4 Hz) frequency oscillations—has been proposed for inducing disconnectedness. Local SWA has been found in all those states that present episodic coupling of conscious experience with disconnection from the environment: SWA was recorded during REM sleep at the level of primary sensory and motor cortices (Funk et al., 2016) and has long been found in NREM sleep and anesthesia (Murphy et al., 2011). In propofol anesthesia, SWA saturation has been reported to deactivate the thalamus and primary cortices, interrupting wake-like brain activity to external stimuli, thus probably inducing a state of disconnectedness (Mhuircheartaigh et al., 2013).

Here, we propose to identify unresponsive connected and disconnected dream-like states by delivering auditory stimuli during propofol-induced mild sedation and serially awaking healthy participants. We will collect subjective reports about mental activity prior to awakening (assessing dreaming/consciousness) and stimulus perception (assessing connectedness), while ensuring unresponsiveness throughout the experiment and minimizing the risk of arousals. The cerebral activity of participants will be recorded by means of functional MRI. During the auditory stimulation session, we will play series of sounds following the oddball rule, in which trains of beeps of the same frequency (i.e., standard sounds) are occasionally interrupted by a beep of a different frequency (i.e., the deviant or "oddball" sound). This way, we will be able to investigate not only the difference in the perception of sounds during connected and disconnected consciousness, but also whether standard and deviant sounds are processed differently in the two conditions.

Capitalizing on the enhanced spatial resolution of BOLD fMRI, we will (1) characterize stimulus processing within several thalamic nuclei, and primary and secondary cortices during connected (CC) and disconnected consciousness (DC). To this end, we will conduct a hypothesis-driven ROI analysis, in which we will test various hypotheses on the involvement of thalamic nuclei (*Hypothesis 1*), primary auditory cortex (*Hypothesis 2*) and secondary areas (*Hypothesis 3*) in the processing of auditory stimuli during CC and DC states. The ROI analysis will be complemented by an exploratory whole-brain analysis aimed at identifying other regions potentially involved in sensory disconnection. If *Hypothesis 1–2* prove true, it would indicate that already at a basic level of stimulus processing there is a difference between CC and DC. To investigate this further, we aim (2) to characterize changes in functional connectivity between the thalamus and primary auditory cortices with the rest of the brain in CC and DC. We hypothesize (*Hypothesis 4*) that the functional connectivity between the thalamus [e.g., the pulvinar (Kanai et al., 2015; Sanders et al., 2021)] and primary auditory cortices will be stronger in CC compared to DC. If this hypothesis is confirmed (together with Hypothesis 1–2) it would suggest that some gating mechanism already occurs at the thalamic level. However, this difference in brain activity might be necessary but not sufficient to cause sensory disconnection: that is, a weakened connection between the thalamus and primary auditory cortices does not imply that sensory stimuli are *entirely* blocked from cortical processing via the gating action of the thalamus [thalamic gate hypothesis (Andrillon and Kouider, 2020)]. It would however indicate that differential processing of stimuli in CC and DC already occurs at the thalamic level. This thalamic modulation of sensory inputs could in turn affect cortical processing, leading to a cortical gate, that is, loss of information propagation to higher cortices due to a disruption in functional connectivity (cortical gating hypothesis). In this respect, we hypothesize (*Hypothesis 5*) that the functional connectivity between primary and secondary auditory cortices will be stronger in CC compared to DC. Voxel-to-voxel functional connectivity analysis will be conducted for standard and deviant sounds separately: we hypothesize (*Hypothesis 6*) that different processing for deviant and standard sounds will be present in CC but absent in DC. Finally, we aim (3) to quantify high-frequency BOLD oscillations, which have been shown to track sleep slow waves (Song et al., 2022), in selected thalamic nuclei and in primary and secondary cortices. We hypothesize (*Hypothesis 7*) that high-frequency BOLD oscillations will be lower in high-order/first-order thalamic nuclei and in primary and secondary sensory cortices during CC compared to DC.

In summary, we propose an fMRI experiment that systematically differentiates connected and disconnected conscious states by delivering auditory stimuli and serially awakening participants sedated with propofol to assess the conscious state and stimulus perception through subjective reports. Through activation and connectivity analyses of collected fMRI data, we will investigate whether sensory disconnection is caused by altered activity at the level of thalamus, primary regions, or, higher up, due to a lack of stimulus integration in associative areas. This project will provide fundamental insights on the neural correlates of sensory disconnection.

2 Methods

2.1 Inclusion and exclusion criteria

Participants will be screened through an online form, an in-person interview and a medical examination. The initial phase of screening using the online form will select healthy, right-handed, non-smoking, MRI-compatible subjects without psychiatric and neurological disorders, propensity for nausea, recurrent nightmares, memory and hearing impairments, substance abuse, cannabis use in the three months preceding the study and regular alcohol consumption (i.e., everyday). During the in-person interview it will be verified that the above inclusion/exclusion criteria are fulfilled to avoid oversights or errors in filling out the form. Furthermore, we will select participants who have a low risk of obstructive sleep apnea through the StopBang questionnaire (Low Risk: Yes to 0–2 questions) (Chung et al., 2008, 2012) and low levels of anxiety through the scales "Novelty Seeking" and "Harm Avoidance" of the Temperament and Character Inventory self-rating questionnaire (Cloninger et al., 1993). We will only include participants with average or above average scores on the "Novelty Seeking" scale (i.e., $\geq$16.5 for men and $\geq$16.3 for women) and participants with average or below average scores on the "Harm Avoidance" scale (i.e., $\geq$14.5 for men and $\geq$17.5 for women) (Pélissolo and Lépine, 2000). We control for obstructive sleep apnea because of the known respiratory depression effects of propofol. We take into account the predisposition to anxiety as anxious participants might require higher propofol concentrations to achieve loss of

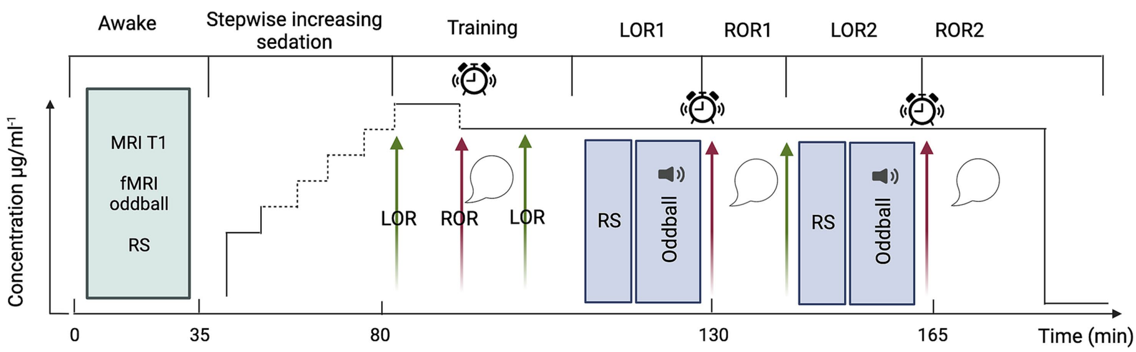

FIGURE 2
Schematic representation of the experimental design. Green arrows indicate loss of responsiveness; red arrows, regain of responsiveness; sound symbol, auditory oddball stimulation; alarm clock, awakening attempt; speech bubble, collection of subjective reports. fMRI acquisitions will be conducted throughout LOR1 and LOR2 and concluded prior to the initiation of the awakening attempt. Created with BioRender.com.

responsiveness compared to non-anxious participants, impairing subsequent recovery of responsiveness and report collection. Participants who meet these criteria will be visited by an anesthesiologist with an evaluation similar to a pre-surgical examination, including a physical examination, full review of the patient's medical, surgical and allergological history, treatment and intubation score and any potential contraindication to propofol sedation. Finally, alcohol consumption will be forbidden for 48 h preceding the experiment. A proper sleep hygiene is encouraged in the 2 to 3 days prior to the study. Participants are required to refrain from drinking and eating six hours before the start of the experiment.

2.2 Experimental setup

In short, the present experiment will comprise four main phases (schematic representation in Figure 2): (1) acquisition of MRI data in awake participants during rest and auditory stimulation, (2) gradual sedation with propofol (~45 min), (3) training session of sedated participants for ~20 min (see below for an explanation) and (4) acquisition of MRI data in sedated participants during rest and auditory stimulation, both repeated twice. During the awake phase a structural (T1) image will be collected prior to the acquisition of functional scans during rest (~10 min) and auditory stimulation (~15 min; for a detailed description of the auditory paradigm, see the section "fMRI experimental design and auditory paradigm"). To prevent the comparison of connected and disconnected consciousness from being contaminated by correlates of (un)responsiveness we will ensure that participants were unresponsive prior to the start of the experimental sessions and awakening. We will also assess their state of wakefulness before and after the experimental sessions to minimise the risk of arousals and confounding correlates of wakefulness with those of (dis)connected consciousness.

From the beginning of the sedation, to monitor responsiveness, participants will be instructed to perform a continuous task of alternately pressing the left and right keys of a box-shaped keypad. Propofol will be infused until loss of volitional motor activity which will be used as a proxy for loss of responsiveness (LOR). LOR will be defined as three consecutive minutes in which the participant has stopped pressing keys and neither speaks nor moves spontaneously.

Propofol will be administered by a computer-controlled continuous infusion (target-controlled infusion—TCI) using a pharmacokinetic model to achieve stable plasma and effect-site propofol concentration (Schnider et al., 1999). The initial target for induction will be set at 1 µg.ml^{-1}, and progressively increased (waiting five minutes between each increase in concentration) until LOR as follows: from the initial target, propofol concentration will be increased by a step of 0.5 µg.ml^{-1} to 1.5 µg.ml^{-1}; if LOR will not be reached at 1.5 µg.ml^{-1}, propofol will be increased by a step of 0.2 µg.ml^{-1} until 1.9 µg.ml^{-1}. If LOR will not be reached at 1.9 µg.ml^{-1} propofol will be increased by steps of 0.1 µg.ml^{-1} until LOR. This will ensure to determine, for each subject, the precise concentration at which LOR occurs. This approach minimizes the risk to exceed the dose required to reach LOR. The higher the dose, the more difficult the recovery of responsiveness may become. Since the aim is to target LOR and not loss of consciousness, the maximum propofol concentration will be set at 4 µg.ml^{-1}. This sedative dosage will also allow the participants to remain in spontaneous ventilation. When participants lose responsiveness, we will wait 5 min for the drug to stabilize, and then we will begin the training session. The goal of the training session is to fine-tune the propofol concentration to maximize the chances of having both LOR and intelligible reports upon regain of responsiveness (ROR) at the same propofol concentration. During the training session (outside the MRI scanner bore, but on the MRI table), we will attempt to awaken participants by performing an arousal protocol that will consist of (1) calling the volunteer aloud through the MRI microphone for up to 2 times, (2) if unsuccessful, lightly shaking volunteer's shoulders for up to 2 times, (3) if unsuccessful, applying moderate painful stimulation, i.e., pinching the skin of the forearm for up to 2 times. If, after three runs of the protocol, participants are still not responsive, we will decrease the propofol concentration by 0.1 µg.ml^{-1}, wait five minutes and repeat the arousal protocol a second time. This process is repeated until participants are able to recover responsiveness. Once responsiveness has been recovered, participants will be asked questions (see "interviews following ROR") to verify their state of (dis)connectedness and (un)consciousness. If participants are unable to speak intelligibly, the concentration of propofol will be reduced by steps of 0.1 µg.ml^{-1} (always waiting 5 min between steps) until reports become intelligible. Once the right concentration has been identified, we will wait for participants to spontaneously lose responsiveness again, and we will end the training session with participants unresponsive (i.e., LOR1).

LOR1 will mark the beginning of the experimental session. After LOR1, we will acquire 10 min of resting-state (RS) fMRI followed by 15 min of task-based fMRI, i.e., passive listening to a sequence of oddball auditory stimuli. After concluding the fMRI acquisition and without altering the drug concentration, we will attempt to awake participants by performing the arousal protocol described above. If the participant does not regain responsiveness after one execution, the protocol will be repeated a maximum of two times. If the participant does not regain responsiveness after three executions, the participant will be defined unarousable. In case of ROR, participants will be questioned about their experience during the period of unresponsiveness to determine whether they were (un)conscious and (dis)connected (see "interviews following ROR"). After the first ROR (ROR1), we will wait a maximum of 10 min for the participant to spontaneously lose responsiveness a second time (LOR2). If the participant does not lose responsiveness in the 10 min following ROR1, we will increase the propofol concentration up to three times in 0.1 $\mu g.ml^{-1}$ increments. If, after increasing the concentration, the participant does not reach LOR2, the LOR attempt will be considered unsuccessful, and the experiment will be terminated. In the case that LOR2 succeeds, the procedure described above for ROR1 will be repeated for ROR2. Finally, the end of the experiment is marked by termination of the drug infusion. This protocol has been fine-tuned based on previous work of our team (Bonhomme et al., 2016).

2.3 Interviews following ROR

After every successful ROR, participants will be subjected to the following 6-question interview (and provided with the following possible answers):

1 Did you have any sensations or thoughts before you were awakened? Yes/No/Not sure/Nothing
2 Did you hear the tones? Yes/No/Not sure
3 Do you think you were awake, having a dream or unconscious? Awake/Dream/Unconscious/Not sure
4 Did you hear one or two different tones? One/Two/Not sure/Nothing
5 Was this experience more centered on yourself or on the environment? Myself/Environment/Nothing
6 Did you rather think, or did you see many things? Think/See/Not sure/Nothing

The first and third questions will verify that participants were (un)conscious during the period of unresponsiveness and the second question will verify the (dis)connectedness of participants during the period of unresponsiveness. The last three questions will serve two purposes: to gather more detailed information on the experience during the period of unresponsiveness and to check the consistency of the reports. Probing the presence of the experience/perception of sounds will strengthen the reliability of the answers to the first two questions (e.g., the participant might answer that he/she did not hear any sounds, but then answer the question "did you hear one or two tones?" with "two"). This interview will identify four different states: (1) awakening without any recall of experiences; (2) connected dreaming; (3) disconnected dreaming and (4) wakefulness. State 1 will be discarded as it cannot be classified in either of the two categories of interest in this study. Connected consciousness (i.e., connected dreaming) will be considered to have occurred during the unresponsive period if participants answered "Yes" to questions number 1–2 and disconnected consciousness if participants replied "Yes" to question 1 and "No" to question 2. In the case of conflicting answers, we will consider participants to have been connected conscious if they provide a positive response to at least one question amongst numbers 1,5,6 or if they respond with "having a dream" to question number 3, in addition to a positive response to at least one question amongst numbers 2 and 4. Participants responding with "awake" to question number 3 will be excluded, as they will be considered not to be in a state of connected consciousness during a dream-like state but rather awake. If participants replied with "No/Nothing" to question number 2–4 and positively to at least one of the questions investigating their experience (i.e., question number 1,3,5,6) they will be considered having been disconnected conscious during the unresponsive period. Participants will be acquainted in advance with the different questions and possible answers to ensure full understanding of each question. To rule out potential arousals during fMRI acquisitions, subjects will be monitored continuously throughout the acquisition via an eye-tracking camera (EyeLink 1000plus system from SR Research, Ltd) – the eye-tracker will be used for online monitoring but not for offline analysis. In case of eye opening, MRI acquisition will be interrupted, and we will wait for participants to spontaneously fall unresponsive again. If, after 15 min, LOR does not occur, we will increase the propofol concentration up to three times in 0.1 $\mu g.ml^{-1}$ increments.

2.4 fMRI experimental design and auditory paradigm

We chose a mixed block/event-related design (Figure 3) in which trials of auditory stimuli are interspersed with blocks of silence of varying durations (15 blocks in total, each lasting 45 s and containing 30 trials on average). This design allows for the simultaneous modelling of the transient, trial-related activity, and the sustained, task-related BOLD activity. That is, by alternating silence blocks with task blocks (i.e., blocks with trials with auditory stimuli) we can optimize the sensitivity for discriminating events within trials and all events combined within a block. Blocks with only standard events will be alternated with blocks containing both standard and deviant events. Standard and deviant blocks will alternate in pseudorandomized order (ABBA…), in which no more than two identical types of blocks can follow one another. The length of the silence blocks will be randomized in 1-s steps in intervals of 7–10 s. Based on previous studies (Bekinschtein et al., 2009), each event (both standard and deviant sounds) will lasts 0.05 s; inter-stimulus interval (ISI) will be fixed at 0.1 s and inter-trial interval (ITI) will be jittered in 0.05-s steps between 0.7 to 1 s. We will use a variant of the "classic oddball" paradigm (Figure 3), in which trials will consist of a randomized number of repetitions (i.e., 3–5) of standard events plus one deviant event. Standard sounds will have a frequency of 100 Hz and deviants of 500 Hz. The total length of the auditory stimulation will be 15 min. Each task block will last 45 s (to have both a frequency still below the recommended 128 s high-pass filter, but also a

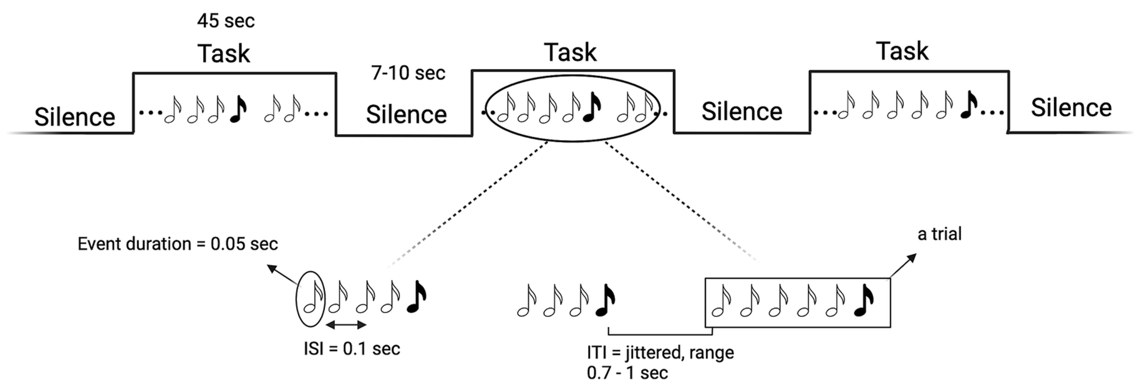

FIGURE 3
Schematic representation of the mixed block/event-related design (top) and of the "Mixed35" oddball rule (bottom). White musical notes denote standard sounds, black notes deviant sounds. Created with BioRender.com.

reasonable number of trials). The parameters selected for the auditory stimulation are the result of the efficiency and collinearity analyses we performed to optimize the efficiency of our design. Sounds will be delivered via a Serene Sound Digital MRI-compatible system.

2.5 Efficiency and collinearity analyses for optimizing the auditory paradigm to fMRI

With the following analyses, we sought to maximize the efficiency of the auditory paradigm which led to the protocol described in the section "fMRI experimental design and auditory paradigm." Estimation of the efficiency can be defined as "a measure of the reliability with which model parameters are estimated"(Mechelli et al., 2003). The efficiency of a design strongly "affects the sensitivity with which experimental effects are detected" (Mechelli et al., 2003). In order to find the (*a priori*) most efficient design to detect our effects of interest, we manipulated the temporal distribution of events, resulting in several designs whose efficiency was estimated *a priori* and then compared. The variables manipulated were the length of the silence blocks (i.e., randomized in 1-s steps in intervals of 7–20 s, 7–10 s, 15–20 s, 10–15 s), the ordering of standard and deviant blocks (i.e., interleaved order or pseudorandomized order in which no more than two identical types of blocks can follow one another), the type of oddball paradigm. We selected four different types of oddball paradigms for comparing their efficiency, of which only one was chosen for the experiment. In the "classic oddball" paradigm (Squires et al., 1975), trials consist of four standard sounds and one deviant, where the deviant is defined by a change in frequency. The "roving oddball" (Garrido et al., 2007) differs with each trial presenting sounds of the same frequency and starting a new trial with a different frequency, making the first event of a trial a deviant. We also designed two "mixed oddballs," in which trials follow the "classic oddball" rule but with the difference that the number of repetitions of standard events is randomized, between three and five ("Mixed Oddball35") or between three and seven repetitions of standards ("Mixed Oddball37"). The efficiency analysis was conducted by comparing all possible combinations of parameters (i.e., ISI, ITI, stimulus and block duration) for the different oddball paradigm designs. Please note that the efficiency calculation is related to the number of scans (i.e., to a given TR and duration of experiment), and specific to a given contrast. We calculated the efficiency for TR = 0.842 s, 900 s duration of the experiment and for the following three contrasts: main effect of the standard response, main effect of the deviant response and the difference between standard and deviant responses. For more information on how we computed efficiency, see our repository on GitHub "Efficiency-Analysis-fMRI-mixed-design," where each step of the analysis is detailed: https://doi.org/10.5281/zenodo.8117861. In addition to the efficiency analysis, we also performed a collinearity analysis in SPM to estimate the extent to which our two events (standard and deviant) were collinear (i.e., whether their responses correlated with each other)—see our GitHub repository for more details on how to compute collinearity in SPM. The design that was most efficient and with least collinearity was the "Mixed Oddball35" with 7–10 s silence blocks and pseudorandomized order (mean efficiency for the difference between standard and deviant events = 0.843). As depicted in Figure 4, "Mixed Oddball35" was found to have comparable efficiency with the "Classic oddball" (mean efficiency for the difference between standard and deviant events = 0.845). We chose the "Mixed Oddball35" for the experiment as randomizing the number of repetitions of the standard sounds has the advantage of decreasing expectation.

2.6 MRI data collection

MRI data will be collected with a 3 T Magnetom Prisma Fit scanner (Siemens, Erlangen, Germany) equipped with a 20-channel array receiver head–neck coil. For rs-fMRI and task based-fMRI, the scanning parameters will be as follows: echo-planar imaging (EPI) with multi-band acceleration factor of 6, 7/8 phase partial Fourier, 2.25 mm slice thickness, no gap between slices, 2.25 mm x 2.25 mm in-plane spatial resolution, 842 ms repetition time (TR), 30 ms echo time (TE), 52° flip angle, 207 mm x 225 mm field of view (FOV) and a matrix size of 92 × 100. For anatomical reference, a high-resolution T1-weighted image will be acquired for each subject during the awake session (T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE) sequence, TR = 1900 ms, TE = 2.19 ms, inversion time (TI) = 900 ms, sagittal orientation, 224 slices, 1 mm slice thickness, FoV = 256×240 mm^2, matrix size = 256x240x224, voxel

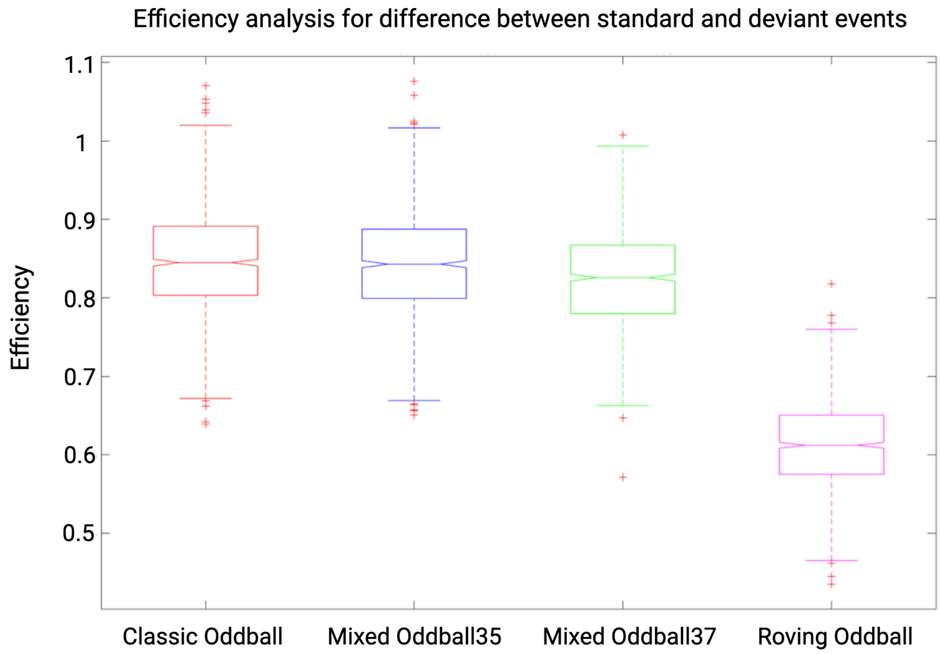

FIGURE 4
Efficiency results for four auditory paradigms. Results shown are for contrast standard-deviant, TR = 0.842 and for an experiment duration of 900 s (i.e., 1,125 scans). Efficiency values are reported for "Classic oddball," "Mixed Oddball35," "Mixed Oddball37," and "Roving oddball" with silence blocks of duration 7–10 s and in pseudorandomized order (ABBA). Whiskers corresponds to approximately +/−2.7σ and 99.3 percent coverage. We selected the "Mixed Oddball35" for its combination of high efficiency and the reduction of expectation achieved through the randomization of standard sounds.

size = 1x1x1 mm^3, GRAPPA R = 2 acceleration factor in phase-encoding direction (AP).

2.7 Variables of interest, randomization, and blinding

In this experiment, the dependent variable will be the BOLD signal (during RS/task sessions). The independent variables will be both the connected/disconnected condition and the type of sounds delivered, i.e., standard or deviant events or silence. The connected/disconnected condition cannot be randomized, given its unpredictability, i.e., it is impossible to predict which individuals will report being connected and which will report being disconnected. In contrast, several parameters of the auditory paradigm will be randomized (e.g., the number of repetitions of standard events). Similarly, the temporal order of the waking and sedation sessions cannot be randomized given the use of anesthetics: data acquisition after the end of sedation would in fact correspond more to acquiring data during the recovery than during the awakening phase. Finally, the order of the RS and task sessions will not be randomized due to time constraints: since the collection of reports must occur immediately after the task session, if the task preceded the RS session, we would be forced to collect reports after the task and then wait for the participants to lose responsiveness again to acquire the 10 min of rs-fMRI, considerably extending acquisition time. Data preprocessing and analysis will be performed blind to the conditions of the experiment.

2.8 Sample size calculation

To the best of our knowledge, no study has investigated cerebral changes between connected and disconnected consciousness using fMRI. Hence, no effect size was available in the literature for a power calculation with a similar setup as the current experiment. However, we were able to make an approximate estimate of the total sample size required based on EEG studies investigating sensory disconnection during propofol anesthesia (Casey et al., 2022a), REM (own data, to be submitted) and NREM sleep (own data, to be submitted). We estimated the effect size with Cohen's *d* from the means and standard deviations of each group, and we performed power calculations (two tailed t-tests) by setting the desired α at 0.05 and power at 0.95. In the first study conducted under propofol sedation, the effect size of the difference between CC and DC in occipital delta power was 1.0 and the allocation ratio N2/N1 was 2.78. In the REM and NREM sleep studies, the effect size of the difference between CC and DC in event related potentials was 0.85 with allocation ratio N2/N1 = 1 and 0.94 with allocation ratio N2/N1 = 0.71, respectively. The total estimated sample size is of 70 sessions according to power calculations based on the effect size of the propofol study; of 64 sessions based on the effect size of the NREM study and of 74 sessions, based on the REM study's effect size (see Figure 5). Taking the most conservative estimate based on the smallest effect size of the three studies, and accounting for an 8% dropout (e.g., impossible to reach LOR or ROR), we plan to collect 40 subjects, for a total of 80 sessions (two per subject). This power analysis was conducted in G*Power 3.1.9.7 software (Faul et al., 2007, 2009) and the results plotted in MATLAB.

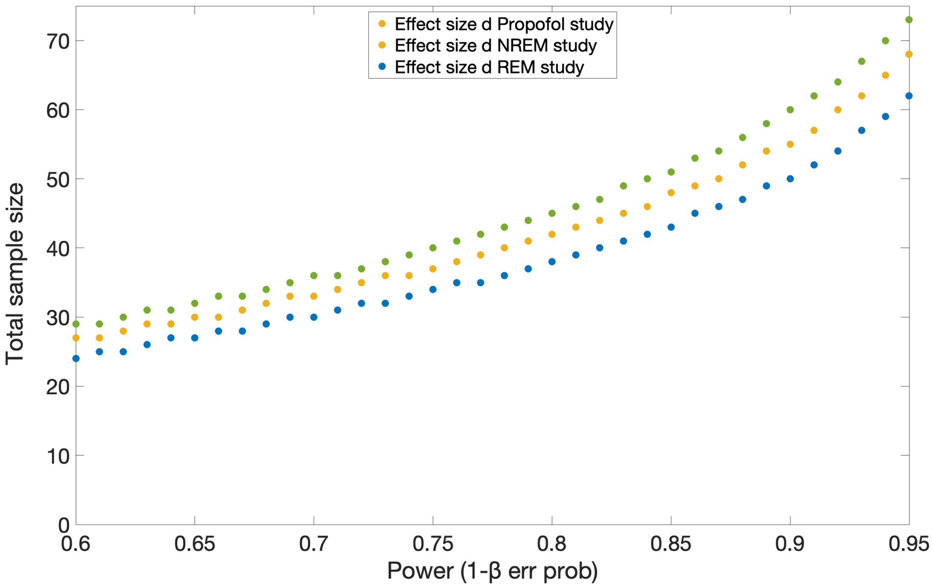

FIGURE 5
Results of the power analysis based on the effect sizes obtained from three studies investigating the difference between connected and disconnected consciousness using EEG. Estimated total sample sizes for power 0.6–0.95 are shown for each of the effect sizes reported in the studies conducted under propofol sedation [yellow circles; Casey et al. (2022a)], during NREM (blue circles) and REM (green circles) sleep. α err prob. = 0.05.

TABLE 1 Hypotheses tested and corresponding analyses. This table provides a comprehensive overview of the analyses, brain areas, atlas employed, and statistical methods used for each hypothesis tested in this study.

Hypotheses tested	Analyses	Considered brain areas	Atlas	Statistical analysis
Hypothesis 1	Activation analysis (ROI-based)	All 7 sub-thalamic regions (L/R)	Oxford thalamic connectivity atlas	Small volume correction
Hypothesis 2	Activation analysis (ROI-based)	Primary auditory cortex (L/R): Heschl's gyrus	Harvard-Oxford cortical and subcortical structural atlases	Small volume correction
Hypothesis 3	Activation analysis (ROI-based)	Secondary regions (L/R): superior temporal gyrus (both anterior and posterior division)	Harvard-Oxford cortical and subcortical structural atlases	Small volume correction
Hypothesis 4	ROI-to-ROI based connectivity	2 sub-thalamic (sensory and temporal) nuclei and Heschl's gyrus	Oxford thalamic connectivity atlas/ Harvard-Oxford cortical and subcortical structural atlases	Subject-level: wGLM Group-level: LME
Hypothesis 5	ROI-to-ROI based connectivity	Heschl's gyrus and superior temporal gyrus (both anterior and posterior division)	Harvard-Oxford cortical and subcortical structural atlases	Subject-level: wGLM Group-level: LME
Hypothesis 6	Voxel-to-voxel connectivity	Whole cortical and subcortical brain matters	No atlas	Subject-level: Intrinsic connective analysis Group-level: LME
Hypothesis 7	Spectrogram analysis	All 7 sub-thalamic regions (L/R); Primary auditory cortex (L/R): Heschl's gyrus; Secondary regions (L/R): superior temporal gyrus (both anterior and posterior division)	Oxford thalamic connectivity atlas; Harvard-Oxford cortical and subcortical structural atlases; Harvard-Oxford cortical and subcortical structural atlases	Subject-level: Short Time Fourier Transform (STFT) for regional spectrogram estimation. Group-level: LME

2.9 Preprocessing and statistical analysis

As we will use a mixed block/event-related design, all the following analyses (see Table 1 for a summary) will be conducted on the BOLD time-series at both the block and trial level: for the block-level analyses, we will contrast (1) task blocks (i.e., standard and deviant blocks combined) with silence blocks and (2) standard vs. deviant blocks. For the trial-level analysis, we will contrast (1) combined deviant and standard trials with BOLD extracted time-series of simulated time points during silence blocks and (2) deviant vs. standard trials. For the purpose of this project and the hypotheses presented, the analysis of RS data will not be discussed in this study.

2.9.1 Preprocessing

fMRI data will be preprocessed in software SPM12 (Statistical Parametric Mapping, version 12, UCL Institute of Neurology, London, Britain, http://www.fil.ion.ucl.ac.uk/spm) and FSL 6.[3] The preprocessing pipeline will include standard steps of realignment, susceptibility-induced distortions correction [FSL topup (Andersson et al., 2003; Smith et al., 2004)], slice acquisition time correction, coregistration, brain tissue segmentation, spatial normalization to the Montreal Neurological Institute stereotaxic template and smoothing using the Gaussian filter method with an isotropic kernel of size 6 mm. Outlier volumes, due to excessive head and body motion, will be detected using Artifact Detection Tools (ART) toolbox[4] and will be regressed out in the first-level general linear model (GLM) analysis. An image will be defined as an outlier image or artifact if the absolute head displacement in the x-, y-, or z-direction is equal or greater than 2.3 mm, if the framewise displacement is greater than 0.4 mm, or if the overall average image intensity is greater than 3 standard deviations from the average intensity of the rest of the images. For each run of the 15-min auditory paradigm, the first five volumes will be discarded to allow magnetization to reach dynamic equilibrium.

2.9.2 Activation analyses (ROI, Hypothesis 1-2-3)

First-level activation analysis will be performed using SPM 12 for block/trial levels. Activation values inside each ROI will be estimated using the Small Volume Correction (SVC) technique. The ROIs will include primary auditory cortex (Heschl's gyrus (HG), including HG1 and HG2) and secondary auditory cortex (planum polare and planum temporale) extracted from the "Harvard-Oxford cortical and subcortical structural atlases"[5], and Thalamic ROIs extracted from the "Oxford thalamic connectivity atlas," in which sub-striatal regions are segmented according to their white-matter connectivity to cortical areas (Behrens T. et al. 2003; Behrens T. E. J. et al. 2003). We will include the areas of the thalamus labeled in the atlas as posterior-parietal, occipital, sensory, and prefrontal. First-level GLM design matrix will include six movement parameters and outlier volumes. Additionally, we will model the effect of elapsed time since awakening by adding a regressor based on the onsets of deviant and standard trials/blocks. The analysis will be considered significant at an alpha of <0.05, corrected for the number of ROIs.

2.9.3 Connectivity analyses (ROI-ROI, Hypothesis 4-5)

Before carrying out the connectivity analysis, functional data will be denoised using a standard denoising pipeline (Nieto-Castanon, 2020) including the regression of potential confounding effects characterized by 5 principal components of white matter and cerebrospinal fluid using the component-based noise correction method (CompCor), 6 motions parameters and their 6 motions parameters and their first order derivatives, outliers volumes, session and task effects and their first order derivatives and linear trends within each functional run, followed by high pass filtering above 0.008 Hz. To assess task-related functional connectivity changes across experiment blocks/trials, the pair-wise ROI-ROI connectivity will be estimated as bivariate correlations using a weighted general linear model (wGLM) as implemented in CONN. The boxcar signal characterizing each block/trial, convolved with an SPM canonical hemodynamic response function and rectified, will be used to weight the ROI signals in the wGLM model. This will lead to block/trial-specific between-ROI correlation coefficients which will be then Fisher-transformed for further analysis. The ROIs will include two sub-thalamic (sensory and temporal) nuclei and the Heschl's gyrus (hypothesis 4) and superior temporal gyrus (both anterior and posterior division; hypothesis 5). The results of this step would be connectivity matrices at the subject-level showing the correlation between the defined ROIs.

2.9.4 Connectivity analyses (voxel-voxel, Hypothesis 6)

Data will be denoised as described in the previous paragraph. An exploratory analysis will be conducted at the whole-brain level, estimating the intrinsic connectivity maps related to each block/trial as implemented in CONN. This parameter characterizes network centrality at each voxel as the root mean square (RMS) of all short- and long-range connections between a voxel and the rest of the brain (Martuzzi et al., 2011).

2.9.5 Group-level analysis

For group-level activation analysis (block/trial) we will perform a linear mixed effect model with nested random effects of subjects within the two experimental sessions, using package lme4 in software R. Fixed effects will be included to control for condition (CC vs. DC), stimulus/block type (deviant vs. standard or silence vs. sound), interaction between stimulus/block type and condition, propofol concentration and time to ROR. The latter allows the investigation of the effect of extending reports back in time, i.e., it is more probable that the subject was in a CC/DC state shortly before awakening as opposed to a long time before awakening. For the group-level connectivity analysis the same procedure will be followed, considering the connectivity matrices or intrinsic connectivity maps as dependent variables.

2.9.6 fMRI spectral analysis

The progressive change in BOLD frequency content during the connected and disconnected consciousness conditions will be assessed using Short Time Fourier Transform. A sliding Hamming window will be used to calculate the spectrogram of each voxel's time series (in order to check the validity of the results, we will repeat the analysis with variable window lengths of 50, 100, and 200 s corresponding to 60, 118, and 238 volumes, respectively). For each ROI, the BOLD power spectrogram will be calculated by averaging the power spectrogram across all voxels within the region. To ensure that different voxels within a brain region contributed equally to the power spectrogram of the region, the power spectrogram of individual voxels will be normalized by its total power. After estimating the ROI spectrograms, the peak frequency at each time point will be estimated based on the method introduced in Song et al. (2022). The time series showing the peak frequencies will be compared between the

3 https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
4 https://www.nitrc.org/projects/artifact_detect/
5 https://identifiers.org/neurovault.collection:262

connected and disconnected consciousness conditions. This analysis will be performed in the ROIs used for hypotheses 1-2-3.

3 Data management and dissemination

Data will be stored in the NIfTI format, organized according to brain imaging data structure (BIDS) (Gorgolewski et al., 2016) and pseudo anonymized by an identification number, identifiable only by the researchers involved in the study. The results of the present study will be published in peer-reviewed journals as original research articles and will be presented at various scientific conferences. In these publications, the privacy of the individuals who took part will be safeguarded through anonymization.

4 Ethics

The study was approved by the University of Liege Hospital Ethics Committee (2020–707) and was registered at the European Union Drug Regulating Authorities Clinical Trials Database (identifier: 2020-003524-17). All study subjects will be informed in writing of the objectives, methods and potential risks of the experiment. They will be given two documents: a general information form on MRI acquisition and a specific form containing information on the study itself. All participants will provide written informed consent according to the Declaration of Helsinki and will receive financial compensation (300 euros). To ensure participant's safety, vital parameters will be continuously monitored, and an anesthesiologist will be present in the MRI room for the entire duration of the experiment. Subjects will receive additional oxygen through a plastic facemask at a rate of 3 L. min^{-1}. Monitored vital signs will include EKG and heart rate, non-invasive blood pressure, peripheral saturation in oxygen, inspired and expired CO_2, thoracic movements amplitude, and respiratory rate. All material and medications needed to ensure safety of the sedation will be immediately available.

5 Discussion

Identifying the neuroimaging signatures of disconnected consciousness during sleep or anesthesia is a particularly difficult undertaking, owing to the inherent challenges in ascertaining the kind of experience (or lack thereof) a sleeping or sedated subject is having. This work aims to advance the investigation of the neural basis of sensory disconnection by achieving a more precise identification of this state, distinguishing it from states of unconsciousness and connected consciousness. In classical anesthesia studies, differentiation between these states was typically overlooked, with anesthetized subjects categorized as either conscious or unconscious based on behavioral responsiveness or explicit recall after anesthesia. In this study, we will overcome previous limitations by awakening participants immediately after auditory stimulation and asking them whether they were connected or disconnected before being awakened. This procedure, by minimizing amnesic effects of anesthesia and relying on subjective reports rather than (un)responsiveness to ascertain the subject's state, enables a more accurate account of the subject's experience.

This work represents also a notable progression in mitigating biases linked to contrastive, between-state paradigms, wherein two physiologically distinct states are compared (Koch et al., 2016; Boly et al., 2017). In this study, we ensure that participants are unresponsive and are in a dream-like state before awakening, in both connected and disconnected consciousness conditions. As a result, the conditions of connected and disconnected consciousness are contrasted within the same physiological state, i.e., both connected and disconnected participants are unresponsive and in a dream-like state.

Notwithstanding the surmounted challenges, certain methodological limitations within the present study design need to be addressed. As stated above, the differentiation between CC and DC is based on subjective reports. The limitations of introspection and thus of subjective reports to verify the state of consciousness have been widely discussed (Irvine, 2012; Tsuchiya et al., 2015). Objective measures of awareness, in which awareness is inferred from (above-chance) performance on a task, have often been advanced as more accurate and reliable measures for tracking changes in experience. However, it has been remarked that objective measures, instead of capturing subjective experience, only capture performance in the task, as below-chance performance does not necessarily imply that the subjects were unaware (Lau, 2008; Ellia et al., 2021). Which measure of awareness is best remains an open question at present. Regardless, our experimental design does not lend itself to the use of objective measures as participants are sedated and expected to remain unresponsive during the auditory stimulation, which makes task performance unfeasible. Furthermore, because the collection of subjective reports occurs during propofol sedation, the amnesic effects of anesthesia could lead participants to forget the experience they were having during the unresponsive period, therefore biasing the reports. Indeed, the absence of dream reports upon awakening does not necessarily imply unconsciousness or disconnectedness (Windt et al., 2016). At present, collecting retrospective reports is the only way to access participants' subjective experience during unresponsive periods such as sleep or sedation. Amnesic effects on subjective reports are, however, significantly reduced the closer they are collected to the experience under investigation. Indeed, compared to post-anesthesia collection of reports, the serial awakening paradigm has been shown to minimize the lack of explicit recall as subjects are awakened and questioned about the experience they were having immediately before regaining responsiveness (Siclari et al., 2013, 2017). The extent to which reports can be extended back in time is however still unknown: e.g., if the participant reported being connected upon awakening, can we infer that (s)he was in a connected state during the entire 15 min of auditory stimulation or only during the last 5, 3 min or 60 s before awakening? This problem can however be partially accounted for by modelling the effect of time passing on the effects of interest. Finally, another limitation of our study pertains to its generalizability to alternative sensory modalities and different anesthetic agents. Our study explores sensory disconnection induced by the anesthetic agent propofol within the auditory modality. The extent to which our findings can be extended to other sensory modalities and anesthetics is presently unknown. Future studies incorporating a range of anesthetics and sensory modalities will therefore be necessary to validate and extend the applicability of these findings.

The findings of this study harbor the potential to disclose biomarkers of intraoperative connected consciousness, profoundly revolutionizing the landscape of anesthesia monitoring. A deeper understanding of the mechanisms that underlie states of disconnected consciousness will not only aid in devising strategies to induce it, as necessitated in instances such as surgical procedures, but also to suppress it, as demanded in scenarios like attention lapses, which pose as potential contributors to car accidents. Finally, these findings may be used to improve diagnosis of patients with disorders of consciousness. Understanding the level of consciousness and the cognitive capacities retained by these patients is, in fact, problematic due to their (often) limited responsiveness. Knowing the neural correlates of (dis)connectedness may allow to innovate the procedures of diagnosis and classification of these patients.

Ethics statement

The study was approved by the University of Liege Hospital Ethics Committee (2020-707) and was registered at the European Union Drug Regulating Authorities Clinical Trials Database (identifier: 492 2020-003524-17). All participants will provide written informed consent according to the Declaration of Helsinki.

Author contributions

BC: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Software, Visualization, Writing – original draft. JM: Methodology, Writing – review & editing. SM: Formal analysis, Software, Writing – review & editing. RP: Formal analysis, Methodology, Writing – review & editing. RS: Methodology, Writing – review & editing. CP: Formal analysis, Methodology, Validation, Writing – review & editing. NA: Methodology, Writing – review & editing. ER: Writing – review & editing. AD: Writing – review & editing. MB: Conceptualization, Methodology, Writing – review & editing. MAB: Formal analysis, Writing – review & editing. LL: Methodology, Writing – review & editing. SL: Funding acquisition, Writing – review & editing. OG: Funding acquisition, Methodology, Writing – review & editing. VB: Funding acquisition, Methodology, Supervision, Writing – review & editing. JA: Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study was supported by the Mind Science Foundation-Tom Slick Research Award granted to BC, JA, VB, and SL; the Belgian National Funds for Scientific Research (FRS-FNRS); the Fondation Leon Fredericq ("Crédit classique", "Crédits forfaitaires" and "Bourse de fonctionnement"); the University and University Hospital of Liège; the Department of Anesthesia and Intensive Care Medicine, Liege University Hospital, Belgium. The study was further supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement no. 945539 (Human Brain Project SGA3), the FNRS PDR project (T.0134.21), FNRS - MIS project F.4521.23, FNRS - MIS project F.4521.23, the ERA-Net FLAG-ERA JTC2021 project ModelDXConsciousness (Human Brain Project Partnering Project), the fund Generet, the King Baudouin Foundation, the Télévie Foundation, the Public Utility Foundation "Université Européenne du Travail", "Fondazione Europea di Ricerca Biomedica", the BIAL Foundation, the European Commission, the Mind-Care foundation, the National Natural Science Foundation of China (Joint Research Project 81471100) and the European Foundation of Biomedical Research FERB Onlus. BC is doctoral candidate at the GIGA Doctoral School for Health Science (University of Liege) and FNRS-ASP. LL is supported by the EU Joint Programme Neurodegenerative Disease Research (JPND) (SCAIFIELD project – FNRS reference: PINT-MULTI R.8006.20). OG is Research Associate at FRS-FNRS. CP and SL are Research Directors at FRS-FNRS. SL is supported by the Human Brain Project, National Natural Science Foundation of China, European Foundation of Biomedical Research FERB Onlus, fund Generet of King Baudouin Foundation, Mind Care International Foundation. JA is postdoctoral fellow funded (1265522N) by the Fund for Scientific Research-Flanders (FWO).

Acknowledgments

The authors are grateful to the advice of Giulio Tononi when planning this project. This work was supported by the Human Imaging platform of the GIGA-CRC-IVI at ULiège, Belgium.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Andersson, J. L. R., Skare, S., and Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20, 870–888. doi: 10.1016/S1053-8119(03)00336-7

Andrillon, T., and Kouider, S. (2020). The vigilant sleeper: neural mechanisms of sensory (de)coupling during sleep. *Curr. Opin. Physio.* 15, 47–59. doi: 10.1016/J.COPHYS.2019.12.002

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. USA* 106, 1672–1677. doi: 10.1073/PNAS.0809667106/SUPPL_FILE/0809667106SI.PDF

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the Back of the cerebral cortex? Clinical and neuroimaging evidence. *J. Neurosci.* 37, 9603–9613. doi: 10.1523/JNEUROSCI.3218-16.2017

Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M. A., Noirhomme, Q., et al. (2012). Connectivity changes underlying spectral EEG changes during Propofol-induced loss of consciousness. *J. Neurosci.* 32, 7082–7090. doi: 10.1523/JNEUROSCI.3769-11.2012

Bonhomme, V., Vanhaudenhuyse, A., Demertzi, A., Bruno, M. A., Jaquet, O., Bahri, M. A., et al. (2016). Resting-state network-specific breakdown of functional connectivity during ketamine alteration of consciousness in volunteers. *Anesthesiology* 125, 873–888. doi: 10.1097/ALN.0000000000001275

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., et al. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.*, 50: 1077–1088. doi: 10.1002/mrm.10609

Behrens, T., Johansen-Berg, H., Woolrich, M., Smith, S. M., Wheeler-Kingshott, C. A. M., Boulby, P. A., et al (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat Neurosci* 6, 750–757. doi: 10.1038/nn1075

Casey, C. P., Tanabe, S., Farahbakhsh, Z., Parker, M., Bo, A., White, M., et al. (2022b). Dynamic causal modelling of auditory surprise during disconnected consciousness: the role of feedback connectivity. *NeuroImage* 263:119657. doi: 10.1016/J.NEUROIMAGE.2022.119657

Casey, C. P., Tanabe, S., Farahbakhsh, Z., Parker, M., Bo, A., White, M., et al. (2022a). Distinct EEG signatures differentiate unconsciousness and disconnection during anaesthesia and sleep. *Br. J. Anaesth.* 128, 1006–1018. doi: 10.1016/j.bja.2022.01.010

Chung, F., Subramanyam, R., Liao, P., Sasaki, E., Shapiro, C., and Sun, Y. (2012). High STOP-bang score indicates a high probability of obstructive sleep apnoea. *Br. J. Anaesth.* 108, 768–775. doi: 10.1093/BJA/AES022

Chung, F., Yegneswaran, B., Liao, P., Chung, S. A., Vairavanathan, S., Islam, S., et al. (2008). STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology* 108, 812–821. doi: 10.1097/ALN.0B013E31816D83E4

Cloninger, C. R., Svrakic, D. M., and Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Arch. Gen. Psychiatry* 50, 975–990. doi: 10.1001/ARCHPSYC.1993.01820240059008

Ellia, F., Hendren, J., Grasso, M., Kozma, C., Mindt, G., Lang, J. P., et al. (2021). Consciousness and the fallacy of misplaced objectivity. *Neurosci. Cons.* 2021:niab032. doi: 10.1093/nc/niab032

Errando, C. L., Sigl, J. C., Robles, M., Calabuig, E., García, J., Arocas, F., et al. (2008). Awareness with recall during general anaesthesia: a prospective observational evaluation of 4001 patients. *Br. J. Anaesth.* 101, 178–185. doi: 10.1093/BJA/AEN144

Esser, S. K., Hill, S., and Tononi, G. (2009). Breakdown of effective connectivity during slow wave sleep: investigating the mechanism underlying a cortical gate using large-scale modeling. *J. Neurophysiol.* 102, 2096–2111. doi: 10.1152/JN.00059.2009

Faul, F., Erdfelder, E., Buchner, A., and Lang, A. G. (2009). Statistical power analyses using G*power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146/METRICS

Feng, L., Motelow, J. E., Ma, C., Biche, W., McCafferty, C., Smith, N., et al. (2017). Seizures and sleep in the thalamus: focal limbic seizures show divergent activity patterns in different thalamic nuclei. *J. Neurosci.* 37, 11441–11454. doi: 10.1523/JNEUROSCI.1011-17.2017

Funk, C. M., Honjoh, S., Rodriguez, A. V., Cirelli, C., and Tononi, G. (2016). Local slow waves in superficial layers of primary cortical areas during REM sleep. *Curr. Biol.* 26, 396–403. doi: 10.1016/J.CUB.2015.11.062

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., and Friston, K. J. (2007). Dynamic causal modelling of evoked potentials: a reproducibility study. *NeuroImage* 36, 571–580. doi: 10.1016/J.NEUROIMAGE.2007.03.014

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., das, S., Duff, E. P., et al. (2016, 2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3, 1–9. doi: 10.1038/sdata.2016.44

Irvine, E. (2012). Old problems with new measures in the science of consciousness. *Br. J. Philos. Sci.* 63, 627–648. doi: 10.1093/BJPS/AXS019

Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 370:20140169. doi: 10.1098/RSTB.2014.0169

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/NRN.2016.22

Lau, H. C. (2008). A higher order Bayesian decision theory of consciousness. *Prog. Brain Res.* 168, 35–48. doi: 10.1016/S0079-6123(07)68004-2

Lennertz, R., Pryor, K. O., Raz, A., Parker, M., Bonhomme, V., Schuller, P., et al. (2023). Connected consciousness after tracheal intubation in young adults: an international multicentre cohort study. *Br. J. Anaesth.* 130, e217–e224. doi: 10.1016/J.BJA.2022.04.010

Leslie, K. (2010). "Dreaming during anesthesia" in *Consciousness, Awareness, and Anesthesia*, Ed. George, A. Mashour. Cambridge University Press. 74–89. https://www.cambridge.org/core/books/consciousness-awareness-and anesthesia/5E3B4B566E7AC244045F687F552F3482

Leslie, K., and Ogilvie, R. (1996). Vestibular dreams: the effect of rocking on dream mentation. *Dreaming* 6, 1–16. doi: 10.1037/H0094442

Leslie, K., Skrzypek, H., Paech, M. J., Kurowski, I., and Whybrow, T. (2007). Dreaming during anesthesia and anesthetic depth in elective surgery patients: a prospective cohort study. *Anesthesiology* 106, 33–42. doi: 10.1097/00000542-200701000-00010

Leslie, K., Sleigh, J., Paech, M. J., Voss, L., Lim, C. W., and Sleigh, C. (2009). Dreaming and electroencephalographic changes during anesthesia maintained with propofol or desflurane. *Anesthesiology* 111, 547–555. doi: 10.1097/ALN.0B013E3181ADF768

Linassi, F., Zanatta, P., Tellaroli, P., Ori, C., and Carron, M. (2018). Isolated forearm technique: a meta-analysis of connected consciousness during different general anaesthesia regimens. *Br. J. Anaesth.* 121, 198–209. doi: 10.1016/J.BJA.2018.02.019

Liu, J., Lee, H. J., Weitz, A. J., Fang, Z., Lin, P., Choy, M., et al. (2015). Frequency-selective control of cortical and subcortical networks by central thalamus. *elife* 4:e09215. doi: 10.7554/ELIFE.09215.001

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/SCIENCE.1117256

Mechelli, A., Price, C. J., Henson, R. N. A., and Friston, K. J. (2003). Estimating efficiency a priori: a comparison of blocked and randomized designs. *Neuroimage* 18, 798–805. doi: 10.1016/S1053-8119(02)00040-X

Mhuircheartaigh, R. N., Warnaby, C., Rogers, R., Jbabdi, S., and Tracey, I. (2013). Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans. *Sci. Transl. Med.* 5:208ra148. doi: 10.1126/SCITRANSLMED.3006007

Murphy, M., Bruno, M. A., Riedner, B. A., Boveroux, P., Noirhomme, Q., Landsness, E. C., et al. (2011). Propofol anesthesia and sleep: a high-density EEG study. *Sleep* 34, 283–291. doi: 10.1093/SLEEP/34.3.283

Martuzzi, R., Ramani, R., Qiu, M., Shen, X., Papademetris, X., and Constable, R. T. (2011). A whole-brain voxel based measure of intrinsic connectivity contrast reveals local changes in tissue connectivity with anesthetic without a priori assumptions on thresholds or regions of interest. *NeuroImage*, 58. doi: 10.1016/j.neuroimage.2011.06.075

Nieto-Castanon, A. (2020). Handbook of functional connectivity Magnetic Resonance Imaging methods in CONN. *Hilbert Press*. doi: 10.56441/hilbertpress.2207.6598

Nielsen, T. A. (1993). Changes in the kinesthetic content of dreams following somatosensory stimulation of leg muscles during REM sleep. *Dreaming* 3, 99–113. doi: 10.1037/H0094374

Noreika, V., Jylhänkangas, L., Móró, L., Valli, K., Kaskinoro, K., Aantaa, R., et al. (2011). Consciousness lost and found: subjective experiences in an unresponsive state. *Brain Cogn.* 77, 327–334. doi: 10.1016/J.BANDC.2011.09.002

Pélissolo, A., and Lépine, J. P. (2000). Normative data and factor structure of the temperament and character inventory (TCI) in the French version. *Psychiatry Res.* 94, 67–76. doi: 10.1016/S0165-1781(00)00127-X

Radek, L., Kallionpää, R. E., Karvonen, M., Scheinin, A., Maksimow, A., Långsjö, J., et al. (2018). Dreaming and awareness during dexmedetomidine- and propofol-induced unresponsiveness. *Br. J. Anaesth.* 121, 260–269. doi: 10.1016/j.bja.2018.03.014

Sanders, R. D., Casey, C., and Saalmann, Y. B. (2021). Predictive coding as a model of sensory disconnection: relevance to anaesthetic mechanisms. *Br. J. Anaesth.* 126, 37–40. doi: 10.1016/J.BJA.2020.08.017

Sanders, R. D., Gaskell, A., Raz, A., Winders, J., Stevanovic, A., Rossaint, R., et al. (2017). Incidence of connected consciousness after tracheal intubation: a prospective, international, multicenter cohort study of the isolated forearm technique. *Anesthesiology* 126, 214–222. doi: 10.1097/ALN.0000000000001479

Sanders, R. D., Tononi, G., Laureys, S., and Sleigh, J. W. (2012). Unresponsiveness ≠ unconsciousness. *Anesthesiology* 116, 946–959. doi: 10.1097/ALN.0b013e318249d0a7

Scheinin, A., Kantonen, O., Alkire, M., Långsjö, J., Kallionpää, R. E., Kaisti, K., et al. (2021). Foundations of human consciousness: imaging the twilight zone. *J. Neurosci.* 41, 1769–1778. doi: 10.1523/JNEUROSCI.0775-20.2020

Schnider, T. W., Minto, C. F., Shafer, S. L., Gambus, P. L., Andresen, C., Goodale, D. B., et al. (1999). The influence of age on propofol pharmacodynamics. *Anesthesiology* 90, 1502–1516. doi: 10.1097/00000542-199906000-00003

Schredl, M., Atanasova, D., Hörmann, K., Maurer, J. T., Hummel, T., and Stuck, B. A. (2009). Information processing during sleep: the effect of olfactory stimuli on dream content and dream emotions. *J. Sleep Res.* 18, 285–290. doi: 10.1111/J.1365-2869.2009.00737.X

Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., Riedner, B., et al. (2017). The neural correlates of dreaming. *Nat. Neurosci.* 20, 872–878. doi: 10.1038/nn.4545

Siclari, F., LaRocque, J. J., Postle, B. R., and Tononi, G. (2013). Assessing sleep consciousness within subjects using a serial awakening paradigm. *Front. Psychol.* 4:542. doi: 10.3389/FPSYG.2013.00542/BIBTEX

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23, S208–S219. doi: 10.1016/J.NEUROIMAGE.2004.07.051

Song, C., Boly, M., Tagliazucchi, E., Laufs, H., and Tononi, G. (2022). fMRI spectral signatures of sleep. *Proc. Natl. Acad. Sci. USA* 119:e2016732119. doi: 10.1073/PNAS.2016732119

Squires, N. K., Squires, K. C., and Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr. Clin. Neurophysiol.* 38, 387–401. doi: 10.1016/0013-4694(75)90263-1

Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. A. F. (2015). No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* 19, 757–770. doi: 10.1016/J.TICS.2015.10.002

Valli, K., Radek, L., Kallionpää, R. E., Scheinin, A., Långsjö, J., Kaisti, K., et al. (2023). Subjective experiences during dexmedetomidine- or propofol-induced unresponsiveness and non-rapid eye movement sleep in healthy male subjects. *Br. J. Anaesth.* 131, 348–359. doi: 10.1016/j.bja.2023.04.026

Windt, J. M., Nielsen, T., and Thompson, E. (2016). Does consciousness disappear in dreamless sleep? *Trends Cogn. Sci.* 20, 871–882. doi: 10.1016/J.TICS.2016.09.006

frontiers | Frontiers in Psychology

TYPE Hypothesis and Theory
PUBLISHED 18 March 2024
DOI 10.3389/fpsyg.2024.1329259

OPEN ACCESS

EDITED BY
Antonino Raffone,
Sapienza University of Rome, Italy

REVIEWED BY
Jonas Coelho,
São Paulo State University, Brazil
Louis Neal Irwin,
The University of Texas at El Paso,
United States
Ken Mogi,
Sony Computer Science Laboratories, Japan
Wieslaw Galus,
Institute of Plasma Physics and Laser
Microfusion, Poland

*CORRESPONDENCE
Bruno Forti
✉ brunoforti4@gmail.com

RECEIVED 28 October 2023
ACCEPTED 07 March 2024
PUBLISHED 18 March 2024

CITATION
Forti B (2024) Approaching the nature of consciousness through a phenomenal analysis of early vision. What is the explanandum?
Front. Psychol. 15:1329259.
doi: 10.3389/fpsyg.2024.1329259

Approaching the nature of consciousness through a phenomenal analysis of early vision. What is the explanandum?

Bruno Forti*

Department of Mental Health, Azienda ULSS 1 Dolomiti, Belluno, Italy

Loorits (2014) identifies the solution to the hard problem of consciousness in the possibility of fully analyzing seemingly non-structural aspects of consciousness in structural terms. However, research on consciousness conducted in recent decades has failed to bridge the explanatory gap between the brain and conscious mind. One reason why the explanatory gap cannot be filled, and consequently the problem remains hard, is that experience and neural structure are too different or "distant" to be directly compatible. Conversely, structural aspects of consciousness can be found in phenomenal experience. One possible alternative, therefore, is to seek the structure of seemingly non-structural aspects of consciousness not in the neural substrate, but within consciousness itself, through a phenomenal analysis of the qualitative aspects of experience, starting from its simplest forms. An essential premise is to reformulate the explanandum of consciousness, which is usually attributed to qualia and what it is like to be in a certain state. However, these properties do not allow us to identify the fundamental aspects of phenomenal experience. Sensations such as the redness of red or the painfulness of pain are inseparable from the context of the experience to which they belong, making qualia appear as phenomenal artifacts. Furthermore, the simplest qualitative aspects can be found in early vision. They are involved in perceptual organization and necessarily have relational significance. The unitary set of qualities found in early vision—such as those related to being an object, background or detail—constitutes the explanandum of the simplest forms of consciousness and seems to imply a justifying structure. Although early vision is characterized by interdependent qualitative components that form a unitary whole, we cannot find in it the structure of seemingly non-structural aspects of consciousness. Phenomenal appearance alone does not seem sufficient to identify a unitary structure of consciousness. However, the closeness of these characteristics to a unitary structure prompts us to delve into less explored territory, using the components of experience also as possible explanans.

KEYWORDS

explanatory gap, conscious structure, phenomenal analysis, explanandum, qualia, early vision

Introduction

In a 2014 paper, Loorits stated that "one possible way to present the hard problem of consciousness is to consider three seemingly plausible theses that are in an interesting tension. First, all the objects of physics and other natural sciences can be fully analyzed in terms of structure and relations, or simply, in structural terms. Second, consciousness is (or has) something over and above its structure and relations. Third, the existence and nature of consciousness can be explained in terms of natural sciences." In other words, if we want consciousness to be explained in terms of natural sciences, we should be able to analyze it in structural terms. However, consciousness seems to be something that goes beyond its structure and relations. Loorits sees the possibility of analyzing in structural terms seemingly non-structural aspects of consciousness like qualia as the solution to the hard problem of consciousness.

Loorits founds his arguments on Crick and Koch's work on consciousness (Crick and Koch, 1998). The idea is that the structure of a quale is a network of nodes (neurons) in the brain. A fully structural account of consciousness answers the question of how phenomenal consciousness could possibly "rise" from neural activity: if the hypothesis is correct, then the phenomenal consciousness simply is a certain complex pattern of neural activity. On this account a person experiences a particular quale when a given ensemble of neurons reaches a certain threshold.

However, this hypothesis does not seem capable of solving the hard problem, which is basically bridging the explanatory gap between physical properties and experience (Levine, 1983). It does not explain how a sensation could emerge from the activity of a network of neurons in the brain. Even in the way Loorits (2014) poses it, the hard problem seems to remain unresolved. In the years since Crick and Koch's (1998) pioneering research, numerous authors have sought to identify the structure of Phenomenal Consciousness in neuronal organization (Seth and Bayne, 2022).

During the last three decades, the advent and development of new scientific procedures, such as functional magnetic resonance imaging and positron emission tomography, have allowed neuroscientists to study the activity of the living brain. These methods have been extensively used to identify with an acceptable degree of accuracy the neural correlates of any aspect of mental activity (Nani et al., 2019). Tracking the correlations between brain processes and states of phenomenal consciousness (neural correlates of consciousness) is the basic method of scientific consciousness research (Tononi and Koch, 2008; Polák and Marvan, 2018). Many potential neural correlates have been investigated. A classic example of an attempt to identity neural correlates of consciousness comes from the study by Sheinberg and Logothetis (1997), who used the phenomenon of binocular rivalry and significant correlation between neuronal activity and the conscious percept in infero-temporal cortex but not V1. My previous article (Forti, 2021) provides a detailed description of the correlations between brain processes and phenomenal consciousness. In short, one could say that consciousness is dependent on the brainstem and thalamus for arousal; that basic cognition is supported by recurrent electrical activity between the cortex and the thalamus at gamma band frequencies; and that some kind of working memory must, at least fleetingly, be present for awareness to occur (Calabrò et al., 2015).

With regard to subcortical structures, the cerebellum has four times more neurons than the cortex but has little effect on consciousness and its contents (Lemon and Edgley, 2010). By contrast, brainstem lesions typically cause immediate coma by damaging the reticular activating system and its associated neuromodulatory systems. However, neurological patients with a severely damaged cortex, but with relatively spared brainstem function, typically remain in a vegetative state. This suggests that brainstem activity is insufficient to sustain consciousness. Rather, it is likely that the activity of heterogeneous neuronal populations within the brainstem, hypothalamus and basal forebrain, which project diffusely to thalamic and cortical neurons and promote their depolarization, provides an important background condition for enabling consciousness by facilitating effective interactions among cortical areas (Parvizi and Damasio, 2001).

The role of the thalamus in consciousness remains controversial. Small bilateral lesions in the intralaminar nuclei of the thalamus can lead to coma, and chronic thalamic electrical stimulation may promote recovery in some patients with disorders of consciousness. Although the so-called core neurons in primary thalamic nuclei have focused connectivity, several higher-order thalamic nuclei are rich in widely projecting matrix cells, which may facilitate interactions among distant cortical areas. Thus, some thalamic cells may represent critical enabling factors for consciousness (Van der Werf et al., 2002; Koch et al., 2016).

With regard to cortical activity, according to the Global Neuronal Workspace model (Dehaene et al., 1998), when a stimulus is presented but not consciously perceived, activation can be seen mainly in the associated primary sensory cortices. When the stimulus is consciously perceived, however, activation in primary cortical areas is followed by a delayed 'neural ignition' in which a sustained wave of activity propagates across prefrontal and parietal association cortices (Noel et al., 2019) and send top-down signals back to all processors (Maillé and Lynn, 2020).

Other evidence across lesion, stimulation, and recording studies consistently point to regions in the "back" of the cortex, including temporal, parietal, and occipital areas, as a "posterior hot zone" that seems to play a direct role in specifying the contents of consciousness (Koch et al., 2016). By contrast, evidence for a direct, content-specific involvement of the "front" of the cortex, including most prefrontal regions, is missing or unclear (Boly et al., 2017). Although most prefrontal regions may be "mute" as regards to consciousness, it remains possible that some prefrontal regions, such as ventromedial areas (Koenigs et al., 2007) or premotor areas, may contribute specific conscious contents, such as feelings of reflection, valuation, and affect.

Recent neuroscientific findings challenge the widely held assumption that similar neural mechanisms underlie different types of conscious awareness, such as seeing, feeling, knowing, and willing. Even within a single modality such as conscious visual perception, the anatomical location, timing, and information flow of neural activity related to conscious awareness vary depending on both external and internal factors (He, 2023). For example, whether the prefrontal cortex is involved in conscious perception might depend on the characteristics of the sensory input: if it is simple and unambiguous, the prefrontal cortex might not be needed (DiCarlo et al., 2012); if it is complex or ambiguous, at least ventral prefrontal cortex appears to be recruited.

Some pathological conditions such as Contralateral Neglect syndrome could provide a window into consciousness. Jerath and Crawford (2014) suggest that the thalamus generates a dynamic default three-dimensional space by integrating processed information from corticothalamic feed-back loops, creating an infrastructure that may form the basis of our consciousness.

The impact of the circadian rhythms on spectral characteristics of EEG signals and on consciousness fluctuations has been investigated for more than half a century (Lehnertz et al., 2021). An activated or desynchronized EEG, one of the oldest electrophysiological indices of consciousness, is still one of the most sensitive and useful markers available (Koch et al., 2016). Spontaneous activity in the alpha-band may index, or even causally support, conscious perception (Gallotto et al., 2017). Low gamma-band (30–50 Hz) synchronization between neural groups coding the various features of objects currently populating experience has been proposed as a mechanism for dynamic functional integration in the brain and has been suggested to be the biological basis of perceptual experience and feature binding (Doesburg et al., 2009).

The time course of conscious perception has been studied using event-related potential components associated with awareness. Railo et al. (2011) argue that the visual awareness negativity component that occurs around 200 ms after stimulus presentation might be associated with conscious perception, and late positivity that occurs around 300–400 ms after stimulus presentation might be associated with conscious access (Raffone et al., 2014). Different event-related potentials likely correspond to different aspects of phenomenal consciousness—not all of consciousness—which may explain some of the disagreements in the literature (Friedman et al., 2023).

Recently, a number of theories have proliferated attempting to explain phenomenal experience and qualia based on the activity of electromagnetic field (Jones and Hunt, 2023). Field theories have arguably made real progress in explaining how fields integrate colors to form unified pictorial images (McFadden, 2020, 2023; Ward and Guevara, 2022). Theories of consciousness rooted in quantum physics are also well known (Hameroff and Penrose, 2014; Tuszynski, 2020). A major problem for quantum mind theories is to explain how quantum effects can occur in the brain at a sufficient scale to be useful (Tegmark, 2000; Bond, 2023).

However, all these studies do not seem to be able to bridge the explanatory gap between physical phenomena and phenomenal experience (Skokowski, 2022; Jones and Hunt, 2023; Sanfey, 2023). Neuroscientists track how light impinging on the retina is transformed into electrical pulses, relayed through the visual thalamus to reach the visual cortex, and finally culminates in activity within speech-related areas causing us to say "red." But how such experience as the redness of red emerges from the processing of sensory information is utterly mysterious (Kanai and Tsuchiya, 2012). In other words, these studies do not seem capable of explaining the phenomenal and qualitative, seemingly non-structural aspects of consciousness. That is to say, they do not seem capable of bridging the explanatory gap between experience and physical substrate as is the case with the "qualitative" properties of wood and stone (Loorits, 2014).

In my opinion, a possible alternative is to look for the structure of seemingly non-structural aspects of consciousness not in the neuronal substrate, but in consciousness itself, through a phenomenal analysis of the qualitative aspects of experience, starting from its simplest forms. An essential prerequisite for this hypothesis is to define the explanandum in terms that can be useful for research. This article is aimed at defining the explanandum, i.e., what about consciousness we find useful to explain. In particular, I will try to highlight that qualia, which many authors identify as the main explanandum of consciousness, do not have a phenomenal existence as isolated entities and that the qualitative aspects analyzed in the literature must be placed in a more complex structural context than is commonly believed. Furthermore, the simplest qualitative aspects belong to early perception and necessarily have relational significance. This is a first step of a phenomenal analysis that I will develop further elsewhere, hypothesizing a hidden structure of consciousness.

The problem of the specificity of consciousness

An often underestimated problem is the specificity of the aspects of consciousness that constitute the explanandum. In this sense, a theory of consciousness cannot avoid referring to qualia or, as I call them in this paper, the qualitative aspects of experience. The idea that consciousness has some features over and above its structural and relational properties has been strongly criticized by many (for example by most of the functionalists, behaviorists, and representationalists). However, most of the attempts to analyze consciousness in fully structural terms have ended up eliminating or simply ignoring certain (qualitative) aspects of consciousness whose existence is considered as absolutely obvious by many (Loorits, 2014). By eliminating or ignoring certain aspects of consciousness, these approaches to consciousness propose a correlation with something that is not necessarily conscious. In other words, one could say that they fail to identify the specificity of consciousness.

What aspects of consciousness that we recognize as such are useful in formulating a theory of consciousness? One way of asking this question is to ask what aspects of consciousness are specific, in order to avoid referring to "false positives," i.e., states that are not conscious even though they exhibit some features typical of consciousness. The properties most often associated with consciousness (James, 1890; Tononi and Edelman, 1998; Zeman, 2001; Edelman, 2003; Searle, 2004) are the following: qualitative character; subjective; unitary; intentional; selective, with a foreground and background. According to Searle (2000), the essential trait of consciousness that we need to explain is unified qualitative subjectivity. Tononi and Koch (2015) identify five essential properties that belong to conscious experience, namely intrinsicality, composition, information, integration, and exclusion.

A fundamental distinction is the one between "Phenomenal" Consciousness and "Access" Consciousness (Block, 1995, 2005). Access consciousness can be considered a non-specific form of consciousness, as it can belong to consciousness, but also to many other non-conscious states (Tyler, 2020). Many theories of consciousness, as was historically the case with binding (Feldman, 2013), fail from square one precisely because they refer to something that is not specific to consciousness. Specificity is not fulfilled in the case of the unity of consciousness either, even though this is a characteristic that almost all authors attribute to consciousness. The unity of consciousness at a single time (Bayne, 2010), related to the ability to integrate information from all senses into one coherent whole—e.g. unified images (Jones and Hunt, 2023), can apply to different non-conscious systems. In Recurrent Processing Theory, the unconscious visual functions of feature extraction and categorizations are mediated by the feedforward sweep, while conscious functions related to perceptual organization are mediated by recurrent cortico-cortical connections (Lamme, 2010). However, these latter functions - that only occur when conscious percepts are present—are candidate neuronal correlates of consciousness. They are not conscious by themselves.

The Higher Order Theory of consciousness claims that a mere first-order representation is not sufficient for conscious experiences to arise (Brown et al., 2019). However, even a first-order state being in some ways monitored or meta-represented by a relevant higher-order representation is in no way sufficient for a state of consciousness to occur. The Global Neuronal Workspace model (Baars, 1997; Dehaene et al., 1998; Dehaene, 2014) is a model according to which conscious access occurs when incoming information is made globally available to multiple brain systems through a network of neurons with long-range axons. Why should global accessibility give rise to conscious experience (Chalmers, 2007)? Intentionality, as a quality of being directed toward an object, has often been associated with consciousness. But even a non-conscious system like an automaton can relate to something external to it. Not even the ability to select one region of the field as the object rather than another (Schwarzkopf and Rees, 2015) guarantees the occurrence of conscious experience. Therefore, there are aspects that do belong to consciousness, but not in a specific way. In the absence of specific features of consciousness, there is a risk of formulating a theory that refers to something that is compatible with the absence of consciousness.

Conversely, the specific characteristics of consciousness can be attributed to its phenomenal aspect. It is precisely this aspect of consciousness that is extremely difficult to explain in relational and structural terms. Phenomenal Consciousness (PC) seems to represent what is unique to consciousness, which exists exclusively in the presence of consciousness and not in other situations. Consequently, if a property such as unity undoubtedly applies to consciousness, then we should understand how the unity that manifests itself on the phenomenal level differs from other forms of unity (Wiese, 2017).

Difficulty arises when we try to better define the meaning of PC. How can we characterize phenomenal consciousness? Specificity is fulfilled if one experiences something in being an organism. According to Nagel (1974), a being is conscious just if there is “something that it is like” to be that creature, i.e., some subjective way the world seems or appears from the creature’s mental or experiential point of view. This is a vague and imprecise concept, presumably referring to a set of several closely intertwined components, such as more or less complex qualitative aspects, subjectivity and value connotations. As Loorits (2014) points out, “the most common ways to introduce the hard problem are intuitively appealing but rather obscure in meaning.”

A similar way of characterizing phenomenal consciousness is the notion of qualia (Kind, 2008). Qualia seem to fully meet the specificity criterion. The sheer qualitative feel of pain is a very different feature from the pattern of neuron firing that causes the pain (Searle, 1997). We shall see that the concept of quale, as interpreted by many authors, also appears questionable. In view of these limitations, in this paper I will refer to the concept of qualitative aspect rather than the concept of quale.

Phenomenal analysis: investigating consciousness “from within”

The seemingly insurmountable difficulty of explaining the phenomenal aspects of consciousness must prompt us to reflect. We look for the structure of PC in the brain substrate, apparently without succeeding. However, we must ask ourselves whether the problem lies in consciousness itself rather than in the substrate. The extreme difficulty of explaining qualia in terms of brain structure could be considered an anomaly in the sense described by Lightman and Gingerich (1992). An anomalous fact is one that is unexpected and difficult to explain within an existing explanatory framework. According to Kuhn, awareness of anomaly is “the recognition that nature has somehow violated the pre-induced expectations that govern normal science.” In this sense, the structure of seemingly non-structural aspects of consciousness could be sought not in the neuronal substrate, but in consciousness itself. While it is known that consciousness has structural aspects, it is underestimated that many of them are related to its qualitative aspects. As I will try to highlight in this paper, the relational and unitary nature of its qualitative aspects cannot be ignored.

Experience and brain structure are too different or “distant” to be directly compatible. On the contrary, structural aspects of consciousness can be found in phenomenal experience. We can “perceive” the relational characteristics of PC. As will be discussed further below, despite the supposed intrinsic nature of qualia, many phenomenal aspects of experience—if not all—appear relational to us. At the same time, we can experience the unity of PC. The components of the perceptual field, such as the part and the whole, appear dependent on each other (Wagemans et al., 2012; Tononi and Koch, 2015).

Consequently, an analysis in structural terms of consciousness could be carried out not by searching for the structural features of the brain that can account for the phenomenal characteristics of consciousness (Tononi and Koch, 2015), but starting from the phenomenal properties of consciousness. There are phenomenal aspects that we do not usually take into account. It is important to point out that in almost all theories of consciousness, phenomenal aspects are either ignored altogether or are analyzed in a very cursory and superficial way.

The hypothesis of a structure of consciousness can only be explored by correctly identifying the starting point. This paper is devoted to the search for the explanandum—what about consciousness we find useful to explain, both in terms of specificity and simplicity. The explanandum of consciousness is usually traced back to qualia and what it is like to be in a certain state. However, the explanandum must be reformulated, since qualia, taken alone, are a phenomenal artifact. In addition, these properties do not make it possible to identify the basic aspects of phenomenal experience. Sensations such as the redness of red or the painfulness of pain must be placed in a more complex structural context than is commonly believed. The simplest qualitative aspects—such as those related to being an object, background or detail—can be found in early vision. Such phenomenal qualities, which are manifold and different from each other, are perceived in relation to each other and seem to form a unitary whole.

As I will explain later in the text, I am not referring to the most frequent definitions of early vision, which can start from retinal vision (Tomasi, 2006; Ghosh, 2020). Here I am referring to it as the simplest form of visual experience, related to perceptual organization. In this sense, early vision corresponds to Kanizsa’s (1979, 1980) “primary vision.” Early vision does not involve recognition, semantic interpretation, or other higher cognitive processing of visual information.

I call the method I adopt in this paper phenomenal analysis. Quite simply, its objective is to identify the structure of consciousness on the basis of the analysis of the phenomenal and qualitative aspects of experience, starting from its simplest forms. I call this analysis phenomenal rather than phenomenological because, while my approach has aspects in common with phenomenology in the observation of conscious phenomena, it does not aspire to embrace a methodological apparatus as complex as the one of phenomenology. My analysis primarily addresses very simple forms of experience, trying to prioritize the aspects that seem to belong to the fundamental "framework" of consciousness and might be involved in the formation of its structure.

Moreover, phenomenology investigates what characterizes perceptions, judgments or feelings. Its goals do not involve the search for an explanation of consciousness, as phenomenology addresses phenomena as they manifest themselves in the intentional consciousness of the subject. "Phenomenology is concerned with attaining an understanding and proper description of the experiential structure of our mental/embodied life; it does not attempt to develop a naturalistic explanation of consciousness, nor does it seek to uncover its biological genesis, neurological basis, psychological motivation, or the like" (Gallagher and Zahavi, 2008).

With respect to the matter of simplicity, it is worth noting that, in addition to identifying the specific aspects of consciousness, a theory of consciousness should identify the simplest forms of phenomenal consciousness. There are several reasons for this. First of all, in any theory it is important to identify the fundamental aspects of the phenomenon under study. The identification of elementary units has been a key in many fields of science and could also be a key in the field of consciousness research (Kanai and Tsuchiya, 2012). Secondly, it is necessary to identify the simplest level at which consciousness manifests itself. Edelman (2003) distinguishes between primary consciousness, which concerns sensations, images and perceptual experiences in general, and higher-order consciousness, which includes self-consciousness and language. However, the main problem is the description of primary consciousness, because higher-order consciousness emerges from processes that are already conscious. Thirdly, the simplest forms of consciousness might have been the first to appear in the course of evolution and the primary significance of its appearance should be traced to them. Finally, the most difficult aspects to explain seem to be the apparently less complex ones. In this sense, the mystery of consciousness seems to boil down to the impossibility of explaining the fact that we experience sensations (Chalmers, 1995). Simple aspects such as the redness of red or the painfulness of pain help identify the problem of consciousness very effectively (Humphrey, 2006).

I will focus phenomenal analysis not on qualia and raw feelings, but on the qualitative aspects of the simplest forms of visual experience taken as a whole. This way, phenomenal analysis makes it possible to highlight the relational nature of the qualitative aspects of perceptual experience. As we shall see in the course of the analysis, at some point there comes the problem of explaining how the qualitative components of the conscious field form a totality of interdependent parts. In fact, the different qualitative components of the phenomenal field appear to be both distinct and dependent on each other at the same time, without it being possible to identify which structure is responsible for this.

This appears to be a limitation of an analysis that considers only the apparent aspects of visual experience. However, the "closeness" of these characteristics to a unitary structure prompts us to delve into less explored territory, using the components of experience also as possible explanans. In a separate paper, starting from the nature of appearance itself, I will consider the need to postulate the existence of non-apparent aspects.

Qualia are a phenomenal artifact

One of the main problems in the approach to consciousness is that we tend to identify the simplest aspects of experience with qualia. It is a common view that simple qualia could be a useful starting point for a theory of consciousness. Koch (2004) wonders how the elemental feelings and sensations making up conscious experience arise from the concerted actions of nerve cells and their associated synaptic and molecular processes. The assumption is that if we explain the neuronal substrate of pain, sweetness and the redness of red we lay the foundation for explaining consciousness.

However, identifying the simplest aspects of experience with qualia is erroneous. According to the majority of authors, considering qualia as a possible starting point for a theory of consciousness means being able to think of them as isolated, or extrapolating them from objects and other components of the field of experience as fully representative of experience itself. Then, it means being able to look for the simplest possible explanation of consciousness at the level of brain organization. It should be noted that, although Lewis separates the properties of qualia from those of objects, he does not identify them with conscious experience: "This given element in a single experience of an object is what will be meant by 'a presentation.' Such a presentation is, obviously, an event and historically unique. No identification of the event itself with the repeatable content of it is intended" (Lewis, 1929). However, the way in which literature on consciousness has defined the concept of quale over time has coincided with a tendency to separate it from anything having to do with the idea of relationship and structure. Qualia are intrinsic, i.e., non-relational (Dennett, 1988; de Leon, 2001; Siddharth and Menon, 2017). As Loorits (2014) points out, qualia are some features of consciousness over and above its structural and relational properties.

The meaning of non-relational is not univocal. We must distinguish between internal relations and external relations. Regarding the former, Dennett (1988) states that "qualia … are *intrinsic* properties—which seems to imply … that they are somehow atomic and unanalyzable." Simple qualia such as blueness or sweetness have no obvious signs of an internal structure (Haun and Tononi, 2019). According to Loorits (2014), in the classical view, qualia would be monadic, not compositional, and with no internal structure: "when I have a visual perception of a red apple, I have a direct epistemic access to many structural features of my visual experience: the size and shape of the perceived apple, for instance. I do not have similar direct epistemic access to the structure of the perceived redness of my visual experience."

However, it should be noted that the non-analyzability of qualia is related to the fact that they are characterized by an internal homogeneity, which Metzinger (2004) calls ultrasmoothness, in the sense that they have a grain structure. We should keep in mind that at the conscious level we can make a phenomenal distinction only by contrasting one region with another. If there is no contrast within a red surface, we perceive it as homogeneous and cannot make any phenomenal distinction. However, its supposed non-analyzability, which we perceive phenomenally as homogeneity, is a piece of information about the region of the perceptual field that differs from the possibility of any point or part of that region not being red. Experiencing the redness of red means seeing the red color distributed homogeneously over an object. This is information that we receive from experience and that we ignore if we speak abstractly about the redness of red. Therefore, the unanalyzability of qualia is at least questionable with regard to its internal relations.

With regard to external relations, according to de Leon (2001), "that qualia are intrinsic means that their qualitative character can be isolated from everything else going on in the brain (or elsewhere) and is not dependent on relations to other mental states." According to the standard view, qualia are not in themselves, representational or intentional (Loar, 2002). According to Dennett (1988), intrinsic means that they are non-relational properties, which do not change depending on the experience's relation to other things. Consequently, qualia would not be related:

1 with other mental states and behavioral output, so they are not mental states in the functional sense of the term (de Leon, 2001; Van Gulick, 2017);
2 with the stimulus and, in a broad sense, with the external reality to which they refer (see inverted or absent qualia), so they are non-intentional and non-representational (Loar, 2002);
3 with other components within the experiential field; or at least, they can be separated from them, e.g., from the object, so they are universals (Lewis, 1929; Dennett, 1988).

While the first two statements concern undeniably important aspects, specifically the functional and the intentional ones, the third is crucial for a phenomenal conception of consciousness. Claiming that qualia are not characterized by their relations to other components of the field has three implications: first, the idea that extrapolation from other components of the field can allow the phenomenal properties of qualia to be preserved; second, that everything within the field that has to do with relation is not, in the strict sense of the word, phenomenal; and third, that everything that has to do with relation, and more broadly with structure, can be explained in terms of cerebral or other organization.

However, relations with other components of the experiential field have to do with the very nature of experience, of what is phenomenal. In the absence of such relations, qualia risk being incompatible not only with a functional and intentional view of mind (Loar, 2002), but also with the essence of PC. Since qualia are extrapolated from the phenomenal experience in which they are placed, they give no guarantee of retaining phenomenal qualities, so they cannot be considered fully representative of the experience itself. The universality of qualia, i.e., the possibility of their being recognized from one experience to another, must be distinguished from their phenomenal nature, which is related to their relations in each individual experience. At the same time, it is difficult to deny the phenomenal nature of the relational aspects of consciousness, such as seeing the object place itself in the foreground and the background extend behind it.

If we limit ourselves to vision, some of the most frequently described qualia are the ones that refer to colors. Scholars refer to the redness of red, using terminology that is different from the one of common sense and referring to a visual experience that is distant from the usual ones. Interestingly, scholars do not refer to the way we see a face, which is much closer to the reality of conscious vision, and which is used as a prototype of conscious experience in many experiments on neural correlates of consciousness (Koch et al., 2016). This is probably because it would be much more difficult to describe the phenomenal experience of a face in non-relational terms.

Dennett (1988) defines qualia as the ways things seem to us. As an example, he cites the way we see a glass of milk at sunset. According to Dennett, "the particular, personal, subjective visual quality of the glass of milk at sunset is the *quale* of your visual experience at the moment." However, it is very difficult to have this kind of experience and describe it in the absence of its internal relations: the whiteness and liquidity of milk, the fact that the milk is contained in the glass, the convexity and transparency of the glass, the table on which the glass of milk is standing, the sun next to the glass that disappears over the horizon, the particular light of sunset that affects the way the glass looks, the feeling that this vision can arouse, and our state of mind when we see the glass. What would this experience be without these relations? Would it be an experience in an absolute sense? Is it possible to really separate the elements that, in relation to each other, make up our phenomenal reality from the way they seem to us? Are we assuming that there is a conscious quale of the vision of the glass of milk at sunset that is associated with the non-conscious vision of the glass of milk at sunset?

Or are we assuming that qualia give to a perceptual state the particular qualities that would make it phenomenal, whereby the phenomenal character would be determined in the relation between qualia and perceptual state? In other words, "qualia … are properties of sensations and perceptual states, namely the properties that give them their qualitative or phenomenal character—those that determine 'what it is like' to have them" (Shoemaker, 1991). This could mean that, in order to be conscious, the vision of an object must have particular qualities. But the conscious nature of perception is either an expression of the set of relations existing between the components of the field—without our being able to confidently assign a particular status to any of them—or we must assume that something similar happens to when the magic dust from the wand of Cinderella's fairy godmother turns the pumpkin into a carriage.

The intrinsic nature of qualia can be traced to the supposed simplicity of some of them, such as the ones related to color. But this misunderstanding stems from a phenomenal simplicity of the perception of a color which, in fact, is not so simple. Let us try to replace red with black in a black-and-white world, made up of black, white and a range of grays. It is a simpler world, but it is to all intents and purposes a phenomenal world. In a black-and-white world, it becomes much more difficult to speak of the blackness of black as an intrinsic element. Dark gray is phenomenally dark gray because it differs from the white background more than light gray and less than black. Black is black because it equals black and differs from white more than any shade of gray. It is hard to imagine that this does not apply to a color like red that is immersed in a more complex range of relations, including, in addition to the light–dark dimension, saturation and relation to other colors.

In this sense, a certain shade is necessarily related to something that is outside the field of the stimulus. Perception of so-called elemental qualities implies the involvement of memory in the conscious field, as Edelman (2001) eloquently expressed with the concept of *remembered present*. Perceiving a color implies similarities and differences with reference patterns that cannot derive solely from the present stimulus and that must consciously manifest themselves somehow, e.g., "in the background." In other words, not only the premises of the perceived quality, but also the perceived quality itself, in the way it is perceived, imply the involvement of elements that are not present in the stimulus. One of the properties of qualia, which gives them their universal character, is precisely the fact of being recognized from one experience to another (Lewis, 1929).

Briefly, it is more correct to speak of *qualitative aspects as components of experience* rather than qualia. They cannot be analyzed independently of the experiential field to which they belong. If we consider qualitative aspects taken in isolation as fully representative of experience, we distort their phenomenal essence. It is an operation that creates a phenomenal artifact (de Laguna, 1916).

If, on the contrary, we admit that the qualitative aspects of experience cannot be extrapolated from the context to which they belong without undermining their phenomenal nature, an important consequence is that they are necessarily relational. Being relational is an integral part of the nature of what is qualitative. Unless we assume that *the entire* field of experience is something intrinsic, monadic, and nonrelational, the object of phenomenal analysis can only be the field of experience in its totality and in its internal relations.

The simplest aspects of consciousness should be researched in perception

Another consequence of considering qualitative aspects of experience as relational is that we will not necessarily focus primarily on the qualitative aspects of what we might call the classical qualia. Although it can occur in the simplest forms of consciousness, a qualitative feel is something that characterizes a conscious experience but is not identified with it. None of us perceives the quale of green, of sweetness, of pain *alone*. We perceive *something* green, something sweet, we perceive pain in a part of the body and therefore in relation to it. We cannot help but perceive these sensations in relation to *something*. There is no evidence that by eliminating what green belongs to, it would retain those phenomenal properties or that it would retain phenomenal properties in general. Also in common usage, in addition to having a positive or negative connotation, a quality is a characteristic or feature that someone or something has (Encyclopædia Britannica, 2023).

Moreover, if the quality is inevitably the quality of something, *this something is in turn always in relation to a background*. In other words, a phenomenal quality cannot but belong to something, and this something cannot but belong to a background. Green belongs to the leaf, pain to the knee. In turn, the green leaf is on the tree, the painful knee is in the leg. As Merleau-Ponty (1945) points out, "at the outset of the study of perception, we find in language the notion of sensation, which seems immediate and obvious: I have a sensation of redness, of blueness, of hot or cold. It will, however, be seen that nothing could in fact be more confused, and that because they accepted it readily, traditional analyses missed the phenomenon of perception … When Gestalt theory informs us that a figure on a background is the simplest sense-given available to us, we reply that this is not a contingent characteristic of factual perception, which leaves us free, in an ideal analysis, to bring in the notion of impressions. It is the very definition of the phenomenon of perception, that without which a phenomenon cannot be said to be perception at all. The perceptual 'something' is always in the middle of something else, it always forms part of a 'field'… The pure impression is, therefore, not only undiscoverable, but also imperceptible and so inconceivable as an instant of perception." A qualitative feel, insofar as it relates to a perceptual "something" that belongs to a "field," merely adds a sensory aspect to this dyad.

There is a philosophical tradition that tends to attribute the primitive aspects of experience to sensation. According to Reid (1764/1997), if sensation is a simple, subjective datum, perception is a complex cognitive act that actively unifies a set of sensations by ascribing them to an object. It is widely believed that the most relevant aspect of perception is the extraction of relevant information from sensation: detecting, identifying, recognizing (Fesce, 2023). The idea that sensations precede perception (Gärdenfors, 2006) has been somewhat reframed by the attribution of the simplest forms of phenomenal experience to qualia and raw feelings. However, perception is not a more complex and organized form of sensation. The formation of the object is the *sine qua non* for the occurrence of experience. In my view, sensations can only occur in a perceptual context that is, *ab initio*, multisensory (Bennett and Hill, 2014; Bayne and Spence, 2015; O'Callaghan, 2015) and in which sensations are in a way dependent on perceptual aspects. In other words, they can only occur within a conscious perceptual experience (Hardin, 1992). On this basis, rather than with classic qualia and simple sensations, basic consciousness might coincide with perception and the qualitative aspects associated with it.

It could be argued that our experience does not necessarily refer to an object. Even without making reference to the Eastern disciplines (Srinivasan, 2020), it is enough to close our eyes to experience darkness. But in these cases we cannot help but experience our body: if we focus on the visual experience, our body will act as a background to the darkness we perceive and will in turn be perceived in the background of the perceptual space in which our body is located (Jerath et al., 2015). The conscious perception of light and dark, which is identified as one of the simplest things we can perceive (Edelman and Tononi, 2000), is only possible at a level comparable to that of the perception of an object.

If we keep in mind that, in the classical sense of the term, quality is such in relation to a reference pattern existing in memory, that it is in relation to an object, and that the object is in relation to a contrasting background, it is clear that the simplest aspects of phenomenal experience can be detected most easily in a simple figure and thus in early vision. In this sense, early vision is what gestaltists call "primary vision," which occurs even before object recognition (Kanizsa, 1979, 1980, 1991). Kanizsa (1980) states that "visual perception is a complex cognitive activity, in which it is possible to distinguish at least two levels or moments: the moment of the formation of the visual object, i.e., the primary process by which sensory input is organized and segmented, and a secondary process that includes the more properly intellectual operations of categorization, signification, and interpretation that the mind performs on the results of primary segmentation." So, I am referring to early vision as the simplest form of visual experience, related to perceptual organization. It does not involve interpretation or other strictly cognitive processing of visual information.

The figure/background organization is often listed among the properties of consciousness, with similar but not entirely overlapping meanings such as foreground/background, situation, figure/background, center/periphery, selection or choice (James, 1890; Zeman, 2001; Edelman, 2003; Searle, 2004; Northoff et al., 2023). After all, vision—which I will address here in its phenomenal aspects—is the preferred field of investigation of consciousness for many authors (Koch, 2004; Jerath et al., 2015; Lamme, 2020; Ludwig, 2023). It is worth noting that for gestaltists perception is not preceded by sensation but is a primary and immediate process. Structured wholes or Gestalts, rather than sensations, are the primary units of mental life (Wagemans et al., 2012). According to Lamme (2020), perceptual organization is the visual function that is central to understanding the transition from unconscious to conscious seeing. Processes of grouping and figure-ground segregation depend strongly on the stimulus that is evoking these operations being consciously perceived.

It could be argued that, by investigating the principles that determine the grouping or the choice of a region of the field as an object rather than as a background, the Gestalt approach somewhat circumvents the hard problem, since it limits itself to the so-called "functional" aspects of perceptual organization (Lamme, 2010, 2020). However, it should be pointed out that the perception of a figure against a background cannot be equated with the mere result of an operation like the assignment of borders, to which cognitive science attributes the choice of the object (Williford and von der Heydt, 2013). Ever since Rubin's first descriptions, it has been clear that a figure seen against the background of something has purely phenomenal characteristics. The figure has an object-like character, and there is a tendency to see the figure as positioned in front, and the ground at a further depth plane and continuing to extend behind the figure. Furthermore, the border separating the two segments is perceived as delineating the figure's shape as its contour, whereas it is irrelevant to the shape of the ground (Todorovic, 2008).

These characteristics are not taken into account in identifying the basic phenomenal aspects of consciousness. However, they are no less qualitative than the redness of red and the painfulness of pain. Moreover, in the visual field there are not only the figure and the background. A visual object is not such if, in addition to differentiating itself from the background, it does not have an inhomogeneity that underlies its details, its constituent parts and its surface texture. Secondly, in addition to the object and background there are secondary objects and backgrounds, elements that come together to form Gestalts, and so on. Likewise, being an object, a detail, a Gestalt or a secondary object involves attributing a certain phenomenal quality to that part of the field.

The qualities of the field components that result from perceptual organization appear even simpler than the ones usually identified with qualia, with raw feelings and seemingly elementary aspects of phenomenal experience: redness, sweetness, painfulness, roundness, distinction between light and dark. In contrast to classical qualitative aspects, the quality related to being an object can be derived exclusively from features present in the stimulus. There is no need to bring up anything from memory to see an object against the background of something. Although there is no unanimous agreement on this point, it can be argued that in most cases the relation of the object to the background depends on autochthonous factors, that is, on factors that are all in the stimulus, thereby they do not depend on previous knowledge, expectancies, voluntary sets, intentions of the observer (Luccio, 2011).

The Gestalt approach is for all intents and purposes a phenomenological approach. However, in studying perceptual organization, it has addressed very simple aspects of conscious experience. The perceptual field is made up of figure and background, main objects and secondary objects, clear components and other less clear components. One reason why it is difficult to conceive of the perceptual field in its entirety is the progressive fading of its components. However, this is an aspect that is part of consciousness and that cannot be ignored. It is therefore necessary to formulate a conception of experience that includes its fading. One problem lies in the fact that perceptibility declines progressively, with no clear boundary between what we see clearly and what we do not see at all. It is worth noting that in very simple stimulus conditions, as in many of those studied by Gestaltists, we can sufficiently perceive all the relationships in the field, partially overcoming this difficulty.

Galus and Starzyk (2020) and Galus (2023a,b) propose the Reductive Model of the Conscious Mind. It is based on the distinction among different aspects of consciousness served by independent neural processes. According to the authors, attempts to define the phenomenon of consciousness have encountered difficulties. They seemed insurmountable because they strived to explain a multifaceted phenomenon, realized by completely different neural, biophysical, and behavioral phenomena, using one definition, one process or property of matter. The basic structure of consciousness includes three main aspects: Perceptual Consciousness, Executive Consciousness, and Reporting Consciousness. This complex view includes perceptions, the manipulation of the world and of objects, the sensations we derive from this manipulation, emotions, interoception of states of the organism that deviate from a condition of homeostasis. Embodiment requires having a body equipped with senses of external and internal signals reporting on the state of the environment and the state of homeostasis. This body must also be able to respond to detected signals from the environment and its own body.

It is worth noting the hypothesis of how secondary perception can visualize thoughts as well as imagery, memories, and dreams. As Galus (2023a) underlines, "the more important aspect of secondary signal transmission up-down is the dramatic increase in the ability to learn and analyze situations quickly. Thanks to the visualization of one's thoughts, it was possible not only to react directly to sensory stimuli but also to imagine the sequence of actions and plan the reactions optimally. Moreover, it is less about the logical analysis of possible responses and making appropriate decisions but about the idea of how one's body functions, muscle tension, the position of the limbs, and the dynamics of movements."

The scope of my paper is much more limited. I focus on a simpler level. I refer neither to classically defined qualia, nor to interoception. Of course, emotions and qualia of internal states play a fundamental role for the mental states aimed at maintaining homeostasis. However, as we have seen with regard to the phenomenal nature of the perception of object and background, even a simple visual perception is conscious and must be explained and justified as such. Following the distinction of Galus and Starzyk (2020), I think that direct perception can be conscious even if it is not accompanied by phenomenal feelings.

The role of the relationship between subject and object in basic consciousness remains to be clarified. The question is whether this apparently obvious role (Searle, 2004; Damasio, 2010; Damasio and Damasio, 2022) is a fundamental aspect of PC. As phenomenologists argue (Gallagher and Zahavi, 2008), even in the absence of self-consciousness in the full sense of the term, consciousness would be characterized by pre-reflective self-consciousness, which is involved in having experiences as one's own and can be construed as a kind of low-level self-consciousness (Flanagan, 1992). In a similar sense, Kriegel (2004) speaks of peripheral self-consciousness.

If the basic aspect of consciousness is perception in its simplest forms, it appears less intimately linked to subjectivity than sensations. It is certainly true that our conscious experiences are subjective. However, it is one thing to take an interest in the world around us; it is another to observe ourselves as we observe the world. It is one thing to have an egocentric perspective; it is another to have an allocentric perspective, such as when we look at a map. Sensations—such as heat or pain—that directly concern the subject and its relations to the outside world are one thing; "distal" features of the outside world that are such because of the relations between the elements that make it up—such as the roundness of an object or the number of trees in the forest in front of us—are another. When we turn our attention to the outside world, our conscious experiences are not characterized by introspective awareness (Seager, 2002). When we become absorbed in some intense perceptual task, we are vividly conscious but, often, we may lose the sense of self (Tononi and Koch, 2008).

If we hypothetically eliminated the subjective component of consciousness, the phenomenal problem of vision—about why a red triangle appears as such and it is not just a configuration eliciting a response—would still remain unsolved. The fact that a red triangle appears *to us* cannot be the only element accounting for its appearance and for its phenomenal ontology (Forti, 2009). It is therefore possible to temporarily set aside the problem of subjectivity. As Merleau-Ponty (1945) points out, "it is the very notion of the immediate which is transformed: henceforth the immediate is no longer the impression, the object which is one with the subject, but the meaning, the structure, the spontaneous arrangement of parts."

In my view, early vision can represent a form of experience that, by allowing subjectivity to be temporarily put in abeyance, provides a pathway to consciousness that may facilitate the formulation of third-person theoretical constructs. Experimental situations in which gestalt laws are tested represent experiences that feature characteristics of phenomena observable in the third person perspective. Or, at least, the role of the subject can be considered irrelevant. In these situations, what we see seems to depend phenomenally on the relationships between the components of the field rather than on the relationships between object and percipient subject. Most Gestalt laws concern the organization of conscious vision. They are based exclusively on the relations existing in the perceptual field, starting with the relation of the object to the background (Luccio, 2011). Of course, vision necessarily implies a point of view, but it is the same with many recording and measuring instruments. Moreover, perception can be considered a public mode of observation. In this sense, visual perception has aspects in common with the scientific approach, of which, through observation of the world around us, it is the basis (Gallagher and Zahavi, 2008).

The explanandum is a unitary whole of qualities

What do we find if we analyze the simplest forms of visual perceptual experience? The first observation might be in some respects obvious and in others questionable: the simplest aspects of consciousness can be seen in the perception of a simple figure against the background of something (Merleau-Ponty, 1945). Unlike classical qualia, a figure has a clear relation to the background, which is essential for the perception to have the phenomenal characteristics that are well known to us. The "quale" of the object can only be perceived or conceived in relation to the "quale" of the background.

But not only the figure against the background of something is relational. We have seen that all qualitative aspects of consciousness are relational. These aspects include the ones that are usually attributed to qualia, whereby the quality is such in relation to a reference pattern in memory and is in relation to an object that is in turn in relation to a contrasting background. If we simply examine the relations existing in early perception, the relational aspect is even more evident. Any content can only have phenomenal characteristics in relation to other contents or aspects of the perceptual field, starting from the object and background. Being the main object implies at least a background, other objects over which it prevails, as well as the details and parts of which it is made up.

Another fundamental aspect of phenomenal experience, related to the previous one, is the unity we experience in all perceptions. Since Descartes and Kant, unity has been considered by almost all authors to be among the fundamental characteristics of consciousness. Often, the attribution of unity to consciousness has implied a monadic conception of consciousness. It should be noted that identifying consciousness with a simple and intrinsic unity is not the exclusive prerogative of classical qualia. In fact, it includes most qualitative conceptions—or conceptions referable to the idea of "what it is like to be"—insofar as reference is made to something that does not appear to be analyzable in its internal structure.

In my opinion, it is preferable to adopt a conception—like the gestalt—whereby unity is not monadic but is such through the interdependence of the parts that make up the field of experience (Kanizsa, 1980; Wagemans et al., 2012; Tononi and Koch, 2015). Unity is clearly found in the visual experiences described by gestaltists, starting from the relationship between part and whole. In a simple perceptual situation, the relations between the elements of the field are characterized by mutual dependence, in the sense that each component of the field is such in function of the others, e.g., object-background, gestalt-constituent elements, object-detail, main object-secondary object. Interdependence seems to involve multiple elements of the field at the same time. A detail could not be perceived as part of an object if at the same time the object were not perceived as belonging to a background. We thus move from a monadic conception of consciousness to a conception whereby the qualitative aspects of consciousness are necessarily manifold and at the same time closely related to each other. The phenomenal analysis of perceptual experience highlights that its qualitative aspects are relational and that consciousness appears to us as unitary through the mutual dependence of these relations. Consequently, we can say that the explanandum is a *unitary set of qualities*, i.e., a set of qualities closely dependent on each other, which we can find in its simplest forms in early vision. Such an explanandum may appear insufficient, but it certainly cannot be reduced to something that does not include these features taken together. This conception is clearly different from the mosaic of qualia, which entails a mere combination of different qualities (Jansen, 2017). The relationship between the various qualitative aspects is something more complex. It entails relationships of interdependence and on different hierarchical levels—not only between objects, but also between contiguous regions.

Above I stated that unity *per se* is not specific to consciousness, as it could belong to many non-conscious organizations, and that, if a property such as unity undoubtedly applies to consciousness, then we should understand how the unity that manifests itself on the phenomenal level differs from other forms of unity (Wiese, 2017). The concept of a unitary set of qualities is well suited to this statement, as unity concerns specific properties of consciousness such as the qualitative aspects. In this sense, the qualities that characterize consciousness are necessarily interdependent parts of a whole that encompasses the entire field. The co-presence of the qualitative aspect and the unity aspect is thus crucial in identifying the explanandum of consciousness.

Unlike Gestaltists and other authors (Tononi and Koch, 2015), this totality should not be identified in the object as a structured whole, but in the total field of experience, which includes background, fringe parts and progressively fading components. We often consider only the most salient contents of consciousness, disregarding the progressively fading field and ignoring other components even when they are sufficiently perceptible. If we do not limit ourselves to the main object, its main features, and the gestalts present in the field, but we also take into account elements such as the background, secondary objects, parts of an object, components of a gestalt, and less important elements, the unitary set of qualities that we identify in a phenomenal analysis of early perception becomes progressively evanescent. The difficulty of dealing with such situations can be partly overcome by limiting ourselves to the simplicity of many stimulus situations analyzed by Gestaltists, in which the progressive fading of the perceptual field is negligible.

Discussion: in search of the unitary structure of consciousness

At first glance, one might think that identifying the explanandum in a unitary set of qualities is equivalent to identifying the structure of consciousness, at least in such elementary forms as early vision. But things are not so simple. I started from the need to analyze in structural terms qualia or, rather, the qualitative aspects of consciousness. The analysis of the simplest forms of perceptual consciousness led us to point out that these qualitative aspects are not only relational, but also form a unitary whole. Thus, the existence of a unitary set of qualities does not allow us to limit ourselves to analyzing in structural terms a single quality. We must also explain their relational nature, the way their relations form a unitary whole and their interdependence in perceptual organization. On the one hand, this explanation may seem more difficult. On the other hand, we can assume that quality and interdependence are somehow related, at least in early vision.

Jones and Hunt (2023) approach this issue in a similar way, but do not challenge the phenomenal reality of qualia. According to these authors, the main problems in neuroscience's accounts of qualia seem to fit into three categories: the coding/correlation problem, the qualia-integration problem, and the hard problem. In my view, these are not three distinct questions, even though they are interrelated; they constitute a single fundamental question, which is to explain the unitary set of qualities encountered in early perception.

With regard to the unity of visual experience, it is not sufficient to say that the various qualitative aspects of consciousness are perceived as interdependent. The perceived interdependence does not explain the qualities of perceptual experience, but neither does it explain how these qualities form a unitary whole. Saying that the explanandum is a unitary whole of qualities is not the same as identifying the structure of the consciousness, i.e., how that whole is organized into a unitary whole. Consciousness should have a structure that justifies such unity.

Therefore, Loorits' argument that consciousness should have a structure must be completed by stating that the structure of consciousness should have that unitary character that is typical of consciousness. The goal is to look not for a series of separate structural aspects, but for a unitary structure. We must ask ourselves whether the relational aspects we identify in experience are compatible with the unity we feel in all perceptual experiences. We cannot separate these aspects. It is neither sufficient to identify on its own the unity we all feel in our experience, nor to identify relational or structural aspects that do not ensure unity by themselves. In a way, a phenomenal analysis goes over the two poles of conscious experience: its being composite, in that it is made up of multiple qualitatively characterized contents or phenomenal distinctions, and at the same time unitary, so much so that, through qualia, it recalls the idea of a monad. How is it possible to reconcile these two poles?

The unity manifested through the interdependence of the parts of the field of experience can be interpreted as a form of integration. Tononi and Koch (2015) propose the Integrated Information Theory (IIT) and list structure (composition) and unity (integration) among the properties of consciousness. In this sense, as a result of the interdependence of phenomenal distinctions, integration is phenomenal evidence rather than a theory. Historically, this has been clear to many authors who have tried to define consciousness (Brogaard et al., 2021; Hirschhorn et al., 2021; Solms, 2021). The problem is to explain how integration, as manifested in conscious experience, can come about. The IIT postulates an organization of the neuronal substrate characterized by complexity and by the presence of high levels of integration and differentiation. This proposal appears to be an almost tautological and overly general explanation to justify the *particular kind* of integration that we observe in experience. Life also involves a complex organization of organic molecules, but postulating a high level of complexity is not sufficient to explain it. Moreover, the IIT does not address the specific qualitative aspect (Cooke, 2021), so it is precisely the qualitative aspects that are integrated into experience. Even if the IIT proposes an explanation for the qualitative aspect (Tononi, 2008), it does not correlate it with the integration that occurs in the perceptual field. By not including an explanation of the qualitative aspects and their relations, a complex system such as the one postulated by the proponents of the IIT may belong to non-conscious organizations.

Moreover, structure should be constitutive, not just reflecting relations existing in the stimulus field. We should identify a structure that is not contingent, but constitutive of each experience and somewhat independent of the type of stimuli (Buzsáki, 2007; Bayne et al., 2016; Smith, 2018; Kent and Wittmann, 2021; Northoff and Zilio, 2022). Many structural aspects highlighted in the literature seem to reflect the organization in the apparent reality of specific contents rather than the internal structure of the conscious field. Of course, we can assume that conscious structure allows us to capture structural aspects of the reality around us, so the ability to capture a structure present in external reality may also be an expression of conscious structure.

A unitary structure can be identified in a simple relationship between figure and background and in their interdependence. The coherence and unity of what we perceive cannot be separated from its belonging to the background: "the background, which need never have been made determinate, affects the appearance of what is determinate by letting it appear as unified, bounded figure" (Dreyfus, 1992). It should be emphasized that this is a phenomenal unitary structure, in that object and background have qualitative characteristics that appear as a function of each other. In essence, there is a unitary structure in the simplest manifestation of consciousness, a phenomenal object in the form of a simple figure. This structure appears constitutive and non-contingent, because we cannot perceive the object without the background. It is constitutive because without this relationship there is no consciousness—even though this relationship reflects a fundamental aspect in the surrounding reality, i.e., the fact that as a rule the world is made up of objects in a space.

However, if we analyze images that are just a little more complex, a unitary structure becomes more difficult to detect. Unity, which manifests itself through the interdependence of the parts, remains perceptible, but we cannot identify the structure underpinning it. It is possible to identify relational qualitative aspects, but they do not seem able to provide phenomenal unity. In their comprehensive approach, gestaltists postulate the unity of the field, but they do not explain it (Wagemans et al., 2012). The various Gestalt laws explain in heterogeneous ways different forms of grouping and the figure-background organization, but not the unity of the perceptual field.

Faced with the heterogeneity of relations between the parts, the apparent unity of perceptual experience leads us to wonder *how* these different relations constitute a unitary whole. Object, background, gestalt, detail, secondary objects are all expressions of the relationship with something else, but, at first glance, they do not allow us to understand *how* they constitute a unitary whole. We might say that the various phenomenal qualities are not all on the same plane. In a perceptual experience the main object stands in the foreground. Other qualities are associated with it in a subordinate way; others are associated with such qualities, and so on, until they completely fade away. However, not even conceiving the various phenomenal qualities as a set of progressively fading hierarchical relationships justifies the apparent unity of the field. Indeed, the phenomenally subordinate relationship of the qualities associated with the main object is not limited to the background's secondary role, but it includes such heterogeneous relationships as the ones involving the secondary objects, parts, details, and elements that form a gestalt. Why do objects, backgrounds, gestalts and details appear as they appear and at the same time are part of a unitary experience?

If we focus on a more complex image than a figure against a homogeneous background, it is not enough to say that on the table there are a bottle, two plates and some glasses, that a picture hangs on the wall, and that we perceive these objects as a unitary whole. There is a gap between the unity we perceive and the possibility of identifying the structure underlying it through relationships that make it possible. We cannot identify the structure that provides the unity we experience and perceive even when the composition of an image seems random. We can put random elements into a visual field (Kanizsa, 1980) and the image will retain its own unity. Thus, unity is not merely contingent.

At a preliminary phenomenal analysis, the problem of the unitary structure of consciousness seems without solution. The fact that we perceive the experience as unitary and perceive the various qualitative aspects as interdependent seems to be a kind of mystery for which we cannot find an explanation, either in brain organization or in experience itself. Phenomenal appearance alone does not seem sufficient to identify a unitary structure of consciousness.

This paper has arguably achieved the goal of identifying an explanandum in terms that can be useful for research, but it has not achieved the goal of identifying the unitary structure of consciousness. The unitary set of qualities that I have identified as the explanandum of consciousness is not a real structure, let alone a unitary structure. While it is a unitary set of qualities, it neither tells us what the structure of seemingly non-structural aspects like the qualities of object and background is, nor does it identify the unitary character of that structure. However, this does not mean going back to the search for the physical substrate that has proven to be dead-end. Elements of "closeness" with the structural aspects of consciousness can be found in appearance itself. The results of the analysis of the simplest forms of perceptual experience, with the presence of closely interdependent qualitative components that form a unitary whole prompt us to go beyond the mere phenomenal appearance, using the components of experience also as possible explanans.

One of the most obvious explananda is appearance, which is nothing else than the etymological meaning of consciousness as a phenomenal entity. In the simplest sense, it implies the possibility of something being perceived consciously. However, it is the very appearance and the way it is structured in perceptual experience that makes us think that the perceptive field contains within itself parts responsible for the appearance, yet they remain imperceptible. Other components of experience that could constitute a possible explanans are generally neglected phenomenal aspects like overlapping of the contents of the field and surroundedness. Surroundedness is a relationship whereby a region is surrounded by or surrounds a contrasting region, and it has a broader meaning than the one we assign to the figure-ground relationship. I will examine the possibility of going beyond the mere phenomenal appearance elsewhere, hypothesizing that the structure of consciousness is somehow conscious, although "hidden" from consciousness itself. Such a structure might provide a kind of link that can bridge—or at least reduce—the explanatory gap between experience and brain processes and thus help solve the hard problem.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

BF: Writing – original draft.

Funding

The author declares that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Baars, B. J. (1997). In the theatre of consciousness: global workspace theory, a rigorous scientific theory of consciousness. *J. Conscious. Stud.* 4, 292–309.

Bayne, T. (2010). *The unity of consciousness*. Oxford: Oxford University Press.

Bayne, T., Hohwy, J., and Owen, A. M. (2016). Are there levels of consciousness? *Trends Cogn. Sci.* 20, 405–413. doi: 10.1016/j.tics.2016.03.009

Bayne, T., and Spence, C. (2015). "Multisensory perception" in *The Oxford handbook of philosophy of perception*. ed. M. Matthen (Oxford: University Press), 603–620.

Bennett, D., and Hill, C. (2014). *Sensory integration and the Unity of consciousness*. Cambridge, MA MIT Press.

Block, N. (1995). On a confusion about a function of consciousness. *Brain Behav. Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188

Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci.* 9, 46–52. doi: 10.1016/j.tics.2004.12.006

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the Back of the cerebral cortex? Clinical and neuroimaging evidence. *J. Neurosci.* 37, 9603–9613. doi: 10.1523/JNEUROSCI.3218-16.2017

Bond, E. (2023). The contribution of coherence field theory to a model of consciousness: electric currents, EM fields, and EM radiation in the brain. *Front. Hum. Neurosci.* 16:1020105. doi: 10.3389/fnhum.2022.1020105

Brogaard, B., Gatzia, D. E., and Chomanski, B. (2021). Consciousness and information integration. *Synthese* 198, 763–792. doi: 10.1007/s11229-020-02613-3

Brown, R., Lau, H., and LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends Cogn. Sci.* 23, 754–768. doi: 10.1016/j.tics.2019.06.009

Buzsáki, G. (2007). The structure of consciousness. *Nature* 446:267. doi: 10.1038/446267a

Calabrò, R. S., Cacciola, A., Bramanti, P., and Milardi, D. (2015). Neural correlates of consciousness: what we know and what we have to learn! *Neurol. Sci.* 36, 505–513. doi: 10.1007/s10072-015-2072-x

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.

Chalmers, D. (2007). "The hard problem of consciousness" in *The Blackwell companion to consciousness*. eds. M. Velmans and S. Schneider (Chichester, UK: Blackwell), 32–42.

Cooke, J. E. (2021). What is consciousness? Integrated information vs. inference. *Entropy* 23:1032. doi: 10.3390/e23081032

Crick, F., and Koch, C. (1998). Consciousness and neuroscience. *Cereb. Cortex* 8, 97–107. doi: 10.1093/cercor/8.2.97

Damasio, A. (2010). *Self comes to mind: Constructing the conscious brain*. New York: Pantheon Books.

Damasio, A., and Damasio, H. (2022). Homeostatic feelings and the biology of consciousness. *Brain* 145, 2231–2235. doi: 10.1093/brain/awac194

de Laguna, G. A. (1916). Sensation and perception. I: the genetic relationship. *J. Philos. Psychol. Sci. Methods* 13, 533–547. doi: 10.2307/2012564

de Leon, D. (2001). The qualities of qualia. *Commun. Cogn* 34, 121–138.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. New York: Viking Penguin.

Dehaene, S., Kerszberg, M., and Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA* 95, 14529–14534. doi: 10.1073/pnas.95.24.14529

Dennett, D. C. (1988). "Quining qualia" in *Consciousness in contemporary science*. eds. A. J. Marcel and E. Bisiach (Oxford: Clarendon Press), 42–77.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

Doesburg, S. M., Green, J. J., McDonald, J. J., and Ward, L. M. (2009). Rhythms of consciousness: binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception. *PLoS One* 4:e6142. doi: 10.1371/journal.pone.0006142

Dreyfus, H. L. (1992). *What computers still Can'T do: A critique of artificial reason*. Cambridge, MA: MIT Press.

Edelman, G. (2001). Consciousness: the remembered present. *Ann. N. Y. Acad. Sci.* 929, 111–122. doi: 10.1111/j.1749-6632.2001.tb05711.x

Edelman, G. M. (2003). Naturalizing consciousness: a theoretical framework. *Proc. Natl. Acad. Sci. USA* 100, 5520–5524. doi: 10.1073/pnas.0931349100

Edelman, G. M., and Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York: Basic Books.

Encyclopædia Britannica (2023). Quality. Available at: https://www.britannica.com/dictionar/quality (Accessed October 11, 2023).

Feldman, J. (2013). The neural binding problem(s). *Cogn. Neurodyn.* 7, 1–11. doi: 10.1007/s11571-012-9219-8

Fesce, R. (2023). Imagination: the dawn of consciousness: fighting some misconceptions in the discussion about consciousness. *Physiol. Behav.* 259:114035. doi: 10.1016/j.physbeh.2022.114035

Flanagan, O. J. (1992). *Consciousness reconsidered*. Cambridge, MA: MIT Press.

Forti, B. (2009). How could phenomenal consciousness be involved in mental function? *New Ideas Psychol.* 27, 312–325. doi: 10.1016/j.newideapsych.2008.10.001

Forti, B. (2021). Subjective experiences matter. What do we know about consciousness? *Evid. Based Psychiatr. Care* 7, 158–169. doi: 10.36180/2421-4469-2021-25

Friedman, G., Turk, K. W., and Budson, A. E. (2023). The current of consciousness: neural correlates and clinical aspects. *Curr. Neurol. Neurosci. Rep.* 23, 345–352. doi: 10.1007/s11910-023-01276-0

Gallagher, S., and Zahavi, D. (2008). *The phenomenological mind*. New York: Routledge

Gallotto, S., Sack, A. T., Schuhmann, T., and de Graaf, T. A. (2017). Oscillatory correlates of visual consciousness. *Front. Psychol.* 8:1147. doi: 10.3389/fpsyg.2017.01147

Galus, W. (2023a). Different aspects of consciousness explained by distinct biophysical processes. *J. Theor. Philos. Psychol.* doi: 10.1037/teo0000236

Galus, W. (2023b). Mind–brain identity theory confirmed? *Cogn. Neurodyn.* 17, 1–21. doi: 10.1007/s11571-023-09992-6

Galus, W., and Starzyk, J. (2020). *Reductive model of the conscious mind*. Hershey, PA: IGI Global isbn:13: 9781799856535.

Gärdenfors, P. (2006). *How Homo became sapiens: On the evolution of thinking*. Oxford: Oxford Academic (Accessed October 13, 2023).

Ghosh, H. (2020). "Early vision" in *Computational models for cognitive vision*. (New York: IEEE), 9–24. doi: 10.1002/9781119527886.ch2

Hameroff, S., and Penrose, R. (2014). Consciousness in the universe: a review of the 'Orch OR' theory. *Phys Life Rev* 11, 39–78. doi: 10.1016/j.plrev.2013.08.002

Hardin, C. L. (1992). "Physiology, phenomenology, and Spinoza's true colors" in *Emergence or reduction?: Essays on the prospects of nonreductive physicalism*. eds. A. Beckermann, H. Flohr and J. Kim (Berlin: De Gruyter), 201–219.

Haun, A., and Tononi, G. (2019). Why does space feel the way it does? Towards a principled account of spatial experience. *Entropy* 21:1160. doi: 10.3390/e21121160

He, B. J. (2023). Towards a pluralistic neurobiological understanding of consciousness. *Trends Cogn. Sci.* 27, 420–432. doi: 10.1016/j.tics.2023.02.001

Hirschhorn, R., Kahane, O., Gur-Arie, I., Faivre, N., and Mudrik, L. (2021). Windows of integration hypothesis revisited. *Front. Hum. Neurosci.* 14:617187. doi: 10.3389/fnhum.2020.617187

Humphrey, N. (2006). *Seeing red: A study in consciousness*. Harvard: Harvard University Press.

James, W. (1890). *The principles of psychology*. London: MacMillan.

Jansen, F. (2017). The hard problem of consciousness from a bio-psychological perspective. *Philos. Stud.* 7, 579–594. doi: 10.17265/2159-5313/2017.11.001

Jerath, R., and Crawford, M. W. (2014). Neural correlates of visuospatial consciousness in 3D default space: insights from contralateral neglect syndrome. *Conscious. Cogn.* 28, 81–93. doi: 10.1016/j.concog.2014.06.008

Jerath, R., Crawford, M. W., and Barnes, V. A. (2015). Functional representation of vision within the mind: a visual consciousness model based in 3D default space. *J. Med. Hypotheses Ideas* 9, 45–56. doi: 10.1016/j.jmhi.2015.02.001

Jones, M. W., and Hunt, T. (2023). Electromagnetic-field theories of qualia: can they improve upon standard neuroscience? *Front. Psychol.* 14:1015967. doi: 10.3389/fpsyg.2023.1015967

Kanai, R., and Tsuchiya, N. (2012). Qualia. *Curr. Biol.* 22, R392–R396. doi: 10.1016/j.cub.2012.03.033

Kanizsa, G. (1979). *Organization in vision. Essays in gestalt perception*. New York: Praeger.

Kanizsa, G. (1980). *Grammatica del vedere*. Bologna: Il Mulino.

Kanizsa, G. (1991). *Vedere e pensare*. Bologna: Il Mulino.

Kent, L., and Wittmann, M. (2021). Time consciousness: the missing link in theories of consciousness. *Neurosci. Conscious* 2021:niab011. doi: 10.1093/nc/niab011

Kind, A. (2008). "How to believe in qualia" in *The case for qualia*. ed. E. Wright (Cambridge, MA: MIT Press), 285–298.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Englewood, CO: Roberts & Company.

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446, 908–911. doi: 10.1038/nature05631

Kriegel, U. (2004). Consciousness and self-consciousness. *Monist* 87, 182–205. doi: 10.5840/monist20048725

Lamme, V. A. (2010). How neuroscience will change our view on consciousness. *Cogn. Neurosci.* 1, 204–220. doi: 10.1080/17588921003731586

Lamme, V. A. F. (2020). Visual functions generating conscious seeing. *Front. Psychol.* 11:83. doi: 10.3389/fpsyg.2020.00083

Lehnertz, K., Rings, T., and Bröhl, T. (2021). Time in brain: how biological rhythms impact on EEG signals and on EEG-derived brain networks. *Front. Netw. Physiol.* 1:755016. doi: 10.3389/fnetp.2021.755016

Lemon, R. N., and Edgley, S. A. (2010). Life without a cerebellum. *Brain* 133, 652–654. doi: 10.1093/brain/awq030

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pac. Philos. Q.* 64, 354–361. doi: 10.1111/j.1468-0114.1983.tb00207.x

Lewis, C. I. (1929). *Mind and the world order: Outline of a theory of knowledge*, New York: Charles Scribners.

Lightman, A., and Gingerich, O. (1992). When do anomalies begin? *Science* 255, 690–695. doi: 10.1126/science.255.5045.690

Loar, B. (2002). "Transparent experience and the availability of qualia" in *Consciousness: New philosophical perspectives*. eds. Q. Smith and A. Jokic (Oxford: Oxford University Press), 173–290.

Loorits, K. (2014). Structural qualia: a solution to the hard problem of consciousness. *Front. Psychol.* 5:237. doi: 10.3389/fpsyg.2014.00237

Luccio, R. (2011). Gestalt psychology and cognitive psychology. *Humana Mente* 17, 95–128.

Ludwig, D. (2023). The functions of consciousness in visual processing. *Neurosci. Conscious.* 2023:niac018. doi: 10.1093/nc/niac018

Maillé, S., and Lynn, M. (2020). Reconciling current theories of consciousness. *J. Neurosci.* 40, 1994–1996. doi: 10.1523/JNEUROSCI.2740-19.2020

McFadden, J. (2020). Integrating information in the brain's EM field: the cemi field theory of consciousness. *Neurosci. Conscious.* 2020:niaa016. doi: 10.1093/nc/niaa016

McFadden, J. (2023). Consciousness: matter or EMF? *Front. Hum. Neurosci.* 16:1024934. doi: 10.3389/fnhum.2022.1024934

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Metzinger, T. (2004). *Losing your self being no-one: The self-model theory of subjectivity*. Cambridge: The MIT Press.

Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450. doi: 10.2307/2183914

Nani, A., Manuello, J., Mancuso, L., Liloia, D., Costa, T., and Cauda, F. (2019). The neural correlates of consciousness and attention: two sister processes of the brain. *Front. Neurosci.* 13:1169. doi: 10.3389/fnins.2019.01169

Noel, J. P., Ishizawa, Y., Patel, S. R., Eskandar, E. N., and Wallace, M. T. (2019). Leveraging nonhuman primate multisensory neurons and circuits in assessing consciousness theory. *J. Neurosci.* 39, 7485–7500. doi: 10.1523/JNEUROSCI.0934-19.2019

Northoff, G., Klar, P., Bein, M., and Safron, A. (2023). As without, so within: how the brain's temporo-spatial alignment to the environment shapes consciousness. *Interface Focus* 13:20220076. doi: 10.1098/rsfs.2022.0076

Northoff, G., and Zilio, F. (2022). Temporo-spatial theory of consciousness (TTC) - bridging the gap of neuronal activity and phenomenal states. *Behav. Brain Res.* 424:113788. doi: 10.1016/j.bbr.2022.113788

O'Callaghan, C. (2015). The multisensory character of perception. *J. Philos.* 112, 551–569. doi: 10.5840/jphil20151121035

Parvizi, J., and Damasio, A. (2001). Consciousness and the brainstem. *Cognition* 79, 135–160. doi: 10.1016/S0010-0277(00)00127-X

Polák, M., and Marvan, T. (2018). Neural correlates of consciousness meet the theory of identity. *Front. Psychol.* 9:1269. doi: 10.3389/fpsyg.2018.01269

Raffone, A., Srinivasan, N., and van Leeuwen, C. (2014). Perceptual awareness and its neural basis: bridging experimental and theoretical paradigms. *Philos. Trans. R. Soc. B* 369:20130203. doi: 10.1098/rstb.2013.0203

Railo, H., Koivisto, M., and Revonsuo, A. (2011). Tracking the processes behind conscious perception: a review of event-related potential correlates of visual consciousness. *Conscious. Cogn.* 20, 972–983. doi: 10.1016/j.concog.2011.03.019

Reid, T. (1764/1977) in *An inquiry into the human mind on the principles of common sense*. ed. D. R. Brookes (Edinburgh: Edinburgh University Press)

Sanfey, J. (2023). Simultaneity of consciousness with physical reality: the key that unlocks the mind-matter problem. *Front. Psychol.* 14:1173653. doi: 10.3389/fpsyg.2023.1173653

Schwarzkopf, D. S., and Rees, G. (2015). "Perceptual organization and consciousness" in *The Oxford handbook of perceptual organization*. ed. J. Wagemans (Oxford: Oxford University Press), 799–819.

Seager, W. (2002). Emotional introspection. *Conscious. Cogn.* 11, 666–687. doi: 10.1016/s1053-8100(02)00027-2

Searle, J. R. (1997). *The mistery of consciousness*. New York: The New York Review of Books.

Searle, J. R. (2000). Consciousness. *Annu. Rev. Neurosci.* 23, 557–578. doi: 10.1146/annurev.neuro.23.1.557

Searle, J. R. (2004). *Mind. A brief introduction*. New York: Oxford University Press.

Seth, A. K., and Bayne, T. (2022). Theories of consciousness. *Nat. Rev. Neurosci.* 23, 439–452. doi: 10.1038/s41583-022-00587-4

Sheinberg, D. L., and Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. USA* 94, 3408–3413. doi: 10.1073/pnas.94.7.3408

Shoemaker, S. (1991). Qualia and consciousness. *Mind* C, 507–524. doi: 10.1093/mind/C.400.507

Siddharth, S., and Menon, S. (2017). "What does it mean for qualia to be intrinsic?" in *Self, culture and consciousness*. eds. E. Menon, N. Nagaraj and V. Binoy (Singapore: Springer)

Skokowski, P. (2022). Sensing Qualia. *Front. Syst. Neurosci.* 16:795405. doi: 10.3389/fnsys.2022.795405

Smith, D. W. (2018). "Phenomenology" in *The Stanford encyclopedia of philosophy (Summer 2018 Edn.)*. ed. E. N. Zalta Available at: https://plato.stanford.edu/archives/sum2018/entries/phenomenology

Solms, M. (2021). *The hidden spring: A journey to the source of consciousness*. New York: Norton.

Srinivasan, N. (2020). Consciousness without content: a look at evidence and prospects. *Front. Psychol.* 11:1992. doi: 10.3389/fpsyg.2020.01992

Tegmark, M. (2000). Importance of quantum decoherence in brain processes. *Phys. Rev. E* 61, 4194–4206. doi: 10.1103/PhysRevE.61.4194

Todorovic, D. (2008). Gestalt principles. *Scholarpedia* 3:5345. doi: 10.4249/scholarpedia.5345

Tomasi, C. (2006). "Early vision" in *Encyclopedia of Cognitive Science*

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242. doi: 10.2307/25470707

Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846

Tononi, G., and Koch, C. (2008). The neural correlates of consciousness: an update. *Ann. N. Y. Acad. Sci.* 1124, 239–261. doi: 10.1196/annals.1440.004

Tononi, G., and Koch, C. (2015). Consciousness: here, there and everywhere? Philos. *Trans. R. Soc. Lond. B Biol. Sci.* 370, 1–18. doi: 10.1098/rstb.2014.0167

Tuszynski, J. A. (2020). From quantum chemistry to quantum biology: a path toward consciousness. *J. Integr. Neurosci.* 19, 687–700. doi: 10.31083/j.jin.2020.04.393

Tyler, C. W. (2020). Ten testable properties of consciousness. *Front. Psychol.* 11:1144. doi: 10.3389/fpsyg.2020.01144

Van der Werf, Y. D., Witter, M. P., and Groenewegen, H. J. (2002). The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Res. Brain Res. Rev.* 39, 107–140. doi: 10.1016/S0165-0173(02)00181-9

Van Gulick, R. (2017). "Functionalism and qualia" in *The Blackwell companion to Consciousnes*. eds. S. Schneider and M. Velmans (Oxford, UK: Blackwell), 430–444.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., et al. (2012). A century of gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychol. Bull.* 138, 1218–1252. doi: 10.1037/a0029334

Ward, L., and Guevara, R. (2022). Qualia and phenomenal consciousness Arise from the information structure of an electromagnetic field in the brain. *Front. Hum. Neurosci.* 16:874241. doi: 10.3389/fnhum.2022.874241

Wiese, W. (2017). How to solve the problem of phenomenal unity: finding alternatives to the single state conception. *Phenom. Cogn. Sci.* 16, 811–836. doi: 10.1007/s11097-016-9478-7

Williford, J. R., and von der Heydt, R. (2013). Border-ownership coding. *Scholarpedia J.* 8:30040. doi: 10.4249/scholarpedia.30040

Zeman, A. (2001). Consciousness. *Brain* 124, 1263–1289. doi: 10.1093/brain/124.7.1263

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 21 March 2024
DOI 10.3389/fpsyg.2024.1283980

Check for updates

OPEN ACCESS

EDITED BY
Luca Simione,
UNINT - Università degli studi Internazionali di Roma, Italy

REVIEWED BY
Tina Lindhard,
International University of Professional Studies, United States
Charlène Aubinet,
University of Liège, Belgium

*CORRESPONDENCE
Yuane Jia
joy.y.jia@rutgers.edu

†These authors have contributed equally to this work and share first authorship

RECEIVED 27 August 2023
ACCEPTED 16 February 2024
PUBLISHED 21 March 2024

CITATION
Jia Y, Schenkman M, O Connor H, Jayanna K, Pearmain R, Van't Westeinde A and Patel KD (2024) Validation of the Awareness Atlas—a new measure of the manifestation of consciousness.
Front. Psychol. 15:1283980.
doi: 10.3389/fpsyg.2024.1283980

Validation of the Awareness Atlas—a new measure of the manifestation of consciousness

Yuane Jia[1*†], Margaret Schenkman[2†], Hester O Connor[3], Krishnmurthy Jayanna[4,5], Rosalind Pearmain[6], Annelies Van't Westeinde[7] and Kamlesh D. Patel[8]

[1]Department of Interdisciplinary Studies, School of Health Professions, Rutgers University, Newark, NJ, United States, [2]Department of Physical Medicine and Rehabilitation, Physical Therapy Program, University of Colorado, Aurora, CO, United States, [3]Retired, Kildare, Ireland, [4]Division of Public Health, Faculty of Life and Allied Health Sciences, Ramaiah University of Applied Sciences, Bangalore, India, [5]Center for Integrative Health and Wellbeing, Bangalore, India, [6]Minster Centre, London, United Kingdom, [7]Department of Women's and Children's Health, Unit for Pediatric Endocrinology, Karolinska Institutet, Stockholm, Sweden, [8]Heartfulness Institute, Kanha Shanti Vanam, Hyderabad, India

Consciousness has intrigued philosophers and scholars for millennia and has been the topic of considerable scientific investigation in recent decades. Despite its importance, there is no unifying definition of the term, nor are there widely accepted measures of consciousness. Indeed, it is likely that consciousness—by its very nature—eludes measurement. It is, however, possible to measure how consciousness manifests as a lived experience. Yet here, too, holistic measures are lacking. This investigation describes the development and validation of the Awareness Atlas, a measure of the manifestation of consciousness. The scale was informed by heart-based contemplative practices and the resulting lived experience with a focus on the impacts of manifestation of consciousness on daily life. Four hundred forty-nine individuals from the USA, Canada, India, and Europe participated in psychometric testing of the scale. Exploratory and confirmatory factor analyses were used for validation, demonstrating excellent validity in measuring manifestation of consciousness. The final model fit exceeded all required thresholds, indicating an excellent fitted model with a single dimensionality to measure the manifestation of consciousness comprised of four subscales: Relationship to Others; Listening to the Heart; Connection with Higher Self; and Acceptance and Letting Go. Number of years meditating and practicing Heartfulness meditation were positively related to the total and subscale scores. Test–retest reliability was excellent for the total scale, and good to excellent for the four subscales. Findings demonstrate that the Awareness Atlas is a well-constructed tool that will be useful in examining changes in manifestation of consciousness with various experiences (e.g., meditation, life-altering conditions).

KEYWORDS

scale development, validation, new measure, manifestation of consciousness, awareness, Heartfulness meditation

Introduction

Consciousness has intrigued philosophers and scholars for millennia and has been the focus of considerable scientific investigation in recent decades (Garcia-Castro and Kodukula, 2022). There is increasing interest in the role consciousness plays with respect to health and well-being (George et al., 2000; Bożek et al., 2020; Jayanna, 2021). Further, there is heightened realization that a shift in human consciousness is necessary to address global challenges such as societal conflicts, climate change, and global warming (Kirsmayer, 2010; Nasibulina, 2015). To these ends, various approaches have been employed in recent years to expand consciousness (e.g., educational, psychological, religious and spiritual) (Vieten et al., 2006; Thiengkamol, 2011; Savelyeva and Douglas, 2017; Patel, 2023).

Exploration of how consciousness manifests in daily life, and of intentional changes in consciousness, requires a spectrum of manifestations from the personal to the universal. Such explorations are limited by the lack of a simple measure that incorporates the range of manifestations of consciousness experienced by individuals. Against this backdrop, we believe research is necessary to develop and validate a tool to assess shifts in human consciousness.

The problem of defining consciousness

Conceptualizing consciousness

Consciousness has been characterized in numerous ways without consensus (Guertin, 2019; Garcia-Castro and Kodukula, 2022). A few examples illustrate the diverse descriptions and resulting difficulties with measuring this construct.

Early discussions of consciousness come from Eastern traditions. For example, four states of consciousness were described in the Mandukya Upandishad (Srinivasan, 2022): wakefulness, dreaming, sleep, and the Turiya state, which transcends the first three states: "*perceiving neither what is inside nor what is outside… neither as perceiving nor as not perceiving… as unthinkable; as indescribable…*" [quoted from Olivelle (Srinivasan, 2022)]. The "state of pure consciousness" described in the Buddhist tradition appears to be similar to the Turiya state described in the Upanishads (Srinivasan, 2022). Heartfulness identifies a fifth state, the Turiyatit state, which extends the Turiya state experienced in meditation out into everyday life (Patel, 2019).

Consciousness can be considered to move along a spectrum from zero to infinity (Wilber, 2000; van't Westeinde and Patel, 2022). Drawing from Patanjali and the Vedas, van't Westeinde and Patel noted that the goal of yoga and other contemplative traditions, "Union with Universal Consciousness," is accomplished through purifying and expanding individual consciousness to such an extent that it merges with the universal. Some investigators use the term "pure awareness" in the context of consciousness. For example, Gamma and Metzinger (2021) suggested that pure awareness is characterized by "an absence of space or time and body sense" and by the experience of "peacefulness" and "unboundedness." These authors used the term "minimal phenomenal experience" to describe the condition. Thomas et al. (2022) suggested that "The original cause [of consciousness] is nothingness." Some contemplative practices talk about pure consciousness as consciousness without content, while another term, "non-conceptual representational content," has been proposed in the context of consciousness (Srinivasan, 2022).

In contrast, from a Western perspective, Kihlstrom (2022) suggested that consciousness has to do with monitoring ourselves and our environment, which is linked to voluntarily starting and ending both mental and behavioral activities. This monitoring allows for memory, thoughts, feelings, emotions, and desires to be represented in phenomenological awareness, suggesting that consciousness and self are inexorably intertwined.

Wilber recognized the importance of integrating Eastern philosophies with contemporary Western approaches to psychological development (Wilber, 2000). To this end, he drew on the work of a variety of scholars (e.g., Loevinger, 1976; Kegan, 1982) and attempted to integrate spiritual philosophies, adult developmental psychology (including development of moral reasoning and ethics), humanistic, transpersonal, and positive psychologies. His spectrum of consciousness explores a span from everyday consciousness to the Turyatit state of yogis (Wilber et al., 1986; Wilber, 2000).

This sample of descriptions above illustrates that consciousness encompasses a spectrum from simple awareness of self and others to pure consciousness that is without content or that has non-conceptual representational content. In other words, consciousness in its most refined state is beyond description. Clearly, there are significant and substantial challenges to measuring such an ill-defined construct, especially if one includes a state without content.

Manifestation of consciousness

An alternative and more feasible approach to understanding consciousness is to study its expression, which we refer to as the "manifestation of consciousness." While consciousness itself cannot be easily defined, its effects on human behavior, thoughts, and feelings can be characterized. The awareness of the thoughts and feelings of oneself and others may grow as one experiences expansion of consciousness. This is one aspect of an expanding consciousness, and one that we will focus upon for this study. From this perspective, expanding consciousness can be considered as an expansion of one's awareness of self and others. By examining the manifestation of consciousness, it is possible to identify relevant phenomena that can be measured quantitatively. This approach has been adopted across multiple disciplines, including medicine and neurology (Teasdale and Jennett, 1974; Sattin et al., 2021), psychology and behavior (Turner et al., 1978; Grant et al., 2002), sociology and human development (Carver and Glass, 1976; Diemer et al., 2017).

As with consciousness itself, the manifestation of consciousness is characterized differently depending on the discipline (Guertin, 2019). For example, anesthesiologists consider consciousness in terms of alertness and awareness when determining whether a patient is sufficiently "unconscious" to perform surgery. To determine the level of consciousness post traumatic brain injury, medical professionals typically assess eye opening, ability to speak, and ability to move on request (e.g., as measured with the Glasgow Coma Scale) (Teasdale and Jennett, 1974). Behavioral scientists can conceptualize and investigate changes in consciousness in the context of awareness/perception of self or others, with various interventions as a means of examining that aspect of increasing awareness or consciousness. Thus, studies to examine changing manifestations of consciousness (e.g., empathy) can be designed and used to draw inferences about changing states of consciousness with different situations, interventions, and life experiences. By examining the manifestation of consciousness in these contexts, it should be possible to begin to appreciate the practical role of expanding consciousness in daily life.

The above approaches do not address the expression of consciousness described in contemplative practices (mostly of Eastern origin) as "non-conceptual representational content," "consciousness without content," or "universality" (Srinivasan, 2022). Literature from transpersonal psychology focuses on consciousness from the perspective of self-transcendence, which comes closer to these concepts (MacDonald and Friedman, 2013; Kitson et al., 2020). Macdonald and Friedman (2013) suggested that transcendence is a state that can be appreciated, but cannot easily be put into words, and yet must be investigated through quantitative methods in order to study its impact. Kitson et al. (2020) pointed out that any investigation of self-transcendence also needs to take into account the discipline investigating its impact (e.g., psychology, gerontology, nursing) because manifestation of consciousness, like consciousness itself, is multi-faceted. Each discipline brings a different focus to the issues.

Meditation and its impact on consciousness

Meditation and yoga have long been used as tools to change states of consciousness in Eastern spiritual traditions (e.g., as described in the Yoga Sutras of Patanjali) (Deshpande, 2021). In the recent past, studies have been undertaken to investigate the effects of meditation on states of mind and consciousness. The state and trait effects of meditation are found to have implications on the neuroscience of attention, consciousness, self-awareness, and empathy (Raffone and Srinivasan, 2010; Sparby, 2015; Tripathi and Bharadwaj, 2021).

During the 20th century, several meditation approaches based on ancient practices have emerged that are specifically focused on the heart (Lindhard, 2018a; van't Westeinde and Patel, 2022; Patel 2023). A heart-based approach can lead a person through different levels of consciousness. Examples include the Arka Dhyana method, developed by Srinivas Arka, which has some similarities to Prayer of the Heart (Lindhard, 2017, 2018a,b) and Heartfulness Meditation (Patel and Pollock, 2018; Patel, 2023). Although each approach recognizes a role of the heart in the development and expression of different levels of consciousness, they have different theoretical frameworks and descriptions of the journey. For example, Arka Dhyana refers to stages of consciousness progressing through six levels from a heart-mind stage to a sixth stage of Pure Self, described as "consciousness without content." The Arka Dhyana method uses developing awareness of intuitive feeling (as opposed to physical feeling) to travel this journey. The six levels convey a distinct separation of states.

In contrast, Heartfulness Meditation, a form of Raja Yoga, refers to a gradual expansion of consciousness through various layers or regions, which are uncovered by the removal of old impressions (referred to by the Sanskrit term *samskaras*), and awakened by the impulse of Transmission during meditation. This expansion is not limited to consciousness—it goes beyond consciousness, although we will stay within the spectrum of consciousness in this study. This expansion of consciousness conveys a fluid and dynamic change in consciousness as well as a progression through distinct and separate states. The goal of arriving at a state of oneness requires the refinement of the mind beyond thinking to states of feeling, being, and non-being; the refinement of the intellect to intelligence, wisdom, and intuition; and the refinement of the ego to ever-increasing states of insignificance, love, and humility.

In concept, Heartfulness Meditation and Arka Dhyana (and the ancient practices from which derive) focus on a similar endpoint, which is acquired differently and named differently (e.g., pure self, a state of contentless content, union with the absolute, oneness with all). It is noteworthy that the methods to reach that goal differ between the practices.

Changes associated with heart-based meditation practices provide a platform for studying expansion of consciousness from a focus on self to a focus on the universal. Further, these practices form a platform for investigating their resulting impact on daily life (van't Westeinde and Patel, 2022).

The Awareness Atlas was framed around experience with Heartfulness Meditation. Heartfulness Meditation arose in the early 20th century under the name Sahaj Marg, meaning "the natural path," and is discussed in detail elsewhere (Patel and Pollock, 2018; van't Westeinde and Patel, 2022; Patel, 2023). While its roots are in the yogic tradition, it is a new approach, distilling the essence of earlier yogic practices. Heartfulness describes consciousness as the degree of awareness, which also means the degree of unawareness. The Heartfulness practices take a practitioner through changes in mental states and gradually expanding consciousness, and these changes occur through a journey that was originally described by Chandra (1989) and more recently by Patel (2023).

The Awareness Atlas reflects the transition during this journey of expanding consciousness from a focus on self, to a focus that includes others, and eventually that encompasses the oneness of all. This expansion can be understood in terms of increasing awareness and decreasing unawareness.

Pearmain et al. (2023) conducted a phenomenological investigation of the Heartfulness practices by asking participants to describe if and how they noticed the impact of practice as lived experiences in everyday life. Their findings revealed that participants experienced changes in patterns of relating to oneself, others, and the world around them. The findings also identified that participants actively sustained meditative awareness in their hearts. Five core intertwined themes emerged: (1) a shift in focus from thinking to feeling (from head to heart); (2) developing a sense of stability and groundedness (because of being anchored in the heart); (3) being flexible and open, rather than resisting change; (4) sustaining a connection with the heart in the midst of life experiences; (5) immersing oneself in the heart as a space to retreat into, where one is immersed in love and feels empowered.

This phenomenological study provides unique insights into the lived experiences of Heartfulness practitioners and brings into focus ideas of how consciousness might expand in response to a heart-based meditation. However, quantitative methodology is also needed to fully appreciate what expansion of consciousness entails, how it evolves, and the impact on daily life. A combination of quantitative and phenomenological approaches will facilitate exploration of the impact of contemplative practices, such as Heartfulness and others, on expansion of consciousness. Such a combined approach can also be used to explore relationships between changes in consciousness and changes in health, well-being, and development. To this end, it was important to find outcome measures that capture consciousness along the spectrum from a focus on individuality to a focus on oneness with the whole.

Measures related to consciousness

A number of quantitative measures exist that could be applied to capture changes in the manifestation of consciousness. However, there is a gap: none captures manifestation of consciousness as expressed in daily life through the broad spectrum from individual focus to universal focus.

MacDonald and Friedman (2013) reviewed measures from the perspective of transpersonal psychology and presented them in the following categories: spirituality; well-being; experience and consciousness; beliefs, orientation, and identity. Kitson et al. (2020) summarized the psychometric properties of specific characteristics (e.g., gratitude, positive emotions) and for the impact of mindfulness meditation more broadly.

Gamma and Metzinger (2021) developed the Minimal Phenomenal Experience questionnaire (MPE-92), designed to measure the altered states of consciousness experienced by meditators. The MPE-92 has 12 factors that characterize the meditation experience, including time effort and desire; thoughts and feelings; emptiness and non-egoic awareness. A few other scales are available for measuring altered states of consciousness and transcendence specifically related to the effects of psychotropic drugs (Griffiths et al., 2006; Kitson et al., 2020).

Lindhard (2017) developed the Feeling Consciousness Scale based on the impact of the Intuitive Meditation (IM) practice, also referred to as the Arka Dhyana method of meditation. This scale focuses predominantly on identifying changing feelings related to intuition, calmness, and bodily sensations, although a few of the questions relate to the person's behaviors.

These different measures each focus on a specific aspect of the characteristics in question (e.g., altered states of consciousness, ego, feelings that arise). None of the measures identified by MacDonald and Friedman (2013), nor those identified by Kitson et al. (2020), encompass the spectrum of awareness associated with consciousness as expressed in daily life. The MPE-92 describes the experience of pure awareness as a phenomenon rather than an expression of daily life, and the Feeling of Consciousness scale focuses predominantly on feelings.

The purpose of this investigation

This investigation was designed to develop a valid measure of the manifestation of consciousness, from the perspective of the Heartfulness practices, that could be applied: (1) to study how shifts in manifestation of consciousness relate to health, well-being, peace, and inclusivity; and (2) to better understand the role of the heart-based practices in expanding manifestation of consciousness.

There is no universally accepted definition of the term consciousness or its manifestations. In this manuscript, we began with Patel's definition of consciousness as "degree of awareness, which also means degree of unawareness" (van't Westeinde and Patel, 2022) and adopted an operational definition of manifestation of consciousness as a "state of awareness of self, others, and the connection among all beings." We took into consideration that consciousness moves along a spectrum, and that expansion of consciousness can result from life experiences that can be viewed as a journey (e.g., normal development, the consequences of life challenges, intentional expansion through activities such as meditation).

Methods

Scale development

Item generation

In developing this scale, we were cognizant of Western concepts of consciousness, as well as concepts of consciousness from yoga generally (Deshpande, 2021) and the Heartfulness practice specifically. Heartfulness meditation is centered on connection with the heart with an objective of purifying and expanding consciousness. This includes the evolution of thinking to feeling, intelligence to wisdom and intuition, and ego to humility and love (van't Westeinde and Patel, 2022). Thus, the scale was anticipated to reflect these concepts.

The specific questions included in this measure were based on transcripts from the phenomenological study (Pearmain et al., 2023). Participants in the study were asked the following question: "How does Heartfulness meditation practice affect or permeate your daily life?" The participants then offered their observations without prompting from the interviewer. The transcripts from the phenomenological study were reviewed by one of the investigators of the research team who identified words, concepts, and constructs that could be included in a consciousness scale. Thus, the questions on the Awareness Atlas were derived from words used by people who had practiced Heartfulness Meditation rather than from any preconceived notions of the investigators of this current study. From their review, two members of the research team developed 48 questions for inclusion, including four reverse scored questions. These questions incorporated a broad perspective about expansion of consciousness from the personal and relational items to those that focused on awareness of something beyond oneself. They also specifically reflected the role of the heart in expansion of consciousness. All 48 questions were reviewed, edited, and modified by the entire research team for clarity. All except one member of the team (YJ) practice Heartfulness meditation with experience ranging from 7 to 48 years (average 33 years).

The initial scale consisted of 48 items with 7 themes: (1) Perception of Self; (2) Trust in Self and Others; (3) Relationship to Others; (4) Acceptance of Self; (5) Listening to the Heart; (6) Connection with Higher Self; (7) Acceptance and Letting Go. Four of the items were reverse scored. These items were located in different themes: one reversed question was in Theme 3 Relationship to Others (I strive to get my own way over the needs of others); two were in Theme 4 (It is hard for me to look at my thoughts, actions, and words; I do not feel good about myself); and one reversed question was in Theme 6 (I do not feel connected with a reality that is larger than myself).

Content validity

Once the questions had been agreed upon, an iterative process of refining each question was conducted by the researchers. To ensure content and face validity, the questions were piloted through 12 one-on-one interviews to determine how the interviewees understood the questions. We included people who had a meditation practice (Heartfulness [$n = 3$], Buddhist [$n = 1$], Mindfulness [$n = 4$]) or had no meditation practice ($n = 4$). The entire research team then reviewed all suggestions and finalized the wording for the initial 48 questions.

The final survey consisted of an informed consent, followed by demographics variables (e.g., age, sex, education level, geographic location, ethnicity, and meditation practices), and then the 48 questions. The participants were instructed to consider how they actually behave, rather than simply whether they agree or disagree with the statement by reflecting on their experiences over the past 2 weeks. A Likert scale of 1 (never) to 6 (always) was used to rate each statement.

Testing of the scale

Procedure and participants

Institutional review board approval was obtained from the University of Colorado for worldwide distribution with the exception of the European Union (EU). A separate informed consent was approved by the University of Colorado for EU participants. All participants gave informed consent in order to access the survey.

A total of 433 adults aged 18–86 from USA, Canada, India, Europe, and other countries participated in the scale validation process. An invitation was sent to coordinators of Heartfulness Meditation in the participating countries for distribution to meditators in that practice. The invitation invited voluntary participation in an anonymous survey. The invitation was also sent to colleagues, friends, and family members of Heartfulness practitioners, inviting participation from people who participated in other meditation practices, or had no meditation practice.

Exploratory and confirmatory factor analysis

Factor analysis is commonly used to establish construct validity. Exploratory factor analysis (EFA) looks at the structure within a large number of items by grouping highly correlated items into factors that reflect different theoretical components of the overall construct (Portney, 2020). After initial screening of the data for missing values and distribution, we conducted an EFA ($n = 203$) using a Principal Axis Factoring as extraction method and promax as rotation method in spss 28. Item screening was based on factor loadings and communalities. Decisions regarding which factors would be retained were made using the following criteria: eigenvalue was greater than 1, inspection of the scree plot, number of items in each factor and factor loadings are all greater than 0.30 on at least one factor, and also taking into account the interpretability of each factor. Internal consistency reliability was assessed using item-scale correlations and the Cronbach alpha (α).

Confirmatory factor analysis (CFA) is used to determine whether the theoretical structure of an instrument fits with current empirical understanding of the construct (Portney, 2020). A second sample of 230 participants was used to conduct CFA in Mplus 7.3. Maximum likelihood with robust error (MLR) estimator was utilized to account for missing and potential non-normal distribution. A second order factor CFA was tested in Mplus as well. We used the most commonly used techniques to evaluate the model fit. This methodology included the chi-square test of exact fit, Root Mean Square Error of Approximation (RMSEA $\leq$0.06), Tucker Lewis Index (TLI $\geq$ 0.95), Comparative Fit Index (CFI $\geq$ 0.95), and Standardized Root Mean Square Residual (SRMR $\leq$0.08) (Hu and Bentler, 1999). Additionally, the average variance extracted (AVE) was used to assess the convergent validity. If the square root of the AVE of each latent factor is greater than the correlation coefficients between that latent variable and other latent variables in the measurement model, then the model shows evidence of discriminant validity (Fornell and Larcker, 1981).

Known-groups validity

Known-groups validity provides further evidence of construct validity, which is provided when the scale can distinguish differences between two or more groups with anticipated differences (Portney, 2020). Known-groups validity of the scale was assessed by examining the association among demographics such as age, sex, type of meditation practice, years of experience of meditation practice and the scale scores. We hypothesized that people with more years of meditation experience would have higher scores on the scale, and other factors would not be highly correlated with scale scores. Further, considering the factors that might affect the scale score collectively vs. independently, a linear regression model with all the above-mentioned factors was entered to explore how those important demographics all together affect the scale score.

Test–retest reliability

We used test–retest reliability to examine stability of responses to the scale over time. For this purpose, the final scale was sent to a group of voluntary participants who were asked to respond to the survey twice within a two-week period. The test–retest reliability was calculated using an intraclass correlation coefficient with two-way mixed-effects model and absolute agreement for the total scale score and subscale scores (Koo and Li, 2016).

Results

Item reduction and exploratory factor analysis (EFA) results

Data from the first 203 participants were used to conduct the EFA. Participants were 60.6% female, with about half the sample from USA, 24% from Canada, and 26% from India. The mean age was 51.6 (SD = 13.1) and median age was 52.0. Years of meditation ranged from 0 to 50. The mean and median years of meditation were 16.7 (SD = 11.2) and 18, respectively. 83.7% of meditators practiced Heartfulness Meditation, 10.3% had a variety of meditation and spiritual practices, while 5.9% of the sample reported they did not have a meditation practice.

Data screening and cleaning were conducted before running the EFA, including reverse coding for the four reverse wording items. Initial EFA yielded 9 factors with eigenvalue >1; from inspection of the scree plot, a 5-factor solution was determined to be reasonable. Four items were dropped due to very low communality resulting in a five-factor model. The rotated factor structure and loadings revealed that several items had very close double loadings or low loadings (<0.3) at any of the five factors, so those items were dropped.

With the iterative process of the EFA and repetitive evaluation of the factor structure pattern and loadings, and also taking into account interpretability, we arrived at the final 30-item, 5-factor model indicated by the footnote in Table 1. We labeled Factor 1 as Relationship to Self and Others (5 items), Factor 2 as Judgment of Self (4 items), Factor 3 as Listening to the Heart (5 items), Factor 4 as Connection with Higher Self (6 items), Factor 5 as Acceptance and Letting Go (10 items), respectively. These five factors are clearly distinct with very little overlap. The inter-factor correlation coefficients ranged from 0.10 to 0.64 (see Table 2). The five-factor solution with initial eigenvalues ranging from 11.43 to 1.39, explained 61.51% of the total variance.

Confirmatory factor analysis (CFA) results

A second sample of 230 participants was used for the CFA. Participants were 68.7% female, with 56% of the sample from Europe, 25% from USA, 13% from India, 4.3% from Canada, and 1.7% from other countries. The mean age was 54.2 (SD = 13.4) and median age was 55.5. Years of

10.3389/fpsyg.2024.1283980

TABLE 1 EFA results with five factor model (30 items).

Item description	F1	F2	F3	F4	F5
1. I am receptive to the feelings, needs, and suffering of others.	0.546				
2. I notice how others react to me at the time of an interaction.	0.651				
3. I notice my reactions to others at the time of an interaction.	0.543				
4. I consider the feelings, needs, and suffering of others.	0.546				
5. Throughout the day I notice how my thoughts, feelings, and perceptions fluctuate.	0.340				
6. It is hard for me to look at my thoughts, actions, and words.		0.750			
7. I do not feel connected with a reality that is larger than myself.		0.645			
8. I do not feel good about myself.		0.573			
9. I strive to get my own way over the needs of others.		0.471			
10. I listen to the wisdom of my heart (the wisdom that arises when my preconceived notions, desires, judgment, and emotions are silenced) and trust what it tells me.			0.837		
11. I feel guided in life, by the wisdom of my heart.			0.814		
12. When making decisions and interacting with others, it is easy for me to connect with the wisdom of my heart.			0.541		
13. To make decisions in any situation, my heart (inner wisdom) guides me from a place beyond emotion and thought.			0.539		
14. I trust my intuition.			0.588		
15. I feel supported by a deeper reality, underlying all of creation.				0.719	
16. I feel that I am part of something greater than myself.				0.885	
17. I feel a spiritual aspect to my identity, beyond my worldly identity.				0.633	
18. I feel that my consciousness is ever expanding.				0.743	
19. I have a feeling of wonder and awe about life.				0.492	
20. I have a sense of being one with all beings in the universe				0.470	
21. I cheerfully embrace situations that are hard, uncomfortable, or challenging.					0.965
22. I cheerfully adapt to life circumstances in order to grow.					0.913
23. I embrace all experiences of my life with joy as they unfold.					0.694
24. I accept the struggles and lessons in life.					0.756
25. I use my self-awareness to realize I have choices in how to respond to situations.					0.608
26. My emotions, feelings, and thoughts remain balanced (stable) no matter what is going on within and around me.					0.585
27. I consider the perspectives of others and learn from them.					0.555
28. As my awareness and consciousness change, I adapt my behaviors in order to be compatible with these changes.					0.527
29. I trust that all will work out as is necessary, even in the most difficult situations.					0.531
30. I have a feeling of gratitude no matter what I encounter throughout the day.					0.516
Initial Eigenvalues	1.513	1.394	1.816	2.302	11.428
% variance explained	5.045	4.647	6.052	7.674	38.092

Factor loadings < 0.3 was not showing in the table; items in factor 2 were reverse coded. F1 = Relationship to Self and Others (5 items), F2 = Judgment of Self (4 Items), F3 = Listening to the Heart (5 items), F4 = Connection with Higher Self (6 items), F5 = Acceptance and Letting Go (10 items).

meditation ranged from 0 to 58. The mean and median years of meditation were 18.1 (SD = 12.1) and 19, respectively. 78.9% of meditators practiced Heartfulness Meditation, 16% had a variety of meditation and spiritual practices, while 5.2% did not have a meditation practice.

A five-factor model was specified initially, and items 5, 6 and 9 had factor loadings <0.30, so all three items were dropped from the CFA analysis. These two dropped items were originally from the factor Judgment of Self in the EFA five-factor model. After dropping two items in the factor Judgment of Self, only items 7 and 8 remained. Following a common practice with a minimum of three items in each factor, a four-factor CFA model was also specified. Moreover, modification indices indicated items 29 and 30 could potentially load on another factor—Connection with Higher Self. An iteration of model comparisons was carried out among four-factor and five-factor

TABLE 2 Inter-factor correlations with EFA five factor model factor correlation matrix.

Factor	1	2	3	4	5
1. Relationship to Self and Others	1.000				
2. Judgment of Self	0.605	1.000			
3. Listening to the Heart	0.644	0.566	1.000		
4. Connection with Higher Self	0.574	0.428	0.470	1.000	
5. Acceptance and Letting Go	0.098	0.185	0.141	0.215	1.000

Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.

models, as well as keeping or dropping items 7 and 8, or moving 29 and 30 to a different factor, resulting in the best model with a total of 23 items with four factors: Relationship to Others (RO, 5 items), Listening to the Heart (LH, 5 items), Connection with Higher Self (CHS, 6 items), Acceptance and Letting Go (ALG, 7 items). See the final scale and factor loadings in Table 3. The final scale with instructions is available in the Supplementary material.

Of note, the four reverse scored items were designed to belong to different themes at the early stage of the scale. However, the EFA results indicated that all the four items converged to one distinct factor, which we labeled Judgment of Self and which was dropped as described above.

The latent factor correlations in the final four-factor model were relatively high, ranging from 0.56 to 0.82, hence a second-order factor model was specified in Mplus to investigate if a general construct exists. The chi-square difference test (=0.65) between the first order four-factor model and the second order factor model did not show significant difference, providing evidence of retaining the second-order factor model. A graph representation of the second-order factor model can be seen in Figure 1.

Confirmatory factor analysis confirmed four factors for the final scale with excellent model fit indices (Chi-square = 355.95, df = 221, RMSEA = 0.052, CFI = 0.95, TLI = 0.94, SRMR = 0.048); similarly, the second-order factor model showed favorable model fit indices as well (Chi-square = 356.60, df = 223, RMSEA = 0.051, CFI = 0.95, TLI = 0.94, SRMR = 0.048). All 23 items loaded as we originally conceptualized, with all loadings greater than 0.60 as presented in Table 3. The final four-factor model with 23 items has excellent overall internal consistency reliability (alpha = 0.95), each factor's consistency reliability ranging from 0.81 to 0.91, each factor's composite reliability ranging from 0.82 to 0.92 (Table 3), indicating excellent internal and composite reliability. This final scale had a mean score of 108.63 ± 16.20 out of a possible 138, with an observed range from 59 to 138. Item-scale correlations ranged from 0.52 to 0.82.

Construct validity was demonstrated by both EFA and CFA, both first order four-factor model and second order factor model showed a favorable model fit and meaningful factor loadings. The second order factor model provided support to the underlying latent construct which we termed "manifestation of consciousness" as a general construct. The subscales were highly correlated with the total scale score ranging from 0.73 to 0.89, but moderately correlated with each other, indicating good construct validity. The average variance extracted (AVE) for each factor was greater than 0.5 (Table 3), except for the first factor, Relationship to Others, which had an AVE of 0.47, indicating good convergent validity overall. Although the AVE of this specific factor is just below 0.5, this factor's composite reliability is above 0.70; hence we deemed this factor acceptable (Fornell and Larcker, 1981). Moreover, the square root of AVE for all factors is greater than their correlations with other factors, indicating good discriminant validity.

Known-groups validity

The known-groups validity of the final scale was examined by comparing the scale scores by demographic groups (Table 4). In particular, participants' years of meditation were found to be significantly correlated with all subscale scores and total score ($p < 0.05$) (Figure 2). Furthermore, participants with more years of meditation scored higher on all subscales scores and the total score. Males and females did not show significant differences on the total score and three of the subscale scores. However, females scored higher in the factor Connection with Higher Self than males ($p = 0.008$). Additionally, age was not correlated with two subscales, and had small correlations (Rs < = 0.15, $p < 0.05$) with subscale of LH and CHS as well as the total score. Participants who practice Heartfulness meditation scored significantly higher than other meditation and no meditation on the total scale and subscale scores, with one exception of subscale Relationship to Others. Finally, participants from Asian descents were found to score higher in the total scale score (Asian = 110.61 vs. non-Asian = 105.73, $p = 0.01$) and the subscales of Listening to the Heart (p = 0.01) and Acceptance and Letting Go ($p = 0.005$). Finally, we used a regression model with age, sex, years of meditation, Heartfulness vs. other, and Asian vs. non-Asian as predictors to predict total scale score. The results indicated that only years of meditation and practicing Heartfulness mediation were significantly associated with the total scale score. Whether respondents were of Asian descent no longer significantly associated with the total score.

Test–retest reliability

Thirty-six participants took the survey twice within a two-week period (average time period between the test and retest was 13.7 days). The test–retest reliability (i.e., intraclass correlation coefficient) for the total score was 0.90, and subscale scores test–retest reliability ranged from 0.71 to 0.92.

Discussion

This manuscript describes the development and testing of a scale, known as the Awareness Atlas, which measures the manifestation of consciousness in daily life. The final model fit exceeded all the required thresholds, indicating an excellent fitted model with a single

TABLE 3 CFA results with four factor model (23 items).

	Final item number	Loadings	Alpha	Test–retest reliability	Composite reliability	AVE
Relationship to Others (5 items)		**0.691#**	0.814	0.713	0.817	0.473
1. I am receptive to the feelings, needs, and suffering of others.	RO1	0.767				
2. I notice how others react to me at the time of an interaction.	RO2	0.663				
3. I notice my reactions to others at the time of an interaction.	RO3	0.656				
4. I consider the feelings, needs, and suffering of others.	RO4	0.749				
27. I consider the perspectives of others and learn from them.	RO5	0.626				
Listening to the Heart (5 items)		**0.937#**	0.919	0.911	0.918	0.694
10. I listen to the wisdom of my heart (the wisdom that arises when my)	LH1	0.823				
Preconceived notions, desires, judgment, and emotions are silenced and trust what it tells me.						
11. I feel guided in life, by the wisdom of my heart.	LH2	0.872				
12. When making decisions and interacting with others, it is easy for me to connect with the wisdom of my heart.	LH3	0.856				
13. To make decisions in any situation, my heart (inner wisdom) guides me from a place beyond emotion and thought.	LH4	0.869				
14. I trust my intuition.	LH5	0.713				
Connection with Higher Self (6 items)		**0.821#**	0.890	0.920	0.885	0.568
15. I feel supported by a deeper reality, underlying all of creation.	CHS1	0.841				
16. I feel that I am part of something greater than myself.	CHS2	0.741				
17. I feel a spiritual aspect to my identity, beyond my worldly identity.	CHS3	0.765				
18. I feel that my consciousness is ever expanding.	CHS4	0.764				
19. I have a feeling of wonder and awe about life.	CHS5	0.609				
20. I have a sense of being one with all beings in the universe.	CHS6	0.744				
Acceptance and Letting Go (7 items)		**0.878#**	0.897	0.762	0.898	0.562
21. I cheerfully embrace situations that are hard, uncomfortable, or challenging.	ALG1	0.776				
22. I cheerfully adapt to life circumstances in order to grow.	ALG2	0.817				
23. I embrace all experiences of my life with joy as they unfold.	ALG3	0.838				
24. I accept the struggles and lessons in life.	ALG4	0.620				
25. I use my self-awareness to realize I have choices in how to respond to situations.	ALG5	0.674				
26. My emotions, feelings, and thoughts remain balanced (stable) no matter what is going on within and around me.	ALG6	0.746				
28. As my awareness and consciousness change, I adapt my behaviors in order to be compatible with these changes.	ALG7	0.708				
Overall scale			0.950	0.900	0.900	0.700

Bold indicates the second-order factor loadings. AVE, Average variance extracted.

dimensionality to measure the manifestation of consciousness with four domains: Relationship to Others (RO), Listening to the Heart (LH), Connection with Higher Self (CHS), and Acceptance and Letting Go (ALG). Additionally, the final scale explained more than 50% of the total variance and all the factor loadings were greater than the recommended threshold of 0.4 (Boateng et al., 2018). Together, our findings indicate a measurement model for everyday aspects of manifestation of consciousness.

This manuscript also offers new insights into the current discussion of consciousness. We considered a wide range of definitions drawing from Eastern and Western perspectives, philosophical and modern scientific perspectives, experts and everyday people who experience shifts in consciousness in daily life. We then used this background to support us in developing an operational definition with an aim to develop a simple, comprehensive, and globally-relevant tool.

10.3389/fpsyg.2024.1283980

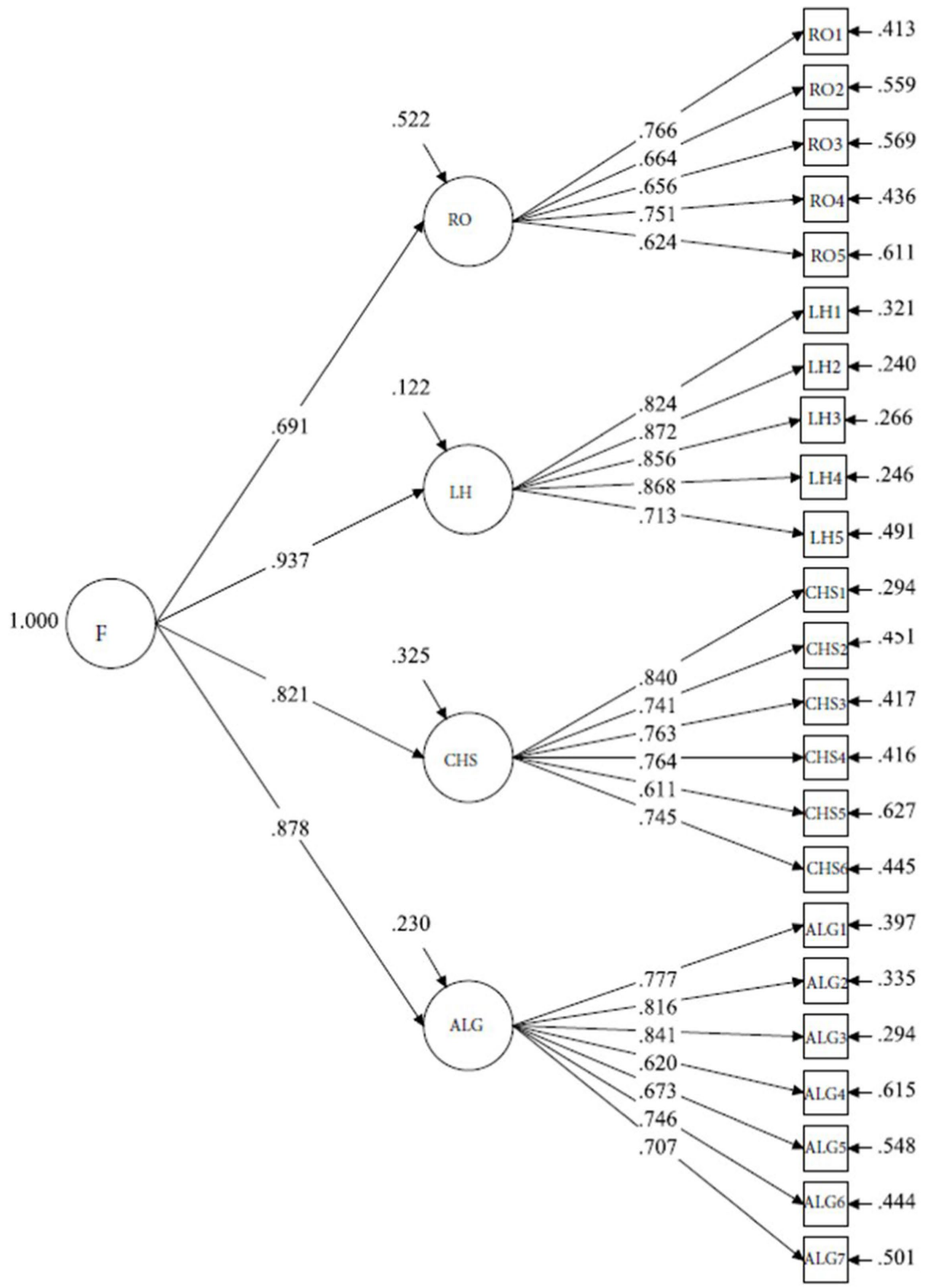

FIGURE 1
Graph representation of the CFA of the second order factor model. The second order factor CFA model fit indices (Chi-square = 356.60, df = 223, RMSEA = 0.051, CFI = 0.95, TLI = 0.94, SRMR = 0.048) was excellent. All 23 items loaded as we originally conceptualized, with all loadings greater than 0.60. Two pairs of residual correlations (CHS1 & CHS2, ALG1 & ALG2) were added to improve the model fit.

Construction of the scale

To develop the scale, we began with concepts of consciousness from Heartfulness Meditation, a set of simple practices based on the philosophy underlying Raja Yoga meditation (Chandra, 1989; van't Westeinde and Patel, 2022; Patel, 2023). The scale emphasizes a connection with the heart. The questions included in the scale were informed by transcripts from participants in a phenomenological study of the lived experience of Heartfulness (Pearmain et al., 2023). The Awareness Atlas was developed from responses of participants in the preliminary phenomenological study. Further, the four factors were derived by statistical analysis from the data.

With the intention of covering as many aspects of lived experiences narrated by respondents as possible, the initial scale started off with an item pool of 48. The number was gradually reduced by following specific criteria at different steps of the factor analysis models, as

TABLE 4 Known-groups validity evidence with the combined sample ($n = 449$).

	N	RO	LH	CHS	ALG	Total score
Age	430	0.059	0.104*	0.154**	0.075	0.124*
Years Med	426	0.107*	0.293**	0.303**	0.272**	0.308**
Sex		M ± SD	M ± SD	M ± SD	M ± SD	M ± SD
Female	281	24.61 ± 3.20	23.49 ± 4.39	30.13 ± 5.34	31.01 ± 5.94	109.17 ± 16.00
Male	149	24.17 ± 3.21	23.17 ± 4.95	28.68 ± 5.44	31.40 ± 5.89	107.52 ± 16.62
		–	–	***p* = 0.008**	–	–
Years Med						
< = 5 Years	103	23.92 ± 3.20	21.20 ± 4.70	27.04 ± 6.59	28.57 ± 5.92	100.56 ± 16.78
> 5 Years	323	24.64 ± 3.17	24.13 ± 4.22	30.50 ± 4.59	32.00 ± 5.65	111.26 ± 14.90
		***p* = 0.047**	***p* < 0.001**	***p* < 0.001**	***p* < 0.001**	***p* < 0.001**
0–10 Years	145	24.12 ± 3.21	21.65 ± 4.73	27.71 ± 6.43	28.97 ± 5.78	102.36 ± 16.63
11–20 Years	120	24.15 ± 3.14	23.65 ± 4.30	29.65 ± 4.86	31.55 ± 5.57	109.00 ± 15.42
20+ Years	161	25.01 ± 3.14	24.85 ± 3.92	31.43 ± 3.84	32.86 ± 5.63	114.16 ± 13.78
		***p* = 0.021**	***p* < 0.001**	***p* < 0.001**	***p* < 0.001**	***p* < 0.001**
Ethnicity						
Asian	194	24.61 ± 3.18	24.03 ± 4.49	29.96 ± 5.29	32.01 ± 6.01	110.61 ± 16.46
Non-Asian	101	24.57 ± 2.75	22.56 ± 4.74	28.71 ± 6.28	30.01 ± 5.25	105.73 ± 14.92
		–	***p* = 0.010**	–	***p* = 0.005**	***p* = 0.013**
Heartfulness						
Heartfulness	338	24.64 ± 2.91	24.09 ± 4.38	30.46 ± 4.79	31.82 ± 5.86	111.00 ± 15.53
Other	94	23.97 ± 3.20	24.70 ± 4.61	26.51 ± 6.45	28.80 ± 5.21	99.80 ± 15.07
		–	***p* < 0.001**	***p* < 0.001**	***p* < 0.001**	***p* < 0.001**
All	449	24.40 ± 3.20	23.23 ± 4.69	29.41 ± 5.57	31.03 ± 5.93	108.06 ± 16.46

*$p < 0.05$, **$p < 0.01$. RO, Relationship to Others; LH, Listening to the Heart; CHS, Connection with Higher Self; ALG, Acceptance and Letting Go. – indicates $p > 0.05$.

detailed in the results section. The final scale with 23 items was comprehensively representative of our definition of manifestation of consciousness, which is evident from the final factor structure, factor loadings (all >0.6), and total variance explained (>50%) (Figure 1). The emerging four domains of the new Awareness Atlas represent distinct areas of contribution to the overall manifestation of consciousness. These domains are Relationship to Others (5 items), Listening to the Heart (5 items), Connection with Higher Self (6 items), and Acceptance and Letting Go (7 items). Each of these content areas has counterparts in the focus of scales developed by others, e.g., the Feelings, Reactions, Beliefs Summary (Cartwright and Mori, 1988); Ego Grasping Orientation (Knoblauch and Falconer, 1986); Self-Expansiveness level Form (Friedman, 1983), however, no scale that we found encompassed all areas in a single scale.

Psychometric properties

By following all the necessary steps of scale development (Boateng et al., 2018), the newly developed scale demonstrated excellent validity and reliability in measuring our definition of the manifestation of consciousness. Specifically, in the confirmatory factor analysis testing dimensionality, the final second-order factor model fit exceeded all the required thresholds, indicating an excellent fitted model to measure a general construct—the manifestation of consciousness—with four domains.

Although all four factors demostrated acceptable to excellent test–retest reliability, it is noteworthy that Factor 1 (Relationships to Others) had a relatively lower test–retest reliability (0.71) and higher residuals for the items than the other factors. Nevertheless, the reliability and loadings exceeded the recommended thresholds. This finding, which should be further examined in future studies, might reflect the day-to-day fluctuation in a person's relationship to others. It is possible that life stresses or a particularly relaxed period of time might have influenced how the respondents rated their relationships to others more than other factors, which might be more stable over time.

The Awareness Atlas compares favorably in its psychometric properties with existing tools. For example, the Intellectual Humility Scale (Krumrei-Mancuso and Rouse, 2016) had comparable psychometric properties (internal consistency coefficient $\alpha = 0.82$ to 0.88, with subscales at 0.70 to 0.89 across samples); the MPE-92 (Gamma and Metzinger, 2021) reported internal reliability coefficients for 12 factors between 0.52 and 0.82; the State Mindfulness Scale (Tanay and Bernstein, 2013) with sound internal consistency where the total score coefficient $\alpha = 0.95$ with two subscales 0.90 and 0.95 but small to moderate test–retest reliability, with control condition ranging 0.46 to 0.59 and intervention condition ranging 0.22 to 0.68 with a two-week period. In contrast, both internal consistency reliability and test–retest reliability of our final scale favorably meet the criteria for scale development, as the overall scale reliability was greater than 0.90, and all subscale reliability scores were above 0.70. Similarly, the CFA fit

10.3389/fpsyg.2024.1283980

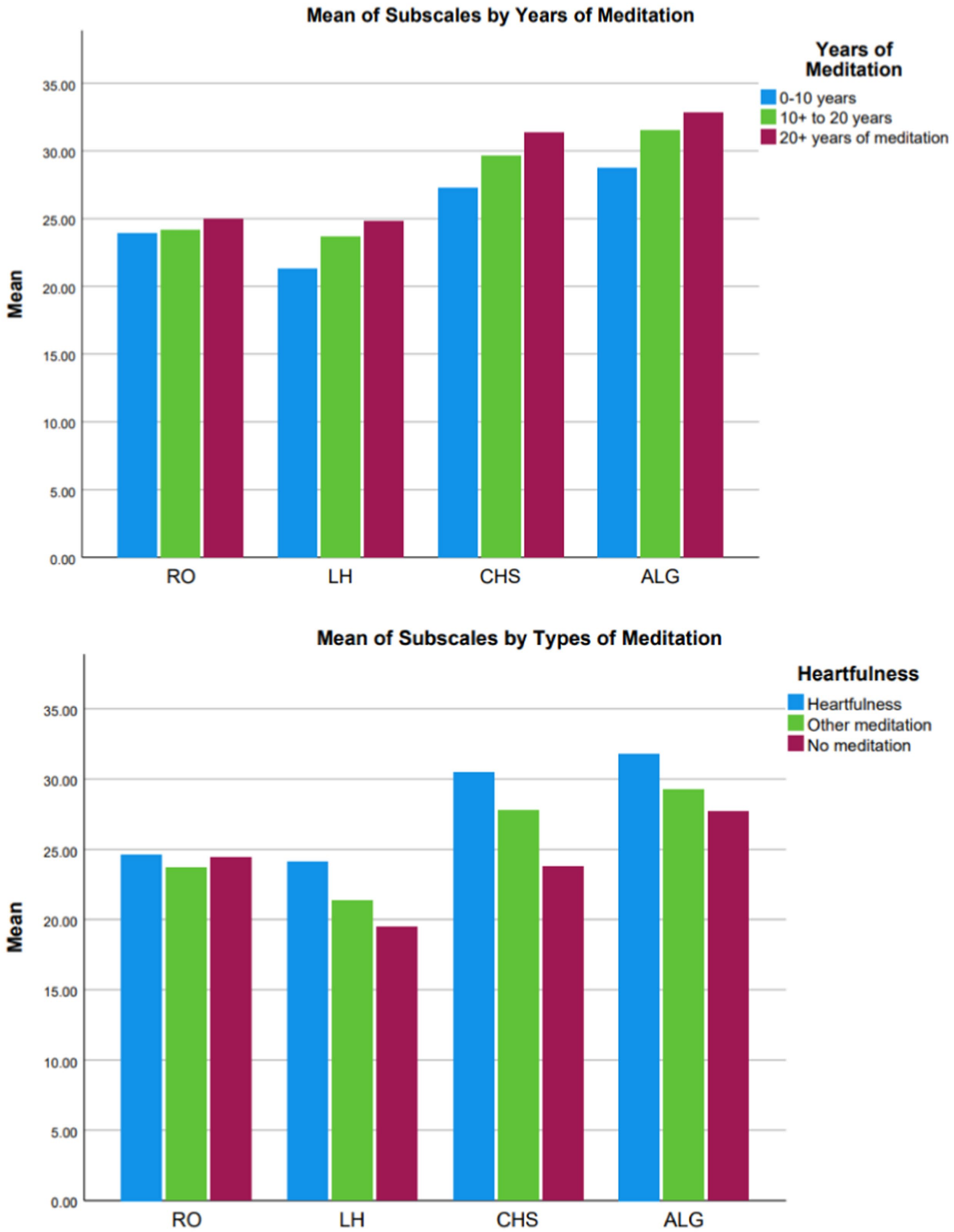

FIGURE 2
Comparison subscale means by years of meditation and types of meditation. RO, Relationship with Others; LH, Listening to the Heart; CHS, Connection with Higher Self; ALG, Acceptance and Letting Go. The subscales of LH, CHS, ALG showed significantly different among <= 10, 10–20, and > 20 years of meditation, while only 20+ years of meditators showed significant higher scores on subscale RO than <=10 and 10–20 years, and no difference was found between <=10 and 10–20 years of meditation. Similarly, heartfulness meditators showed significantly higher scores than other meditators and other meditators showed significantly higher score than no meditation on LH, CHS, ALG score, while no difference was found on RO score.

indices (CFI & TLI > =0.94, SRMR & RMSEA around 0.05) of our final scale were excellently comparable to the abovementioned scales, while the existing related scales with acceptable fit where CFI & TLI around 0.90 and RMSEA & SRMR around 0.08.

Of note, the Feeling Consciousness Scale (FCS), developed by Lindhard (2017, 2018a), focuses exclusively on the third level of consciousness, i.e., feeling-mind, one of the six levels outlined in Arka's theory (Lindhard, 2017). In their study, the FCS was employed to assess pre and post training effects of intuitive meditation conducted over a 6-week period. The author has acknowledged the need for further refinement and validation of the scale, due to a small sample size ($n = 31$) and limited evidence of validity and reliability.

We also examined known-groups validity by examining how the scale scores related to important demographics subgroups. Age and sex were not related to the total scale score, while participants who had more years of meditation were found to have significantly higher scores for all subscale scores and the total score than those who has fewer years of meditation, indicating good construct validity. Of note, those practicing Heartfulness meditation scored significantly higher than other meditation and no meditation groups. Given that the scale was constructed using Heartfulness meditation (a practice focused on heart-based awareness) as a basis, these findings add further support for known-groups validity of the Awareness Atlas.

In summary, our findings indicated a well-constructed instrument for measuring the manifestation of consciousness. The final scale has very strong fit indices and psychometric properties, consists of a short number of items, yet is sufficient to measure the desired construct. This new scale differs from existing scales related to consciousness (Mylonas et al., 2012; Gamma and Metzinger, 2021), in that it focuses on how consciousness manifests in relation to the experience of daily life. This new scale positions us well for studies examining how consciousness manifests in individuals whether or not they have experience with meditation.

Factors affecting scale scores

We conducted exploratory analyses of the data in this initial study to further understand the scale and to guide future research. Given that meditation practices originated in the East and are generally well-accepted by people of Asian descent, we examined scores of people of Asian descent compared to those who were not of Asian descent. We also examined the relationship between years of meditation and both total and subscale scores. Participants who had a longer history of meditation scored higher on the scale scores, those of Asian descent scored higher than those of non-Asian descent on the total score, and participants who practiced Heartfulness meditation scored significantly higher than other meditation and no mediation groups. Given the difference in numbers of participants in each of these groups, and the fact that the participants in "other meditations" and "no meditation" were recruited by Heartfulness practitioners, we acknowledge the need for a future study with more balanced sample sizes in each group. When we examined a number of factors in a multivariate analysis (i.e., age, sex, years of meditation, Heartfulness vs. other, Asian vs. non-Asian) to predict the total scale score, we found that only years of meditation generally, and Heartfulness meditation specifically, significantly predicted the total scale score. It is noteworthy that age was not a significant contributor in the regression analysis, indicating that it was not a factor of importance in this sample. These findings lend support to the good validity of the scale, suggesting it can be used with people having a wide range of backgrounds and ages.

An outcome of Heartfulness meditation is expanded consciousness (van't Westeinde and Patel, 2022). The finding that those who meditated for more years had higher scores than those who meditated for fewer years (Table 4; Figure 2) suggests that meditation does indeed alter manifestation of consciousness. Further, future longitudinal studies are needed to determine the extent to which evolution occurs with respect to feeling, wisdom and intuition, and humility with a meditation practice focused on expanding consciousness. Furthermore, future studies are needed to determine whether specific aspects of expanding consciousness are affected differently depending on the nature and purpose of the person's meditation practice.

Findings from this study demonstrate that the Awareness Atlas provides a tool that can further explore how reliance on the heart impacts a person's inner sense and feelings. Specifically, findings from the previous phenomenological study (Pearmain et al., 2023) suggest that participants found a sense of security, a groundedness within themselves, and a feeling of being held within something larger than self. They also identified a heart space that was very important to them. Those participants described improved relationships with themselves and with others. Taken together, these findings suggest there may be a growing capacity to relate with oneself and with others as a feeling of security increases. This is an interesting observation when viewed from the perspective of the innate need to feel that one has a place of belonging within and in relationship to others (Ainsworth, 1963; Bowlby, 1973). While we acknowledge the limitations in the current study, we believe that the Awareness Atlas can be used to explore such relationships in future studies including further exploration of the concept of a "heart space" to fully appreciate this new dimension to the understanding of health, wellness, and consciousness. Such investigations should explore the relationship between the heart space, the feeling of security, and people's ability to relate and manage life circumstances differently (e.g., relationship with self and others, the ability to navigate through difficult circumstances). Combined use of the Awareness Atlas with a phenomenological approach may help to elucidate new and important areas of investigation. Additionally, it will be important to further examine whether this scale has similar performance characteristics in other heart-based practices. Such studies could lead to a new understanding of the role of heart awareness and expansion of consciousness beyond conscious cognitive awareness.

Possibly, when one feels more grounded and secure within oneself it is easier to be more open and receptive to others, and to better navigate life's difficulties, including health challenges. A body of literature has examined the role of religion and spirituality in health with findings suggesting a positive association between meditation, health behaviors, and subjective well-being (e.g., Jayanna, 2023). Evidence is still needed regarding causality (Bożek et al., 2020; Ransome, 2020). George et al. (2000) considered possible reasons why spirituality and religion might contribute to health, including the encouragement of health promotion by religions, as well as social support that comes with religious and spiritual practices. These authors suggested that a third possibility was most compelling to them, namely the possibility that religion provides a sense of coherence and meaning, so that people develop a better understanding of their place in the universe, their purpose in life, and are better able to endure difficulties when they occur. Ransome (2020) has called for new tools that would be appropriate to epidemiological studies to further understand these relationships in a broader context. The Awareness Atlas provides such an option.

The Awareness Atlas is also applicable to investigations beyond religious and spiritual practices. For example, this scale can be used to explore how consciousness relates to the emotional impact of acute and chronic conditions. It is worth exploring whether reliance on the wisdom of the heart and having an expanded consciousness contributes to why certain people manage disease and aging more easily than others. Further, if expanded consciousness is positively associated with successfully navigating chronic conditions, there may be opportunities to assist people living with such conditions to improve their quality of life through approaches that expand consciousness. These and related issues are worthy of scientific exploration, utilizing both quantitative and qualitative approaches.

There is a growing awareness of the importance of heart rate variability (HRV) in creating emotional stability and a sense of inner calm (McCraty and Zayas, 2014; McCraty et al., 2017). Specifically, heart rate variability provides a measure of physiological coherence. Social coherence has been described as the organization or regulation of groups of individuals that are connected. It has been proposed that individual coherence (measured through HRV) and collective heart rate coherence can be used to increase group coherence (McCraty et al., 2017).

Finally, the operational definition used in this manuscript, "state of awareness of self, others, and the connection among all beings" encompasses the idea that consciousness is a dynamic state of mind, akin to awareness, which expands as one grows and evolves. Various schools of yoga and meditation aim to expand this awareness so that individuals begin to think more broadly of other members in the society, other life forms, the planet, and the universe. This idea appears to be more relevant today when societies are faced with challenges such as wars and terrorism, self-centered economy and development, growing inequities between regions, global warming, etc. Expansion of consciousness appears critical to improving these difficult societal and personal problems.

There are limitations to this study. Participants enrolled in the study were self-selected, which may introduce a potential selection bias. Most of the participants had practiced some form of meditation and many practiced Heartfulness meditation specifically. Additionally, those who meditated in other practices or did not have a meditation practice were recruited through friends and family members of Heartfulness meditators. We are aware of the biases inherent in this approach and will use an alternate recruitment method in further studies.

In this first investigation, we explored only a few factors for their contributions to responses (e.g., years of meditation, sex). A robust examination of potential contributors to scores on the Awareness Atlas should be undertaken in the future including a more comprehensive examination of the role of geographic location and the role of ethnicity in score responses. In this first investigation we did not examine concurrent validity by assessing the association of response on the newly developed scale with other existing scales. This should be evaluated in future studies to further establish the validity of the scale.

Additionally, the final scale does not include reverse scored items. Although the four reverse scored items were designed to belong to different themes at the early stage of the scale, the EFA results indicated that all the four items formed one distinct factor which we labeled 'Judgment of Self.' With an iterative comparison of the model fit, the final scale dropped all the reversed items resulting in a four-factor model. Future investigations should examine why the reverse scored items performed very differently than their counterparts of positively scored items.

It should be noted that the EFA sample did not include European participants, whereas the CFA incorporated participants from Europe. Despite this distinction, the EFA and CFA samples exhibited similarities in sex, age, educational level, and average years of meditation practices. Further validation among culturally diverse groups via measurement invariance will be the next step for future validations.

Despite the limitations above, our study managed to attract a diverse range of participants from various continents. This diversity and the size of the overall sample enhances the generalizability of the findings.

Conclusion

The Awareness Atlas was developed to measure the manifestation of consciousness through the lens of lived experience. This new measure quantifies manifestation of consciousness across a spectrum from simple awareness of self and others to an awareness of that which is beyond self. The Awareness Atlas has strong psychometric properties. It positions scientists to explore questions about changes in the manifestation of consciousness with practices designed to expand awareness (e.g., yoga, meditation) and with life experiences that affect awareness (e.g., life-altering conditions), as well as in response of communities to global challenges (e.g., climate change, wars, famines). Finally, the newly developed Awareness Atlas also can be used by individuals for self-reflection.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Institutional review board approval was obtained from the University of Colorado for worldwide distribution with the exception of the European Union (EU). A separate informed consent was approved and obtained from the University of Colorado for EU participants. All participants gave informed consent in order to access the survey.

Author contributions

YJ: Data curation, Formal analysis, Methodology, Project administration, Validation, Writing – original draft, Writing – review & editing. MS: Conceptualization, Data curation, Investigation, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. HC: Conceptualization, Investigation, Writing – review & editing. KJ: Conceptualization, Investigation, Writing – review & editing. RP: Conceptualization, Investigation, Writing – review & editing. AV: Writing – review & editing, Conceptualization, Investigation. KP: Writing – review & editing, Conceptualization, Investigation.

Funding

The author(s) declare that financial support was received for the publication of this article. Rutgers University paid the publication fee.

Acknowledgments

We thank Elizabeth Denley, PhD for her thoughtful recommendations in the final stages of this manuscript and her assistance with the final editing. We thank the 12 colleagues each of whom reviewed each statement and gave invaluable input during the early stage of scale development.

Conflict of interest

KP is the global guide of the Heartfulness Institute.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1283980/full#supplementary-material

References

Ainsworth, M. D. S. (1963). "The development of infant-mother interaction among the Ganda" in *Determinants of Infant Behavior*. ed. B. M. Foss (London: Methuen), 57–112.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., and Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front. Public Health* 6:149. doi: 10.3389/fpubh.2018.00149

Bowlby, J. (1973). *Attachment and Loss. Vol 11. Separation: Anxiety and Anger*. New York: Basic Books.

Bożek, A., Nowak, P. F., and Blukacz, M. (2020). The relationship between spirituality, health-related behavior, and psychological well-being. *Front. Psychol.* 11:1997. doi: 10.3389/fpsyg.2020.01997

Cartwright, D., and Mori, C. (1988). Scales for assessing aspects of the person. *Person Cent. Rev.* 3, 176–194.

Carver, C. S., and Glass, D. C. (1976). The Self-Consciousness Scale: A Discriminant Validity Study. *J. Pers. Assess.* 40, 169–172. doi: 10.1207/s15327752jpa4002_8

Chandra, R. (1989). *Complete Works of Ram Chandra*, Volume *1*. Toronto, CA: SRCM.

Deshpande, P.Y. (2021). *The Authentic Yoga—Yoga Sutras of Patanjali*. *2nd*. Pragati Offset Pvt Ltd: Hyderabad, India.

Diemer, M. A., Rapa, L. J., Park, C. J., and Perry, J. C. (2017). Development and Validation of the Critical Consciousness Scale. *Youth Soc.* 49, 461–483. doi: 10.1177/0044118X14538289

Fornell, C. D., and Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J. Mark. Res.* 18, 39–50. doi: 10.1177/002224378101800104

Friedman, H. L. (1983). The self-expansive level form: A conceptualization and measurement of a transpersonal construct. *J. Transpers. Psychol.* 15, 37–50.

Gamma, A., and Metzinger, T. (2021). The Minimal Phenomenal Experience Questionnaire (MPE-92M): Towards a phenomenological profile of "pure awareness" experiences in meditators. *PLoS One* 16:e0253694. doi: 10.1371/journal.pone.0253694

Garcia-Castro, J., and Kodukula, P. (2022). Why we need a whole new approach into the study of consciousness. *J. Conscious. Exp. Res.* 13, 397–414.

George, L. K., Larson, D. B., Koenig, H. G., and McCullough, M. E. (2000). Spirituality and health: What we know, what we need to know. *J. Soc. Clin. Psychol.* 19, 102–116. doi: 10.1521/jscp.2000.19.1.102

Grant, A. M., Franklin, J., and Langford, P. (2002). The self-reflection and insight scale: a new measure of private self-consciousness. *Soc. Behav. Personal. Int. J.* 30, 821–835. doi: 10.2224/sbp.2002.30.8.821

Griffiths, R. R., Richards, W. A., McCann, U., and Jesse, R. (2006). Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology* 187, 268–283. doi: 10.1007/s00213-006-0457-5

Guertin, P. A. (2019). A novel concept introducing the idea of continuously changing levels of consciousness. *J. Conscious. Explor. Res.* 10, 406–412.

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118

Jayanna, K. (2021). *Meditation and Expanded Consciousness: Implications for Sustainable Health and Wellbeing*. Consciousness and Meditation: Benefits, Applications and Research: Bangalore, India.

Jayanna, K. (2023). Integrative approach to lifestyle management: implications for public health research & practice in the context of SDG-3. *J. Ayurveda Integr. Med.* 14:100796. doi: 10.1016/j.jaim.2023.100796

Kegan, R. (1982). *The Evolving Self*. Harvard University Press, Cambridge, MA.

Kihlstrom, J. F. (2022). Consciousness, the unconscious, and the self. *Psychol. Conscious. Theory Res. Pract.* 9, 78–92. doi: 10.1037/cns0000285

Kirsmayer, L. J. (2010). Peace, conflict and reconciliations: contributions of cultural psychiatry. *Transcult. Psychiatry* 47, 5–19. doi: 10.1177/1363461510362037

Kitson, A., Chirico, A., Gaggioli, A., and Riecke, B. E. (2020). A review on research and evaluation methods for investigating self-transcendence. *Front. Psychol.* 11:547687. doi: 10.3389/fpsyg.2020.547687

Knoblauch, D. L., and Falconer, J. A. (1986). The relationship of a measured Taoist orientation to Western personality dimensions. *J. Transpers. Psychol.* 18, 73–83.

Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012

Krumrei-Mancuso, E. J., and Rouse, S. V. (2016). The development and validation of the comprehensive intellectual humility scale. *J. Pers. Assess.* 98, 209–221. doi: 10.1080/00223891.2015.1068174

Lindhard, T. (2017). Experiencing peace through heart-based meditation on the Self. *Open Psychol. J.* 10, 27–40. doi: 10.2174/1874350101710010027

Lindhard, T. (2018a). The Theory of Six Main Levels of Consciousness: A Study of the Third Level. *J. Conscious. Explorat. Res.* 9, 40–61.

Lindhard, T. (2018b). Levels of Consciousness: The Role of the Heart and Pulsation. *Int. J. Soc. Work Hum. Serv. Pract.* 6, 65–74.

Loevinger, J. (1976). *Ego Development*. Jssey-Bass, San Francisco, CA.

MacDonald, D. A., and Friedman, H. L. (2013). "Quantitative Assessment of Transpersonal and Spiritual Constructs" in *The Wiley Handbook of Transpersonal Psychology*. eds. H. L. Friedman and G. Hartelius. *1st* ed (Chichester: John Wiley and Sons, Ltd)

McCraty, R., Atkinson, M., Stolc, V., Alabdulgader, A. A., Vainoras, A., and Ragulskis, M. (2017). Synchronization of human autonomic nervous system rhythms with geomagnetic activity in human subjects. *Int. J. Environ. Res. Public Health* 14:770. doi: 10.3390/ijerph14070770

McCraty, R., and Zayas, M. A. (2014). Cardiac coherence, self-regulation, autonomic stability, and psychosocial well-being. *Front. Psychol.* 5:1090. doi: 10.3389/fpsyg.2014.01090

Mylonas, K., Veligekas, P., Gari, A., and Kontaxopoulou, D. (2012). Development and psychometric properties of the scale for self-consciousness assessment. *Psychol. Rep.* 111, 233–252. doi: 10.2466/08.02.07.PR0.111.4.233-252

Nasibulina, A. (2015). Education for sustainable development and environmental ethics. *Procedia Soc. Behav. Sci.* 214, 1077–1082. doi: 10.1016/j.sbspro.2015.11.708

Patel, K.D. (2019). *Designing destiny: heartfulness practices to find your purpose and fulfill your potential*. Hay House, Inc., Carlsbad.

Patel, K.D. (2023). *Spiritual Anatomy: Meditation, Chakras and the Journey to the Center*, Balance, Hachette Book Group. New York.

Patel, K. D., and Pollock, J. (2018). *The Heartfulness Way* New Harbinger Publications, Inc.

Pearmain, R., O'Connor, H., Schenkman, M., Jayanna, J., and Patel, K. D. (2023). An exploration of the lived experience of a modern heart-focused meditation practice. *J. Transpers. Psychol.* 54, 182–198.

Portney, L.G. (Ed.) (2020). *Foundations of Clinical Research: Applications to Evidence-Based Practice, 4e*. Philadelphia: FA Davis.

Raffone, A., and Srinivasan, N. (2010). The exploration of meditation in the neuroscience of attention and consciousness. *Cogn. Process.* 11, 1–7. doi: 10.1007/s10339-009-0354-z

Ransome, Y. (2020). Religion, spirituality, and health: new considerations for epidemiology. *Am. J. Epidemiol.* 189, 755–758. doi: 10.1093/aje/kwaa022

Sattin, D., Magnani, F. G., Bartesaghi, L., Caputo, M., Fittipaldo, A. V., Cacciatore, M., et al. (2021). Theoretical Models of Consciousness: A Scoping Review. *Brain Sci.* 11:535. doi: 10.3390/brainsci11050535

Savelyeva, T., and Douglas, W. (2017). Global consciousness and pillars of sustainable development: A study on self-perceptions of the first-year university students. *Int. J. Sustain. High. Educ.* 18, 218–241. doi: 10.1108/IJSHE-04-2016-0063

Sparby, T. (2015). Investigating the depths of consciousness through meditation. *Mind Matter* 13, 213–240.

Srinivasan, N. (2022). Consciousness without content: a look at evidence and prospects. *Front. Psychol.* 11:1992. doi: 10.3389/fpsyg.2020.01992

Tanay, G., and Bernstein, A. (2013). State Mindfulness Scale (SMS): development and initial validation. *Psychol. Assess.* 25, 1286–1299. doi: 10.1037/a0034044

Teasdale, G., and Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale. *Lancet* 304, 81–84. doi: 10.1016/S0140-6736(74)91639-0

Thiengkamol, N. (2011). Development of a model of environmental education inspiration of public consciousness influencing to global warming alleviation. *Eur. J. Soc. Sci.* 25, 506–514.

Thomas, C., Holz, T., and Bhatia, S. (2022). Heartfulness Meditation with Yogic Transmission: Creating the spiritual environment for relaxation, wellness and the evolution of consciousness. *J. Transpers. Psychol.* 54, 168–181.

Tripathi, V., and Bharadwaj, P. (2021). Neuroscience of the yogic theory of consciousness. *Neurosci. Conscious.* 7, 1–15. doi: 10.1093/nc/niab030

Turner, R. G., Carver, C. S., Scheier, M. F., and Ickes, W. (1978). Correlates of Self-Consciousness. *J. Pers. Assess.* 42, 285–289. doi: 10.1207/s15327752jpa4203_10

van't Westeinde, A., and Patel, K. D. (2022). Heartfulness meditation: a yogic and neuroscientific perspective. *Front. Psychol.* 13:806131. doi: 10.3389/fpsyg.2022.806131

Vieten, C., Amorok, T., and Schlitz, M. M. (2006). I to We: the role of consciousness transformation in compassion and altruism. *J. Relig. Sci.* 41, 915–931. doi: 10.1111/j.1467-9744.2006.00788.x

Wilber, K. (2000). *Integral Psychology: Consciousness, Spirit, Psychology, Therapy*. Shambhala. Boston, MA.

Wilber, K., Engler, J., and Brown, D. (1986). *Transformations of Consciousness: Conventional and contemplative perspectives on development*. Shambhala, Boston, MA.

frontiers | Frontiers in Psychology

TYPE Hypothesis and Theory
PUBLISHED 08 April 2024
DOI 10.3389/fpsyg.2024.1344033

Check for updates

OPEN ACCESS

EDITED BY
Antonino Raffone,
Sapienza University of Rome, Italy

REVIEWED BY
Giorgio Marchetti,
Mind, Consciousness and Language Research Center, Italy
William Wong,
Monash University, Australia

*CORRESPONDENCE
Bruno Forti
✉ brunoforti4@gmail.com

RECEIVED 24 November 2023
ACCEPTED 26 March 2024
PUBLISHED 08 April 2024

CITATION
Forti B (2024) The hidden structure of consciousness.
Front. Psychol. 15:1344033.
doi: 10.3389/fpsyg.2024.1344033

The hidden structure of consciousness

Bruno Forti*

Department of Mental Health, Azienda ULSS 1 Dolomiti, Belluno, Italy

According to Loorits, if we want consciousness to be explained in terms of natural sciences, we should be able to analyze its seemingly non-structural aspects, like qualia, in structural terms. However, the studies conducted over the last three decades do not seem to be able to bridge the explanatory gap between physical phenomena and phenomenal experience. One possible way to bridge the explanatory gap is to seek the structure of consciousness within consciousness itself, through a phenomenal analysis of the qualitative aspects of experience. First, this analysis leads us to identify the explanandum concerning the simplest forms of experience not in qualia but in the unitary set of qualities found in early vision. Second, it leads us to hypothesize that consciousness is also made up of non-apparent parts, and that there exists a hidden structure of consciousness. This structure, corresponding to a simple early visual experience, is constituted by a Hierarchy of Spatial Belongings nested within each other. Each individual Spatial Belonging is formed by a primary content and a primary space. The primary content can be traced in the perceptibility of the contents we can distinguish in the phenomenal field. The primary space is responsible for the perceptibility of the content and is not perceptible in itself. However, the phenomenon I refer to as *subtraction of visibility* allows us to characterize it as *phenomenally negative*. The hierarchical relationships between Spatial Belongings can ensure the qualitative nature of components of perceptual organization, such as object, background, and detail. The hidden structure of consciousness presents aspects that are decidedly counterintuitive compared to our idea of phenomenal experience. However, on the one hand, the Hierarchy of Spatial Belongings can explain the qualities of early vision and their appearance as a unitary whole, while on the other hand, it might be more easily explicable in terms of brain organization. In other words, the hidden structure of consciousness can be considered a bridge structure which, placing itself at an intermediate level between experience and physical properties, can contribute to bridging the explanatory gap.

KEYWORDS

explanatory gap, explanandum, hidden conscious structure, phenomenal analysis, early vision, multiple hierarchical segregation, hierarchy of spatial belongings

Introduction

According to Loorits (2014), if we want consciousness to be explained in terms of natural sciences, we should be able to analyze its seemingly non-structural aspects, like qualia, in structural terms. During the last three decades, numerous authors have sought to identify the structure of Phenomenal Consciousness (PC) in classic neuronal organization (Crick and Koch, 1998; Dehaene et al., 1998; Tononi and Koch, 2008; Jerath and Crawford, 2014; Calabrò et al., 2015; Koch et al., 2016; Boly et al., 2017; Gallotto et al., 2017; Polák and Marvan, 2018;

Noel et al., 2019; Maillé and Lynn, 2020; Seth and Bayne, 2022), in the activity of electromagnetic fields (McFadden, 2020, 2023; Ward and Guevara, 2022; Jones and Hunt, 2023), or in quantum physics (Hameroff and Penrose, 2014; Tuszynski, 2020). However, all these studies do not seem to be able to bridge the explanatory gap (Levine, 1983) between physical phenomena and phenomenal experience (Marius, 2014; Skokowski, 2022; Jones and Hunt, 2023; Sanfey, 2023).

Many theories have addressed non-specific aspects of consciousness, such as access-consciousness (Block, 1995, 2005; Baars, 2002; Tyler, 2020), meta-representation (Gennaro, 2004; Brown et al., 2019), global access (Dehaene et al., 1998; Dehaene, 2014), unity (Bayne, 2010), integration, (Tononi, 2008; Tononi and Koch, 2015; Brogaard et al., 2021; Hirschhorn et al., 2021), intentionality (Crane, 2003, 2009), selection (Zeman, 2001; Schwarzkopf and Rees, 2015). In the absence of specific features of consciousness, there is a risk of formulating a theory that refers to something that is compatible with the absence of consciousness. The specific characteristics of consciousness can be attributed to its phenomenal aspect, which are usually traced back to qualia (Dennett, 1988; Searle, 1997) and what it is like to be in a certain state (Nagel, 1974).

I think that correlating phenomenal experience with certain aspects of neuronal processes – even discovering the proper level of organization of the neural activity (Revonsuo, 2006) – is not enough to bridge the explanatory gap and thus solve the hard problem. In my opinion, the nature of the brain structure is such that it cannot explain – at least directly – experience. All the data we have thus far – and probably also those we might have – seem to indicate that a brain in the broad computational sense is unable to account for experience (Toribio, 1993). Phenomenal experience and brain structure are too different or "distant" to be directly compatible. This difference is probably the basis for the very conception of the explanatory gap and the formulation of the hard problem. On the contrary, structural aspects of consciousness can be found in phenomenal experience. Consequently, a possible alternative is to look for the structure of seemingly non-structural aspects of consciousness (Loorits, 2014) not in the neuronal substrate, but in consciousness itself, through a phenomenal analysis of the qualitative aspects of experience that starts from its simplest forms.

An essential premise is to reformulate the explanandum of consciousness. In fact, qualia do not have a phenomenal existence as isolated entities. Furthermore, the qualitative aspects usually analyzed in the literature - such as the redness of red or the painfulness of pain – must be placed in a more complex structural context than is commonly believed. The simplest qualitative aspects – such as those related to being an object, background or detail - can be found in early vision. They are involved in perceptual organization and necessarily have relational significance. Such phenomenal qualities, which are manifold and different from each other, are perceived in relation to each other and seem to form a unitary whole. We can say that the explanandum of consciousness is a unitary set of qualities, i.e., a set of qualities closely dependent on each other, which we can find in its simplest forms in early vision (Forti, 2024). Of course, unity *per se* is not specific to consciousness. However, in this case unity concerns specific properties of consciousness such as the qualitative aspects. The co-presence of the qualitative aspect and the unity aspect is thus crucial in identifying the explanandum of consciousness.

Although early vision is characterized by interdependent qualitative components that form a unitary whole, we cannot find in it the structure of seemingly non-structural aspects of consciousness. Phenomenal appearance alone does not seem sufficient to identify a unitary structure of consciousness. However, the closeness of these characteristics to a unitary structure prompts us to delve into a less explored territory, using the components of experience also as possible explanans. I hypothesize that the structure of consciousness can be found in consciousness itself on the basis of the possibility that the aspects we attribute to Phenomenal Consciousness (PC), in addition to being explananda – whereby we wonder how subjective experience, made up of qualia, sensations and feelings, emerges or is produced by brain activity – may contribute to an explanation of consciousness itself.

A not insignificant consequence of considering the phenomenal aspects of consciousness *only* as explananda is that, in almost all theoretical approaches, the analysis of these aspects is inadequate. When one merely explains the non-specific aspects of consciousness, one does not perform a phenomenal analysis at all. The phenomenal aspects are simply pushed aside or ignored. In other cases, the phenomenal analysis is very sketchy, limited to only a few elementary aspects like the redness of red and the painfulness of pain. From this point of view, the need to restrict and simplify as much as possible what we have to explain is understandable. However, these approaches lead to ignoring the relational aspects of so-called qualia and to underestimating the richness of the internal structure of experience, even in its simplest forms, and thus to an unrealistic view of the experience that one wants to explain. Phenomenologists have highlighted this issue well: "we will not get very far in giving a scientific account of the relationship between consciousness and the brain unless we have a clear conception of what it is that we are trying to relate. To put it another way, any assessment of the possibility of reducing consciousness to neuronal structures and any appraisal of whether a naturalization of consciousness is possible will require a detailed analysis and description of the experiential aspects of consciousness" (Gallagher and Zahavi, 2008). My approach goes beyond the understandable need to better define the explanandum. The possibility that the phenomenal aspects of consciousness may also be useful elements in identifying an explanation prompts us to analyze them carefully and in detail, taking an interest even in secondary or seemingly insignificant phenomenal aspects.

As I will explain in the next sections, I postulate the existence of *non-apparent* parts of experience and hypothesize that consciousness possesses a *hidden* structure, one that comprises both apparent and non-apparent constituents. I call it the Hierarchy of Spatial Belongings (HSB). This structure can explain the unity of early visual experience and its main qualitative aspects, i.e., its being a unitary set of qualities. At the same time, it better lends itself to being correlated with certain physical processes, helping to bridge the gap between experience and brain processes.

The reasons for taking this hypothesis into consideration arise from the analysis of generally neglected phenomenal aspects such as surroundedness and overlapping of the contents of the field. Another element that suggests the possibility that in consciousness we can find elements that can help explain consciousness itself is the problem of appearance. I hypothesize that appearance depends on something which could be responsible for making it appear, but which would not have in itself the property of appearing. Therefore, what appears would be a clue to the existence of what does not appear. This something could belong to the region surrounding the object, to which we attribute a phenomenal nature of background or space.

The multiple hierarchical segregation of the perceptual field

The mechanism underlying a hidden structure of consciousness can be identified in the model of Multiple Hierarchical Segregation (MHS) of the perceptual field, which I presented in detail in another article (Forti, 2015). MHS is an alternative model of perceptual organization to the Gestalt model. A limit of Gestalt theory is the lack of a comprehensive view of perceptual organization. This can be seen in the distinction between figure-ground segregation and grouping, as well as in the proliferation of principles of grouping (Wagemans et al., 2012). The traditional view is that, once the object is identified (Pinna, 2012), grouping takes place among the components of the field which have an object nature through a heterogeneous set of field organization principles. I have proposed a mode of organization in which the spaces of the field play an active role. This model provides a simpler explanation than the traditional principles and is compatible with a unitary structure of the visual field.

The conditions under which we see a simple figure can be derived from the nature of the figure and of the ground. The perception of a black triangle is conditioned by the brightness of the background. The bigger the object-foreground difference is, the more vivid our conscious experience will be. Therefore, these conditions imply the division of a field into two homogeneous regions, one internal to the other and the two contrasting with each other (Todorović, 2008; Wagemans et al., 2012). One may say that the conditions under which we see a simple shape on a homogeneous background consist in the *Surrounding Contrast (SC)* of the structure of the proximal stimulus. All else being equal, figures in which the SC of the proximal stimulus structure is strongest will tend to visually dominate the others. The strongest contrast of the proximal stimulus is the one in which there is the greatest difference in the response of the receptors to two concentric regions of the stimulus field.

According to the MHS model, there is a correlation between the SC gradient of the structure of the proximal stimulus and the progressive segregation of the perceptual field. There is a SC when a spatially extended region of the proximal stimulus contrasts or is inhomogeneous with the whole surrounding region. These conditions can occur, albeit with some differences, in several perceptual modalities. Since all relationships involve the field of the stimulus in its entirety, we have to imagine that several SCs of the structure of the stimulus are overlapping in a complex way.

The fact that all relationships involve the field of the stimulus in its entirety does not occur at a phenomenal level. In fact, it seems to occur only for the main object. As can be seen in Figure 1, the strongest SC corresponding to the black triangle not only causes it to phenomenally prevail over the other elements, but it also brings about a subdivision of the field into two asymmetric areas which we perceive as figure and ground. Unlike the main segregation, the other segregations which derive from the smaller SCs do not seem to affect the field in its entirety, but the areas which formed as a result of the first segregation. We see the gray triangle *in the region which acts as a background* to the black triangle, i.e., inside a space which does not include the whole framed area, but only the white space surrounding the black triangle, and we see the small white circle *inside* the black triangle.

Similarly, a face segregates from the background and it is in its turn affected by a process of segregation. This process does not affect the whole field, but only the object, i.e., the face which acts as a "background" to the eyes, nose and mouth. The pair of eyes segregates from the face; in its turn, each eye segregates from the region occupied by the pair of eyes. This process appears as the most appropriate explanation of what occurs when we perceive objects such as a house or a tree, which are internally complex and which are perceived in a context which is in its turn internally complex.

FIGURE 1
A simple example of multiple hierarchical segregation.

FIGURE 2
A broken line.

As in the pair of eyes, the progressive segregation of the field occurs not only when an object is located inside another. The term SC refers to a region located inside the field, without necessarily identifying it with a continuously contoured figure, and it can correspond to the grouping of several objects into a gestalt. A broken line (Figure 2) perceptually appears as prevalent because, despite the discontinuity of the parts which form it, the line corresponds to the strongest SC of the structure of the stimulus. The SC corresponding to the individual dashes is smaller than the SC of the line due to the presence of the other dashes in the external space, while the individual dashes have more or less an equivalent SC. As a consequence, the region corresponding to the line is secondarily subdivided into the four dashes and we see a broken line, i.e., *a line made up of dashes*. Like the white circle belongs to the black triangle, the dashes belong to the line – and not to the whole image. Each dash, despite not prevailing over the others, is seen "against the background" of a region which includes the other dashes. In general, we can say that the simple rule stating that, given a visual field, the perceived object corresponds to the strongest SC of the proximal stimulus accounts for several aspects of perceptual organization, thus unifying Gestalt laws. My previous article (Forti, 2015) provides a detailed description of the phenomena that are usually explained on the basis of grouping principles.

In short, there is no grouping of the perceived dashes on the basis of their similarity and proximity, as stated in Gestalt laws. Instead, there is a progressive segregation of the structure of the stimulus, i.e., a process of MHS. MHS is correlated with the SC gradient of the proximal stimulus. The term progressive should not be interpreted in a temporal sense, but in a hierarchical sense. The segregation which determines the perception of the line is hierarchically superior, while the segregations which determine the perception of the dashes are subordinate to it.

It is evident that the perceptual situations selected by gestaltists for their analyses favor the possibility of "seeing" groupings of elements instead of the progressive segregation internal to the field, as is the case when observing the most common perceptual situations. However, the progressive segregation internal to the field can be "seen" also in the perceptual situations analyzed by gestaltists. In this perspective, what gestaltists call grouping by proximity and by similarity can be considered a sort of atypical MHS. A broken line is a sort of incomplete segregation, because it includes both what we attribute to matter and what we attribute to space. Nonetheless, this region tends to segregate anyway in the presence of a sufficient SC of the stimulus. Segregation is atypical in that an incomplete segregation such as the one of the broken line, arising from a stronger SC of the stimulus, prevails over the complete segregations of the individual dashes.

According to Searle (2004), there are two aspects to the Gestalt structure of consciousness: (1) the capacity of the brain to organize perceptions into coherent wholes; (2) the capacity of the brain to discriminate figures from backgrounds. Similarly, Wagemans et al. (2012) state that "perceptual grouping and figure-ground organization, although intimately connected, are not the same process." If instead we think of the phenomenal field as a hierarchy of relationships which form following the progressive segregation of the field, these two aspects can be unified.

The MHS model seems consistent with both the possibility of a single mode of field organization and the need to account for the progressively less significant aspects of phenomenal experience. The main relationship concerns the whole field, the less important relationships concern the parts that formed as a result of the first subdivision, and so on. Unity seems to derive not so much from the existence of elements of homogeneity and coherence in the perceptual Gestalt as from the internal subdivision of the conscious field. As a result of the SC gradient of the proximal stimulus, the visual field is gradually segmented within itself, and each subdivision appears to be strictly dependent on the others. This seems to account for the unity of the conscious structure as well.

However, the progressive subdivision of the field within itself highlights a shortcoming of the MHS model. It is the fact that, while appearing homogeneous, several parts of the field would at the same time be composite. The object is contained within a background but, at the same time, it contains details within itself. The background contains the object within itself but, at the same time, it is contained within an additional background. The problem affects most regions of the field, even under the simplest perceptual conditions. The white space inside the box in Figure 1 can have four perceptual properties: it acts as the background of the black triangle, it is part of a larger background, it acts as the background of the gray triangle, and finally – being a box – it is an "object" seen against the background of the external space. If object and background are to some extent composite, what are they composed of? Since we do not see a background superimposed on a figure or, respectively, a figure superimposed on a background, how can we reconcile the background role of the space inside the box with its role as an "object" seen against the background of the space outside the box?

Other authors have also highlighted this problem. According to Peterson and Salvagio (2010), a region can be a ground along some portion of its bounding edges, and a figure along other portions. Even though the white background in Figure 1 is unshaped near the border it shares with the smaller black region, it is shaped by the outline border it shares with the larger surrounding white region. But the subdivision of this region into two juxtaposed parts seems artificial, especially in the case of small backgrounds. Moreover, it is not compatible with further subdivisions of the field. Another way to deal with this problem is to assume that we see these different aspects of a field region at later times (Searle, 2004). In fact, we do not necessarily separate – at least sharply – a region into a part that we liken to an object and a part that we liken to a background, nor do we see the different parts of an image one after the other. This means that the internal organization of the field involves the simultaneous presence of values whose nature we struggle to understand. Is the MHS hypothesis therefore wrong in that it is phenomenally untenable?

The hierarchy of spatial belongings

The MHS model solves the problem of the unity of conscious structure. However, it poses the problem of the composite nature of many regions of the field that we consider homogeneous. A relatively simple solution is that the only difference between the outcomes of the individual segregations is their hierarchical value. The question can be posed in the following way: if we expect a multiplicity of relations which hierarchically overlap each other as a result of the progressive segregation of the field and make us see how we see what we see, what is the nature of the outcome of each segregation?

I propose that the single segregation of the field would not lead to the formation of figure and ground, so we cannot speak of figure-ground segregation. I call Spatial Belonging (SB) the "simple" relationship, a kind of proto-image, produced by each segregation. I use this definition because belonging to a space is a *sine qua non* for any content to be conscious, for it to be perceived. A SB consists of a primary content and of a primary space. They are content and space in the absolute sense of the term if they do not overlap with other contents or spaces.

What are the properties of the two concentric regions of the SB? I propose that the primary space has the property of allowing the primary content to appear, or to be perceived, and that this occurs through a relationship between contrasting outer and inner regions. This means that a content cannot be perceived unless it is surrounded by a primary space and that this space, while making it appear, is not perceivable. All spatial belongings are characterized by these properties. The difference is that Spatial Belongins arising from a stronger SC prevail over the others and contain them within themselves. It should be pointed out that we do not experience content perceptibility at the level of the individual SB, which we cannot access, but at the phenomenal level. While primary content and primary space are at a level we can call sub-phenomenal, our perceptual experience is made up of overlapping Spatial Belongings nested within each other. Since primary space is not perceptible, it remains "hidden" from our experience. However, it is not phenomenologically inert. I will address this issue in the section "Appearance."

Are primary content and primary space consistent with our experience? Can we reconstruct the phenomenal level of early perception from Spatial Belongings? Spatial Belongings are the building blocks with which early vision is constructed in its qualitative and structural aspects. According to the MHS model, the total field segregates into two concentric regions forming the main SB. In the presence of inhomogeneity, each of the two regions thus formed segregates within itself in turn, resulting in smaller Spatial Belongings. Further segregations result in progressively smaller and less phenomenally relevant Spatial Belongings, until fading. MHS causes different Spatial Belongings to largely overlap with each other. For example, the primary content of the main SB in Figure 1, which includes the area of the black triangle together with the white circle, overlaps with the SB which includes the area of the black triangle as primary space and the area of the white circle as primary content. We could say that most of the field regions we perceive, including the seemingly simplest ones, are at least both primary content and primary perceptual space as a result of the – multiple – overlapping of Spatial Belongings in a hierarchical structure which I call the Hierarchy of Spatial Belongings (HSB). Essentially, the Multiple Segregation of the perceptual field determines the individual Spatial Belongings and their hierarchical organization, accounting for the phenomenal nature of early perception. What we see is the effect of the relationship between overlapping regions nested within each other. The HSB corresponds to early vision but, unlike the latter, it cannot be experienced as such because of the presence of hidden components in it (Figure 3).

Primary contents ensure the perceptibility of what we can call phenomenal contents. Of the many terms used to describe what we perceive, the term 'content' appears to be the most generic. Etymologically, content is necessarily inside something. Phenomenally, and thus not as primary content, a content is anything that we can perceive in a phenomenal field and that appears sufficiently separate from other parts of the field. It is the thing on the basis of which we can make a phenomenal distinction. Despite their differences, all regions of the field have this property in common. From this point of view, a background is also a content.

Phenomenal contents coexist in the perceptual field, not only juxtaposed next each other, but also overlapping each other. In Figure 1, the black triangle, the white circle, and the object consisting of the black triangle with the white circle on the inside are all contents. Both the broken line and the dashes that make it up are contents. If a scented-red-rose is a content of our experience, so are its shape, its scent and its red color, since, while belonging to the overall gestalt of the rose, they are sufficiently separate that they can be perceived as contents. Similarly, the word 'rose' is a content as are the individual letters that make it up.

Also the phenomenal nature of the object and of the background derives from the overlapping of the content and space components. But what differentiates the object from the background if both are formed by the overlapping of primary content and space? My proposition is that we perceive a region of the field as background when the role of the space to phenomenally define prevails over the role of the content to be defined. In the main background, the predominance of the space component is due to the fact that it is part of the main SB as a result of its stronger SC. The background has, at the very least, a dual nature: of content, on the basis of which it can be perceived; and of space, which makes us perceive the object it defines. The background is perceived through an external space, but its predominant component is to make us perceive the content it bounds, so it is perceived as *empty space surrounding the figure*.

If we keep in mind that a visual object differs from a simple shape because of its constitutive inhomogeneity, the space component is also present in the perception of the object. We perceive an object when the role of the content to be defined prevails over the role of the space to define. This is why we do not see the black portion of the main figure in Figure 1 as a background. However, a share of the attention we pay to the black triangle is subtracted from it to focus on the white circle it contains. The difference between background and object is that the component of phenomenally defining prevails in the former, while it is of lesser importance in the latter. This means that in the object the perception of the component attributable to the primary perceptual space is more difficult than in the background, although it is unquestionably present. It corresponds to the phenomenal datum whereby we see the white circle on the inside as a detail.

It should also be noted that object and background, in their content and space components respectively, represent the outcome of the main segregation. The intertwining of content and space does not account only for the phenomenal characteristics of object and background. Because of the additional subdivisions that occur within it, the external space defined by the main segregation is more complex than the generic notion of background might suggest. The secondary object, i.e., the grey triangle, is both content and part of the

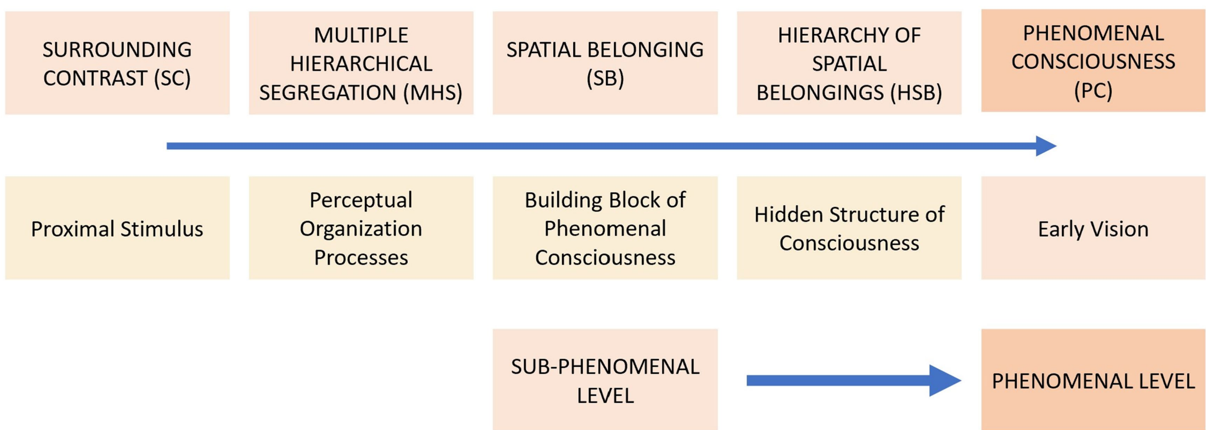

FIGURE 3
Starting from the SC of the proximal stimulus, the processes of MHS produce the hierarchical organization of SBs. This hidden conscious structure (HSB) corresponds to the phenomenal level of early vision (PC).

background, so it is seen against the background of the main object and at the same time is part of it. Moreover, the background of the main object, especially in the immediate surroundings, tends to converge on the secondary object.

Also to understand the nature of a detail it is necessary to take into account the relationships between the Spatial Belongings of the perceptual field. The Spatial Belonging to which the black triangle belongs as content together with the white circle prevails over the Spatial Belonging overlapping with this content, which includes the black triangle as space and the white circle as content. The SB between the black triangle together with the white circle and the overall surrounding space is the main one as a result of the stronger SC of the proximal stimulus. This is the reason why the black triangle is perceived as an object and not as a background and the white circle is perceived as a detail of the triangle and not as a phenomenal object in the full sense of the term. It is both content and part of a larger content that prevails over the former. Because of the limitations of the paper, I will limit myself to analyzing these features of early vision.

The variability of the relationships involved also accounts for all the intermediate situations and varying degrees of prevalence of content over space or vice versa, including the gradual way in which we move from object to background. If we look at the pen lying on a book, the book is more than a background. It may become more important and even become an object within which we recognize the detail of the pen.

Finally, the idea of a HSB appears compatible with the progressive fading of the phenomenal field. This neither means perceiving the entire field equally, nor making experience coincide with focused consciousness alone. The notion of HSB implies that there is a progressive fading of perceptibility from the main content to the contents that are gradually subordinated to it. If we consider the field as a whole, the parts we perceive in relation to others are progressively *fading*, especially – but not only, if we think of change blindness (Noë et al., 2000) – toward the outside of the field. This gradualness is entirely compatible with the richness that characterizes all our phenomenal experience. It is true that in change blindness we cannot see the changes that affect certain parts of the field, so much so that some believe that this phenomenon would demonstrate that we see much less in the perceptual field than we think (Rensink, 2004; Scrivener et al., 2021). However, it is also true that change blindness is based on perceptual situations in which *dozens* of spatial belongings are formed. It is worth noting that in very simple stimulus conditions, as in many of those studied by gestaltists, we can sufficiently perceive all the relations in the field.

Of course, it is well known that consciousness is made up of parts that we can see well and of parts that we can see less well. However, MHS allows us to explain the gradualness of this phenomenon and the structural relationship between focus and fading. The perceptual field is characterized by multiple relationships of surroundedness, the Spatial Belongings, which gradually decline from what we perceive distinctly to what we perceive with increasing difficulty, until gradual disappearance from the phenomenal field. At the same time, the less significant relationships of surroundedness depend on the more significant ones and occur within the subdivisions of the field generated by the latter. In other words, the individual Spatial Belongings are nested to each other on the basis of a hierarchical organization. With its hidden components, the HSB is the structure underlying the perceived unity of the visual field, even in situations where the contents of a scene seem to be arranged randomly. Any element is part of the whole as an outcome of the progressive subdivision of the perceptual field.

Surroundedness

The hypothesis I put forward is based on the analysis of two generally neglected relationship modes present in the perceptual field. The first is the belonging of contents to a space, or surroundedness, and the second is the overlapping of the contents of the field.

One of the difficulties in understanding consciousness stems from the fact that the background and the fringe aspects are underestimated, as James (1890) pointed out with his metaphor of the pails in the river. However, this metaphor should be applied not only to the flowing water of a river, but also to the still water of a lake, because in this regard the important relationships are spatial as well as temporal.

Despite the fact that the relationship between foreground and background has often been included among the properties of consciousness (James, 1890; Zeman, 2001; Edelman, 2003; Searle, 2004; Northoff et al., 2023), many approaches have tended to make consciousness coincide with contents (Schulte, 2023) and focused consciousness, neglecting unfocused aspects. Of course, I do not intend to claim that the background is different from what it appears. The background is phenomenally less significant than the content. The problem is that, taking its phenomenal significance as a starting point, it is considered at best an ancillary element, which accompanies the content. Some authors have considered the background, along with fringe aspects, as degenerate information, or even as something that deceives and misleads us (Dennett, 1991, 2005, 2015; Noë and O'Reagan., 2000; Rensink, 2004; Prinz, 2018). The fact that the background is phenomenally less important does not mean that its role is necessarily negligible. Of course, the background is relational by its very nature. It implies the existence of a relationship with a figure, an object, a foreground. Consequently, what is underestimated is the relationship between figure and background.

The relationship between figure and background concerns not only well-defined shapes. For a visual stimulus to be perceived, a fine-grained representation is not necessary. Let us think of the perception of an indistinct spot. The ability to perceive it depends more on the contrast between content and background than on its definition. The fact that a piece of writing is blurred to the point that it cannot be recognized does not prevent such content from being consciously perceived.

Moreover, the presence of the background is essential not only in vision. Smelling a smell or hearing a sound are considered elementary conscious experiences. But even these contents are invariably perceived against the background of something. Just as we see an indistinct spot against the background of the surrounding visual space, a sound is perceived against the background of the auditory space,[1] and we feel pain against the background of the leg. Coming

1 Of course, the background of a sound can also be temporal. For the sake of simplicity, here I limit myself to spatial backgrounds, both by analogy with vision and because spatial relationships do not require the involvement of memory.

from a region of the perceived space, many elementary sensations have phenomenal characteristics not unlike those of a blurred image. Even in cases where a sensation seems to occupy the entire visual field, such as when we close our eyes to experience darkness, we cannot help but experience our body. If we focus on the visual experience, our body will act as a background to the darkness we perceive and will in turn be perceived in the background of the perceptual space in which our body is located (Jerath et al., 2015). Perception is, *ab initio*, multisensory (Bennett and Hill, 2014; Bayne and Spence, 2015; O'Callaghan, 2015).

Another limitation of the classical approach to perceptual organization is that it almost exclusively analyzes the relationship between the figure and the background – in fact the main figure and the main background. First, there may be a number of backgrounds in a field. Second, it is better if we consider a relationship type like *surroundedness*. I define it as a relationship whereby a region is surrounded by or surrounds a contrasting region. It is a form of juxtaposition that occurs between two contiguous and concentric regions. Surroundedness has a broader meaning than the one we attach to the figure-background relationship. For example, it also applies to the relationship between object and detail or the relationship between primary space and primary content (SB).

The SB is the fundamental surroundedness relationship and it is indispensable for consciousness to exist. Being a relationship between a primary space and a primary content, it is a hidden relationship. It makes it possible to perceive the contents of the perceptual field and it is the basis of all phenomenal surroundedness relationships.

Phenomenal surroundedness does not concern only the background. We can see an object on a table, which has the nature of an object and is seen against the background of the floor. Moreover, in addition to *being surrounded*, the parts of the field can *surround* other parts, such as the dots on the faces of a dice, or such as the eyes and mouth of a person's face. In this case, the relationship is reversed, in the sense that the main element is the object. However, an object is also the "space" in which its details are situated. Surroundedness does not concern only the relationship between the object and its internal components. Even a gestalt can be considered a set of elements *contained within* the region it occupies. We conceive of a gestalt as a group of elements. But, in fact, we perceive a broken line as a salient region containing spaces that interrupt the continuity of the line.

It could be argued that first we see an object against a background that includes the other elements of the field, and that later we shift our attention to other aspects of the field. According to Searle (2004), "I see the pen against the background of the book, the book against the background of the desk, the desk against the background of the floor, and the floor against the rest of the room, until I reach the horizon of my entire perceptual field." Based on a widespread view, this description implies that, by gradually broadening our focus, we *first* see the pen against the background of the book, *then* the book against the background of the desk, and *then* the desk against the background of the floor. In fact, if we focus our attention on the book, we see the book – on which lies a pen – against the background of the table *and simultaneously* the table against the background of the floor. While it is true that we see less well as we move away from the object on which we focus our attention, we cannot even say that we see only what we focus on *and nothing else*. This would be not only a simplistic conception of our experience. Failure to perceive the secondary parts would alter or prevent the perception of the main content. This is

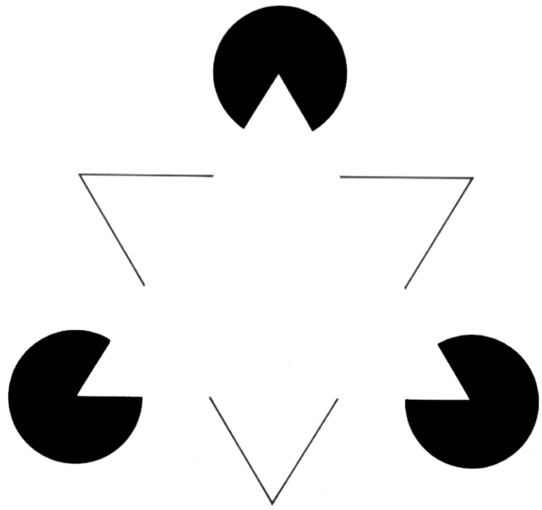

FIGURE 4
Kanizsa's triangle.

evident in illusory figures. We can only perceive an illusory triangle (Figure 4) if we superimpose it on three black disks and a white triangle with a black outline, which in their turn stand out against the surrounding white background. This phenomenon is even more evident in other much more common perceptual situations, such as the ones analyzed in Gestalt psychology. We cannot see a broken line without seeing – at the same time – the dashes that make it up.

Overlapping

Another type of relationship, which is also overlooked, involves overlapping between the components of the field.[2] While evident in some cases, phenomenal overlapping is hardly analyzed for its role in the phenomenal structure (Jerath et al., 2019). I think that overlapping in the conscious field is not only more frequent than we think, but also that it is a fundamental component of conscious organization, without which PC would not be possible.

Overlapping occurs in a number of circumstances, such as in the case of occlusions, when one object extends behind another and through amodal perception we tend to see its shape (Briscoe, 2011; Calabi, 2013; Nanay, 2018). Another well-known case of overlapping, i.e., the one of the background, which tends to perceptually extend behind the object, is even one of the basic properties of visual perception (Todorović, 2008). In fact, overlapping by occlusion is the only one that is taken seriously in the literature.

But phenomenal overlapping does not only occur when visual regions overlap with each other. The characteristics of reality in which we are interested are also auditory, tactile, olfactory and so on. Consequently, multisensory stimuli, such as observing a person talking to us or smelling the scent of a flower we admire while holding

2 In order not to generate confusion, I use the term overlapping also in cases in which the term superimposition might be more appropriate.

it in our hands, come to us simultaneously from the same region of space. This is reflected in our perceptions, which seem to be formed by overlapping sub-images. Overlapping also occurs in the case of different sub-modalities, for example, between form and color. With regard to the property of composition, Tononi and Koch (2015) mention the example of the perception of the blue book, but without explicitly addressing the structural function that the relationship between the two phenomenal distinctions might have at the phenomenal level.

Moreover, as we saw above, contents that can be referred to a part can overlap with contents that can be referred to the whole. Of course, overlapping can also occur with states originating internally, such as emotions, memories, thought processes or simulations (Smit et al., 2023), which I am not going to address in this paper. These few examples highlight a phenomenal reality that can hardly be disputed. However, overlapping is rarely considered a defining characteristic of phenomenal consciousness (Fingelkurts et al., 2009).

The problem of binding is to inquire how the brain binds together multiple multimodal characteristics into the unitary experience of the object, i.e., into what we conceive of as a single conscious "image" (Feldman, 2013). I am not going to delve into the role of the neuronal processes involved (Crick and Coch, 1990; Llinas et al., 1994; Revonsuo and Newman, 1999; Deroy et al., 2014; Walling, 2019), about which there is no consensus and no entirely satisfactory explanation (Isbister et al., 2018; Jerath and Beveridge, 2019; Kesserwani, 2020). But it is worth considering the possibility that binding may *also* occur at the conscious level and that it may occur through the overlapping of a number of sub-images.

The above leads us to think that overlapping is involved in the qualitative aspects usually attributable to qualia. A qualitative sensation is something that overlaps with a region of a perceived image – usually an object – characterizing it and modifying the experience (Jerath et al., 2019). By binding a certain characteristic to an object, overlapping is the way in which the unity of the object is achieved at the conscious level. Thus, the concept of overlapping allows us to place qualitative features in the context of the relations existing in the field and to assign them a structural role.

A qualitative sensation can overlap with an object, as in a yellow triangle or in the taste, pleasantness, color, and cold feeling of pistachio ice cream. Therefore, it is perceived through overlapping with an object which in turn belongs to a background, and thus through secondary or indirect belonging to the background. Or a quality may itself determine the extent of a certain region. In this case, it acquires an object-like function, thus defining a region which in turn belongs to a background.

One might counter that yellow is identified with a certain region. It is true that qualitative aspects define the characteristics of an object that has a certain form and that they can take on the form of the objects they overlap with, as in the case of color. However, qualitative aspects are independent of form, so quality – unlike an object – does not have a form of its own. In conscious perception a color may or may not extend like the object, or it may itself be the object. When yellow overlaps with a definite form such as a banana, we can tell if the yellow color has the same extent as the banana, so seeing a yellow object means seeing it uniformly yellow. Or we can tell if the extent of the yellow color is different from the banana, so we can detect streaks and their actual shape. In the case of a drawing, we can tell if the banana is colored well and, if not, we can identify the form of the color. In the case of a badly colored object, we will say that the extent of the yellow color does not perfectly match the object, but we will not attribute that form to its being yellow. We will attribute it to the region in which the color is located and which is bounded by the surrounding space through contrast. If we see that form on its own, like a spot of color, we will attribute it to the spot, not to the color itself. Even if a rainbow is made "only" of color, at a phenomenal level it characterizes itself as a *colored arc*.

In the case of a tactile or olfactory sensation, such as pain in a knee or a sound from a certain area in the external space, there are no sharp contours. The lack of a definite form seems to depend on the poor degree of definition of pain or sound. As sensations, they help create a kind of formless object, although more or less extended and located in space. In fact, even in this case pain delimits a region of the leg. It is the knee that hurts. We distinguish pain from the region of the leg that hurts, although the latter is defined by the pain itself. Similarly, a sound comes from a region of the visual landscape and presumably from something located in it.

The above analysis highlights that the qualitative aspects usually analyzed in the literature are placed in a more complex structural context than those related to being an object, background or detail. But overlapping occurs also in the simplest, seemingly homogeneous parts. Admitting the existence of multiple segregations poses the problem of overlapping not only between distinct regions of the field – such as between occluding and occluded object or between multimodal sub-images – but also within the same region of the field. MHS leads us to hypothesize that individual segregations result in a kind of proto-images, or SB, that overlap with each other.

HSB is based on the combined role of overlapping and surroundedness. Overlapping and surroundedness are constitutive of consciousness, even in the simplest forms of perceptual experience. What appear to us as juxtaposed components of the perceptual field are actually Spatial Belongings. Their spatial component is not apparent *per se* and they partially overlap with and are nested within each other.

Appearance

Paradoxically, one aspect that can help identify *non-apparent* parts of consciousness is precisely that relating to appearance. Like the relationships of surroundedness and overlapping, it is another fundamental yet neglected aspect. We usually consider it a priority to explain the qualitative aspects of consciousness, but its appearance is something even more fundamental and such that it underlies the qualitative aspects. Aspects related to appearance should be distinguished from strictly phenomenal aspects. According to Nagel (1974), a being is conscious just if there is "something that it is like" to be that creature, i.e., some subjective way the world seems or appears from the creature's mental or experiential point of view (Van Gulick, 2022). The problem lies not only in the way the world appears to us, in the effect the world has on us in its appearance, but also in the mere fact of appearing (Revonsuo, 2006; Whiting, 2016; Merlo, 2020).

In the simplest sense, appearance, which is nothing else than the etymological meaning of consciousness as a phenomenal entity, implies the possibility of something being perceived consciously. This can mean several things: being conscious rather than not being conscious; seeing rather than being blind, despite having other sensory experiences; distinguishing two neighboring points rather than not distinguishing them. In metaphorical terms, if on the inside of Chalmers' zombies all is dark because they have no experience, appearance is that thing that occurs when the "light" of consciousness comes on (Baars, 1997, 2005). This idea is often associated with something magical and inexplicable. As Thomas Huxley states: "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the djinn when Aladdin rubbed his lamp in the story." The notion of global access in the Global Neuronal Workspace theory (Dehaene, 2014; Mashour et al., 2020), linked to brain "ignition," is not that far from this conception and could be considered an updated version of the idea of "conscious light" ignition.

However, at least in a relative sense, appearance is something that is not evenly distributed throughout the conscious field, but it concerns some regions of the conscious field to a greater extent than others. Moreover, the latter seem somehow necessary for perception to occur. "A figure on a background … is the very definition of the phenomenon of perception, that without which a phenomenon cannot be said to be perception at all. The perceptual 'something' is always in the middle of something else, it always forms part of a 'field'" (Merleau-Ponty, 1945). From this point of view, one could say that, since the relationship between object and background involves the existence of contrast, the conditions for the emergence of an elementary form of phenomenal experience do not depend on the – metaphorical – coming on of the light of consciousness, but on the development of a certain kind of relationship between darkness and light. Light certainly illuminates an object in darkness, but darkness also makes light visible. Total darkness, as well as total light, caused by the absence of a contrast between the object and what surrounds it, cannot ever constitute the totality of consciousness, as suggested by the Ganzfeld effect (Schmidt et al., 2020).

According to James (1890), one of the main characteristics of consciousness is that "it is *always* interested more in one part of its object than in another, and it welcomes and rejects, that is, chooses, all the time it is thinking." This phenomenon is not necessarily related to attention (Pitts et al., 2018). It is not so in the case of the perception of a simple figure against the background of something (Kimchi, 2009). James states that "we find it quite impossible to disperse our attention impartially over a number of impressions." In other words, consciousness cannot help but function in this way, so this characteristic is constitutive of consciousness itself. But even before choosing between different contents, whichever way we want to conceive of them – objects, impressions or otherwise – we choose between content and container. Consequently, if we think of the perception of a simple figure, this characteristic of consciousness might imply the very possibility of perceiving.

As we have seen above, appearance derives from SB, a hidden surroundedness relationship. Primary space does not have the property of appearing, but it rather has the property of allowing the primary content to appear. However, as we will see later in the text, it can be traced in the region surrounding any conscious content, to which we often attribute a phenomenal nature of background or space (Forti, 2009). Of course, what is figure and what is background depends on the mutual arrangement of the field regions and it may change over time. According to this hypothesis, in bistable figures we see one figure at a time because the figure is seen thanks to the surrounding region, which thus cannot be seen at the same time. Naturally, bistable perception may depend on attention and neural oscillations (Doesburg et al., 2009; Dieter et al., 2016; Davidson et al., 2018; Zhu et al., 2022).

Is there any evidence for which we can say that the space surrounding an object "makes us see" the object, while remaining unseen, and that it can be traced in PC? Preliminarily, if we assume that "the perceptual 'something' is always in the middle of something else, it always forms part of a 'field'" (Merleau-Ponty, 1945), this cannot apply only to the object which is usually perceived as the main object. It must also apply to other objects and even to the background. To assume that whatever "thing" we see must be in the middle of something else is to assume that what surrounds the thing we see cannot be seen except in the presence of an additional "something else." Consequently, the outermost region of the perceptual field would not be visible. But we can make this argument not only starting from the center to reach the periphery, but also backward, from the periphery to the center. The regions that we see, like the main background, would *also* have the function of making us see and not of being seen.

The hypothesis that the role of the space surrounding the content is to allow perception is consistent with a phenomenal characteristic of the background. While the characteristics of figure and background are well known, it is not sufficiently emphasized that their phenomenal relationship is not one of mere contiguity or *co-occurrence* in the field. Object and ground are closely interdependent, not only because they are foreground and background, respectively. We know that the background is formless and that it is perceived as empty space (Kanizsa, 1980). This description neglects the fact that the background appears in relation to the object and seems to help give it form, pop-out and phenomenal "matter." We cannot simply say that the background is less salient than the object. A secondary object or detail is also less salient, but it does not have the same relationship that the background has with the main figure.

If we try to see the background by focusing our attention on it, it is difficult for us to do so, especially near the object, as we are led, somewhat "pushed," to see the object. Even when we strive to see it as an object, the background still tends to *make us see* the figure it bounds and to make it pop out perceptually. This also means that the background, especially near the object, is phenomenally characterized as a region from which visibility is subtracted. This is why I call this phenomenon *subtraction of visibility*. Naturally, I am referring to a partial subtraction of visibility. The "objective" datum, for which a contrasting surrounding space is necessary for the content to be perceived, is thus consistent with the subjective datum, since a phenomenal property of the background appears to be that of allowing the content to be perceived.

However, the subtraction of visibility is *perceptible* only if this space is bounded by an additional space. About the background, Kanizsa (1980) states that "from a perceptual point of view there are considerable functional differences between the region of the field that takes on the character of figure and the one which plays the role of background. The figure has an object character, it is a 'thing,' whereas for the background this character is much less marked, until it is *almost completely* absent when the background is experienced as empty space." The presence of a residual object character is thus essential to be able to speak of background as an empty space. One can speak of space *in the absolute sense of the term*, and not of background, when the object character is missing *altogether*.

In my view, the loss of the residual object character of a region of the phenomenal field that serves as the background results from the absence of an additional contrasting space bounding it. This loss implies that that space cannot be perceived – not even as empty space. If the space that bounds a content is not bounded by another space – i.e., if it is not also a content – it is not phenomenally defined and, as a result, it is not perceived. The wall surrounding the painting, that we perceive, is also a content. If it is the outermost phenomenal background, it is seen thanks to an additional external space which we cannot see.

When we speak of perceptibility, we usually refer to an *internal region* of the perceptual field, bounded by a contrasting surrounding region. In the periphery of the visual field, where there is a progressive decline of perceptibility and the space component becomes gradually predominant, our ability to define not only objects, but also the spaces to which they belong, progressively diminishes, with no possibility of defining the boundaries of the field. If we consider the outermost region of the perceptual field, we must assume that, while it is necessary to perceive the region it bounds, it cannot be perceived because it is not bounded by another space. This means that the outermost *background* of the visual field, which we barely see and which seems to fade into nothingness without being bounded by any region, is actually bounded by an additional external space which we cannot see, but which somehow makes us see it. It is worth specifying that the outermost background of the visual field is a phenomenal entity and that it is formed by the overlapping of a primary content and a primary space that prevails over the former. The outermost region of the field is not a phenomenal entity, because it is exclusively made up of an absolute space. The outermost region as an additional external space which we cannot see allows us to see the outermost phenomenal background and avoids the endless regression which would occur if we assumed that every background must be surrounded by another - phenomenal - background.

The existence of the subtraction of visibility implies that primary space, while unseen, is not phenomenologically inert. Just as we can say that a primary content – for the way it affects our experience – is phenomenologically positive, we could say that a primary space is *phenomenologically negative*. This characteristic impacts the background through a partial subtraction of visibility. As we have seen above, the phenomenal nature of the background derives from the overlapping of the content and space components when the latter is prevalent. Being phenomenally negative, primary space partially takes away visibility from the region perceived thanks to the surrounding space. This action produces the phenomenal quality typical of the background. Of course, this does not occur only in the background, but in it the phenomenon is more evident. This is a counterintuitive concept, if not contradictory to our idea of the phenomenal world. However, it is interesting to note that, drawing on Gestalt theories, the concept of negative space is used in graphic design and photography. In art and design, negative space is the empty space around and between the subjects of an image (Cave, 2013).

The above is an analysis of the qualitative aspects of early perception. In this paper I am not addressing qualities usually attributable to qualia. I am not explaining the redness of red, but I limit myself to stating that it is something less simple than the quality related to the perception of the phenomenal object. However, the characteristics of the object are no less qualitative than the redness of red and the painfulness of pain.

Discussion

An approach that has made significant contributions to the understanding of conscious perception is experimental phenomenology, i.e., the study of appearances in subjective awareness (Albertazzi, 2019, 2021; Albertazzi et al., 2021). It aims to uncover the principles of organization that guarantee (qualitative) invariants. These phenomena are explicable on the basis of the conditions of their appearance that the phenomenological analysis is able to demonstrate (Kanizsa, 1979). The fact that in the phenomenological experiments there is a manipulation also of physical stimuli is largely irrelevant because the description, manipulation, and demonstration are performed at the level of appearances only (Musatti, 1957). The kind of information that experimental phenomenology uses to perform suitable behavior in conscious perceiving is internally directly given in present awareness, qualitative in nature. My approach partly distances itself from experimental phenomenology, because MHS is correlated with the SC gradient of the proximal stimulus (Forti, 2015). In other words, I have adopted a psychophysical approach (Gescheider, 1997; Fetsch et al., 2013).

Unlike Gestalt laws, MHS guarantees the unity of perceptual organization. It explains not only how we define the main object, but also the relationships between the parts. By posing the problem of the composite nature of apparently homogeneous regions of the field, it is a necessary premise for explaining the phenomenal and qualitative nature of the different components of the perceptual field. This explanation is made possible by assuming the existence of a hidden conscious structure. The qualities of early vision result from the overlapping of appearing and making something appear and from the relationships of surroundedness between the regions that overlap. These relationships also entail a progressive segmentation of the field, which ensures the unity of perceptual experience.

The parts that cannot be perceived affect all aspects of our experience. The primary perceptual space is essential not only – in an absolute sense – to enable us to see, but also to make us see how we see what we see. Early visual experience corresponds to a Hierarchy of Spatial Belongings. This structure, though hidden from experience, appears consistent with the nature of field parts like object, background, and detail. It is also consistent with the way these parts tend to form a unified whole. In other terms, the existence of a HSB explains why in early perception field components have a certain quality and appear as a unitary set of interdependent components, i.e., the reason why the explanandum consists of a unitary set of phenomenal qualities. Several phenomenal qualities can be traced back to just two factors: (1) the relationship between primary content and primary space in Spatial Belonging, and (2) the existence, in the field of consciousness, of a Hierarchy of Spatial Belongings nested within each other. Even if I do not explain how appearance is defined by the relationship between primary content and primary space, this relationship allows us to provide a relatively parsimonious explanation to the different primary qualities and their interdependence (Schurger and Graziano, 2022).

Moreover, the existence of a hidden conscious structure leads us to change our conception of consciousness. The definition of consciousness cannot be based *only* on appearance or a part of it. Consciousness is not only what appears or what we are aware of. On the contrary, it is also made up of non-apparent or non-perceptible parts, in relation to which we cannot make any phenomenal distinction. We are not aware of all that is part of the conscious field, not only because of the existence of unfocused, fringe or progressively fading parts, i.e., as a result of limited capacity (Zeman, 2001). We are not aware because what appears requires something to make it appear, which in itself does not have the property of appearing, even if the phenomenally negative nature of primary space is somehow made manifest at the phenomenal level as subtraction of visibility. I am referring to fundamental components of the conscious field that are an integral part of its structure and that, being related to the apparent ones, are essential for consciousness to appear as such.

The different nature of content and space makes us understand why, although the hierarchical organization of spatial belongings is compatible with the phenomenal datum, we do not see consciousness in this way. This organization can be considered a kind of hidden architecture of the phenomenal field. The hidden structure of consciousness is explained on the basis of the need to make the content appear and not on the basis of the generic idea that a cognitive system is incapable of examining its own structure (Loorits, 2014).

According to some authors, consciousness is deceptive, either in whole or in part (Dennett, 1991, 2015; Noë et al., 2000). We seem to see more than we see. My analysis also leads to the conclusion that consciousness is, to some extent, deceptive. But my conclusion is quite the opposite, in that I claim that consciousness includes not only aspects which we perceive with difficulty, but also aspects that we cannot perceive.

A lot of authors do not even consider in its entirety what is sufficiently distinguishable. Definitions for which consciousness corresponds to a certain qualitative sensation, or to what it is like to be in a certain state, are based on *a part* of experience, leaving out the parts that are considered non-specific, precisely because they are structural. Other authors neglect the parts that cannot be clearly defined, and they theorize that they need to be eliminated from a scientifically acceptable conception of consciousness. In this way, they hope to simplify the object of investigation. But it is as if we wanted to define the cell by considering the nucleus or the phospholipid bilayer and ignoring everything else.

In my approach, consciousness is deceptive not because we think we perceive more than what we actually perceive. On the contrary, it is deceptive because we do not perceive parts that should be considered to all intents and purposes as belonging to the field and that play a major role in defining the phenomenal quality we perceive. At the same time, the existence of non-perceptible parts is entirely compatible with what we perceive, with our experience. In no way does my analysis lead to overturning the fundamental assumption that, in the case of consciousness, appearance is reality (Chalmers, 1995; Searle, 1997; Tononi and Koch, 2015; Whiting, 2016; Merlo, 2020), nor does it lead to questioning its existence (Dennett, 1991). However, consciousness is not *just* appearance.

The presence of non-apparent, even phenomenally negative components implies that consciousness is much more complex and internally structured than we think. The complexity of experience is usually underestimated by seeking it outside of experience. It is generally assumed that, for consciousness to arise, particularly complex processes occur within or outside the classical canons of neuronal architecture (Tononi and Edelman, 1998; Sarasso et al., 2021; Koculak and Wierzchoń, 2022; MacIver, 2022; Hunt and Jones, 2023), or even physics (Hameroff and Penrose, 2014; Zhi and Xiu, 2023), at the non-conscious level. In contrast, little is said about the complexity of a conscious image that we experience, except for the insights provided by the phenomenological approach (Kanizsa, 1979; Gallagher and Zahavi, 2008; Smith, 2018). However, we have seen that all qualities – even the simplest one, related to the object – result not only from relations with juxtaposed regions, but also from overlapping between different field regions.

The phenomenally negative nature of space makes it possible to perceive the spaces surrounding the objects. This results in an "aerial" structure of experience. Indeed, the experience is made up of material objects located in a space which, while appearing phenomenally incorporeal, is in fact part of the experience. It is worth emphasizing the adaptive value of this kind of structure, which somehow reproduces a world made up of objects and regions in space. The salience of the object corresponds to a consistency that testifies to the material nature of the object and contrasts with the incorporeality of the surrounding space. Despite the similarities, this hypothesis is different from that of Jerath and Beveridge (2019), according to which a subconscious, virtual, space–time matrix is the foundation of experience and continuously exists in the conscious mind as a coordinate system for a recreation or simulation of the material world.

To accept this hypothesis is to accept that from the very beginning PC is made up of components that are not only juxtaposed, but also overlapping. In some cases, the composite nature of our experience is quite clear, although its importance is rarely emphasized: some examples are the scent of a flower, the color of a triangle, the voice of a person, the name of an object. Dennett (1991) uses an example of learning to hear fine details of a guitar sound. Guitar sound can be decomposed into overtones, or constituent parts of the sound. In fact, due to overlapping, the structure of consciousness is composite, complex, and counterintuitive. Being phenomenally negative, space allows us to postulate the existence of far more overlapping sub-images than we imagine. However, even in cases where the components are sub-phenomenal, the nature of the combinatorial effect is not very different. The nature of the background seems to derive from the overlapping of the simultaneously space-like and object-like nature of the region surrounding the main object, in a manner not unlike how the nature of a yellow triangle seems to derive from the overlapping of shape and color.

In conclusion, in order to understand how consciousness is made, we have to break it down into the hidden component and the manifest component of which each Spatial Belonging is made. Then, we have to take into account that what we see comes from the "assembly" of the individual Spatial Belongings in a hierarchical structure. It could be argued that this is an assembly of brain components. However, the individual Spatial Belongings correspond to the realization of the property that is needed, at a sub-phenomenal level, to be able to speak of a minimal state of consciousness, i.e., the appearance which, at the phenomenal level, manifests itself in the possibility of perceiving and distinguishing each conscious content.

The hypothesis of a conscious state having components that we cannot experience seems counterintuitive, or even contradictory. However, if sufficiently well-founded insofar as it is compatible with PC, we cannot rule out this hypothesis. Such a structure may provide a kind of link that can bridge – or at least reduce – the explanatory gap between experience and brain processes and thus help solve the hard problem. With their hidden component, appearance-related processes can account for more complex and differentiated aspects, such as phenomenal and qualitative aspects, on the basis of a few simple principles. In this paper, I only analyze those related to early perceptual experience.

At the same time, the hidden structure of consciousness may be more easily explained in terms of brain organization. Phenomenal experience and brain structure are too different or "distant" to be directly compatible. If we think of the difference as a kind of excessive gap, which would imply *direct* non-reducibility, there is an alternative possibility to the dichotomy between dualism and monism. It consists in hypothesizing that the structure of consciousness, while not conscious in the full meaning of the term – and thus equatable to phenomenal experience – nor directly accessible to introspection, is also not equatable to a non-conscious state. The hidden structure of consciousness, which I have identified in HSB, can be considered a bridge structure which places itself at an intermediate level between experience and physical properties.

As stated by Loorits (2014), it is the non-structural nature of qualia that makes them extremely difficult to explain at the brain level. In this model, the qualities of early visual experience correspond to the HSB. The structural nature of the HSB makes it more easily explicable in terms of brain organization, helping to bridge the explanatory gap between physical properties and experience. Although the limitations of the paper do not allow us to address this question, I will briefly mention a possible direction of research.

An important aspect to take into consideration is that the structure of the HSB is unitary. All interactions involving individual Spatial Belongings and their organization on different hierarchical levels are simultaneous, closely integrated and involve the whole field. Therefore, it is likely that this structure cannot be provided by conventional neuronal organization. Most of neurobiological theories of consciousness look primarily to synaptic firing as the physical substrate of consciousness. However, all neurons also produce electromagnetic fields. Various spatiotemporal scales of electromagnetic fields are generated by, but not identical with the anatomy of the brain. Jones and Hunt (2023) "suggest that these fields, in both their local and global forms, may be the primary seat of consciousness, working as a gestalt with synaptic firing and other aspects of neuroanatomy to produce the marvelous complexity of minds."

Field theories have made real progress in explaining how fields integrate colors to form unified pictorial images (Jones and Hunt, 2023). At the same time, these hypotheses do not seem capable of explaining qualia. However, the compatibility of electromagnetic fields with the HSB, which in turn is compatible with the qualities of early vision, could be explored. If we refer to early vision, these fields could be supported by brain areas whose units are linked by a grid-like connectivity (Haun and Tononi, 2019). Some of them are retinotopic maps that retain, at least approximately, the relationships present in the field of the proximal stimulus. But, of course, several other hypotheses can be considered.

A significant limitation of this model is that the HSB does not explain consciousness outside of early visual experience. Other phenomenal aspects like feelings, emotion, imagination or dreaming need to be addressed. As I have argued, their explanation probably requires a higher structural level.

What I have proposed in this paper is a possible explanation of what I have identified as the explanandum, i.e., the unitary set of qualities we find in early vision (Forti, 2024). It is precisely elements of that explanandum that help provide the explanation. In turn, the explanation, which consists in postulating the existence of a Hierarchy of Spatial Belongings nested within each other, is an additional explanandum. Exploring the nature of this explanandum requires further research.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

BF: Writing – original draft.

Funding

The author declares that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Albertazzi, L. (2019). Experimental phenomenology: what it is and what it is not. *Synthese* 198, 2191–2212. doi: 10.1007/s11229-019-02209-6

Albertazzi, L. (2021). Experimental phenomenology, a challenge. *Psychol. Conscious. Theory Res. Pract.* 8, 105–115. doi: 10.1037/cns0000287

Albertazzi, L., Canal, L., Micciolo, R., and Hachen, I. (2021). The perceptual organisation of visual elements: lines. *Brain Sci.* 11:1585. doi: 10.3390/brainsci11121585

Baars, B. (1997). *In the theater of consciousness*. New York, NY: Oxford University Press.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends Cogn. Sci.* 6, 47–52. doi: 10.1016/s1364-6613(00)01819-2

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog. Brain Res.* 150, 45–53. doi: 10.1016/S0079-6123(05)50004-9

Bayne, T. (2010). *The unity of consciousness*. Oxford, United Kingdom: Oxford University Press

Bayne, T., and Spence, C. (2015). "Multisensory perception" in *The Oxford handbook of philosophy of perception*. ed. M. Matthen (Oxford: University Press), 603–620.

Bennett, D., and Hill, C. (2014). *Sensory integration and the Unity of consciousness*. Cambridge, MA MIT Press.

Block, N. (1995). On a confusion about a function of consciousness. *Brain Behav. Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188

Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci.* 9, 46–52. doi: 10.1016/j.tics.2004.12.006

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the Back of the cerebral cortex? Clinical and Neuroimaging Evidence. *J. Neurosci.* 37, 9603–9613. doi: 10.1523/JNEUROSCI.3218-16.2017

Briscoe, R. E. (2011). Mental imagery and the varieties of amodal perception. *Pac. Philos. Q.* 92, 153–173. doi: 10.1111/j.1468-0114.2011.01393.x

Brogaard, B., Gatzia, D. E., and Chomanski, B. (2021). Consciousness and information integration. *Synthese* 198, 763–792. doi: 10.1007/s11229-020-02613-3

Brown, R., Lau, H., and LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends Cogn. Sci.* 23, 754–768. doi: 10.1016/j.tics.2019.06.009

Calabi, C. (2013). Amodal completion, perception and visual imagery. *Phenomenol. Mind* 4, 170–177. doi: 10.13128/Phe_Mi-19599

Calabrò, R. S., Cacciola, A., Bramanti, P., and Milardi, D. (2015). Neural correlates of consciousness: what we know and what we have to learn! *Neurol. Sci.* 36, 505–513. doi: 10.1007/s10072-015-2072-x

Cave, A. (2013). *Shooting with soul: 44 photography exercises exploring life, beauty and self-expression - from film to smartphones, capture images using cameras from yesterday and today*. Beverly, MA: Quarry Books.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.

Crane, T. (2003). "The intentional structure of consciousness" in *Consciousness: new philosophical perspectives*. eds. Q. Smith and A. Jokic (Oxford, UK: Oxford University Press), 33–56.

Crane, T. (2009). "Intentionalism" in *The Oxford handbook to the philosophy of mind*. eds. B. McLaughlin and A. Beckermann (Oxford, UK: Oxford University Press), 474–493.

Crick, F., and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275.

Crick, F., and Koch, C. (1998). Consciousness and neuroscience. *Cereb. Cortex* 8, 97–107. doi: 10.1093/cercor/8.2.971

Davidson, M. J., Alais, D., van Boxtel, J. J., and Tsuchiya, N. (2018). Attention periodically samples competing stimuli during binocular rivalry. *eLife* 7:e40868. doi: 10.7554/eLife.40868

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. New York: Viking Penguin.

Dehaene, S., Kerszberg, M., and Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA* 95, 14529–14534. doi: 10.1073/pnas.95.24.14529

Dennett, D. C. (1988). "Quining qualia" in *Consciousness in contemporary science*. eds. A. J. Marcel and E. Bisiach (Oxford: Clarendon Press), 42–77.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little, Brown and Company.

Dennett, D. C. (2005). *Sweet dreams: Philosophical obstacles to a science of consciousness*. Cambridge, MA: MIT Press.

Dennett, D. C. (2015). "Why and how does consciousness seem the way it seems?" in *Open MIND*. eds. T. Metzinger and J. M. Windt (Frankfurt am Main: MIND Group)

Deroy, O., Chen, Y. C., and Spence, C. (2014). Multisensory constraints on awareness. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 369:20130207. doi: 10.1098/rstb.2013.0207

Dieter, K. C., Brascamp, J., Tadin, D., and Blake, R. (2016). Does visual attention drive the dynamics of bistable perception? *Atten. Percept. Psychophys.* 78, 1861–1873. doi: 10.3758/s13414-016-1143-2

Doesburg, S. M., Green, J. J., McDonald, J. J., and Ward, L. M. (2009). Rhythms of consciousness: binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception. *PLoS One* 4:e6142. doi: 10.1371/journal.pone.0006142

Edelman, G. M. (2003). Naturalizing consciousness: a theoretical framework. *Proc. Natl. Acad. Sci. USA* 100, 5520–5524. doi: 10.1073/pnas.0931349100

Feldman, J. (2013). The neural binding problem(s). *Cogn. Neurodyn.* 7, 1–11. doi: 10.1007/s11571-012-9219-8

Fetsch, C. R., DeAngelis, G. C., and Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensor neurons. *Nat. Rev. Neurosci.* 14, 429–442. doi: 10.1038/nrn3503

Fingelkurts, A. A., Fingelkurts, A. A., and Neves, C. F. H. (2009). Phenomenological architecture of a mind and operational architectonics of the brain: the unified metastable continuum. *New Math. Nat. Comput.* 5, 221–244. doi: 10.1142/S1793005709001258

Forti, B. (2009). How could phenomenal consciousness be involved in mental function? *New Ideas Psychol.* 27, 312–325. doi: 10.1016/j.newideapsych.2008.10.001

Forti, B. (2015). What are the limits of gestalt theory? *Gestalt Theory* 37, 161–188.

Forti, B. (2024). Approaching the nature of consciousness through a phenomenal analysis of early vision. What is the explanandum? *Front. Psychol.* 15:1329259. doi: 10.3389/fpsyg.2024.1329259

Gallagher, S., and Zahavi, D. (2008). *The phenomenological mind*. New York: Routledge

Gallotto, S., Sack, A. T., Schuhmann, T., and de Graaf, T. A. (2017). Oscillatory correlates of visual consciousness. *Front. Psychol.* 8:1147. doi: 10.3389/fpsyg.2017.01147

Gennaro, R. J. (2004). *Higher-order theories of consciousness: an anthology*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Gescheider, G. A. (1997). *Psychophysics: the fundamentals*. Mahwah, NJ: Laurence Erlbaum Associates.

Hameroff, S., and Penrose, R. (2014). Consciousness in the universe: a review of the 'Orch OR' theory. *Phys Life Rev* 11, 39–78. doi: 10.1016/j.plrev.2013.08.002

Haun, A., and Tononi, G. (2019). Why does space feel the way it does? Towards a principled account of spatial experience. *Entropy* 21:1160. doi: 10.3390/e21121160

Hirschhorn, R., Kahane, O., Gur-Arie, I., Faivre, N., and Mudrik, L. (2021). Windows of integration hypothesis revisited. *Front. Hum. Neurosci.* 14:617187. doi: 10.3389/fnhum.2020.617187

Hunt, T., and Jones, M. (2023). Fields or firings? Comparing the spike code and the electromagnetic field hypothesis. *Front. Psychol.* 14:1029715. doi: 10.3389/fpsyg.2023.1029715

Isbister, J. B., Eguchi, A., Ahmad, N., Galeazzi, J. M., Buckley, M. J., and Stringer, S. (2018). A new approach to solving the feature-binding problem in primate vision. *Interface Focus* 8:20180021. doi: 10.1098/rsfs.2018.0021

James, W. (1890). *The principles of psychology*. London: MacMillan.

Jerath, R., and Beveridge, C. (2019). Multimodal integration and phenomenal spatiotemporal binding: a perspective from the default space theory. *Front. Integr. Neurosci.* 13:2. doi: 10.3389/fnint.2019.00002

Jerath, R., Beveridge, C., and Jensen, M. (2019). On the hierarchical Organization of Oscillatory Assemblies: layered superimposition and a global bioelectric framework. *Front. Hum. Neurosci.* 13:426. doi: 10.3389/fnhum.2019.00426

Jerath, R., and Crawford, M. W. (2014). Neural correlates of visuospatial consciousness in 3D default space: insights from contralateral neglect syndrome. *Conscious. Cogn.* 28, 81–93. doi: 10.1016/j.concog.2014.06.008

Jerath, R., Crawford, M. W., and Barnes, V. A. (2015). Functional representation of vision within the mind: a visual consciousness model based in 3D default space. *J. Med. Hypotheses Ideas* 9, 45–56. doi: 10.1016/j.jmhi.2015.02.001

Jones, M. W., and Hunt, T. (2023). Electromagnetic-field theories of qualia: can they improve upon standard neuroscience? *Front. Psychol.* 14:1015967. doi: 10.3389/fpsyg.2023.1015967

Kanizsa, G. (1979). *Organization in vision. Essays in gestalt perception*. New York: Praeger.

Kanizsa, G. (1980). *Grammatica del vedere*. Bologna: Il Mulino.

Kesserwani, H. (2020). The clinical, philosophical, evolutionary and mathematical machinery of consciousness: an analytic dissection of the field theories and a consilience of ideas. *Cureus* 12:e12139. doi: 10.7759/cureus.12139

Kimchi, R. (2009). Perceptual organization and visual attention. *Prog. Brain Res.* 176, 15–33. doi: 10.1016/S0079-6123(09)17602-1

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22

Koculak, M., and Wierzchoń, M. (2022). How much consciousness is there in complexity? *Front. Psychol.* 13:983315. doi: 10.3389/fpsyg.2022.983315

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pac. Philos. Q.* 64, 354–361. doi: 10.1111/j.1468-0114.1983.tb00207.x

Llinas, R. R., Ribary, U., Joliot, M., and Wang, X.-J. (1994). "Content and context in temporal thalamocortical binding" in *Temporal coding in the brain*. eds. G. Buzsaki, R. R. Llinas and W. Singer (Berlin: Springer Verlag)

Loorits, K. (2014). Structural qualia: a solution to the hard problem of consciousness. *Front. Psychol.* 5:237. doi: 10.3389/fpsyg.2014.00237

MacIver, M. B. (2022). Consciousness and inward electromagnetic field interactions. *Front. Hum. Neurosci.* 16:1032339. doi: 10.3389/fnhum.2022.1032339

Maillé, S., and Lynn, M. (2020). Reconciling current theories of consciousness. *J. Neurosci.* 40, 1994–1996. doi: 10.1523/JNEUROSCI.2740-19.2020

Marius, M. (2014). The explanatory gap: 30 years after. *Procedia. Soc. Behav. Sci.* 127, 292–296. doi: 10.1016/j.sbspro.2014.03.258

Mashour, G. A., Roelfsema, P., Changeux, J. P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron* 105, 776–798. doi: 10.1016/j.neuron.2020.01.026

McFadden, J. (2020). Integrating information in the brain's EM field: the cemi field theory of consciousness. *Neurosci. Conscious.* 2020:niaa016. doi: 10.1093/nc/niaa016

McFadden, J. (2023). Consciousness: matter or EMF? *Front. Hum. Neurosci.* 16:1024934. doi: 10.3389/fnhum.2022.1024934

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Merlo, G. (2020). Appearance, reality, and the Meta-problem of consciousness. *J. Conscious. Stud.* 27, 120–130.

Musatti, C. (1957). *Condizioni dell'esperienza e fondazione della psicologia*. Firenze: Giunti Editore.

Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450. doi: 10.2307/2183914

Nanay, B. (2018). The importance of Amodal completion in everyday perception. *Iperception* 9:2041669518788887. doi: 10.1177/2041669518788887

Noë, A., and O'Regan, J. (2000). Perception, Attention and the Grand Illusion. *Psyche* 6, 6–15.

Noë, A., Pessoa, L., and Thompson, E. (2000). Beyond the grand illusion: what change blindness really teaches us about vision. *Vis. Cogn.* 7, 93–106. doi: 10.1080/135062800394702

Noel, J. P., Ishizawa, Y., Patel, S. R., Eskandar, E. N., and Wallace, M. T. (2019). Leveraging non-human primate multisensory neurons and circuits in assessing consciousness theory. *J. Neurosci.* 39, 7485–7500. doi: 10.1523/JNEUROSCI.0934-19.2019

Northoff, G., Klar, P., Bein, M., and Safron, A. (2023). As without, so within: how the brain's temporo-spatial alignment to the environment shapes consciousness. *Interface Focus* 13:20220076. doi: 10.1098/rsfs.2022.0076

O'Callaghan, C. (2015). The multisensory character of perception. *J. Philos.* 112, 551–569. doi: 10.5840/jphil20151121035

Peterson, M., and Salvagio, E. (2010). Figure-ground perception. *Scholarpedia* 5:4320. doi: 10.4249/scholarpedia.4320

Pinna, B. (2012). What is the meaning of shape? *Gestalt Theory* 33, 383–422.

Pitts, M. A., Lutsyshyna, L. A., and Hillyard, S. A. (2018). The relationship between attention and consciousness: an expanded taxonomy and implications for 'no-report' paradigms. *Philos. Trans. R. Soc. B* 373:20170348. doi: 10.1098/rstb.2017.0348

Polák, M., and Marvan, T. (2018). Neural correlates of consciousness meet the theory of identity. *Front. Psychol.* 9:1269. doi: 10.3389/fpsyg.2018.01269

Prinz, J. (2018). "Is consciousness a trick or a treat?" in *The philosophy of Daniel Dennett*. ed. B. Huebner (New York: Oxford Academic)

Rensink, R. A. (2004). Visual sensing without seeing. *Psychol. Sci.* 15, 27–32. doi: 10.1111/j.0963-7214.2004.01501005.x

Revonsuo, A. (2006). *Inner presence: consciousness as a biological phenomenon*. Cambridge: MIT Press.

Revonsuo, A., and Newman, J. (1999). Binding and consciousness. *Conscious. Cogn.* 8, 123–127. doi: 10.1006/ccog.1999.0393

Sanfey, J. (2023). Simultaneity of consciousness with physical reality: the key that unlocks the mind-matter problem. *Front. Psychol.* 14:1173653. doi: 10.3389/fpsyg.2023.1173653

Sarasso, S., Casali, A. G., Casarotto, S., Rosanova, M., Sinigaglia, C., and Massimini, M. (2021). Consciousness and complexity: a consilience of evidence. *Neurosci. Conscious.* 2021:niab023. doi: 10.1093/nc/niab023

Schmidt, T. T., Jagannathan, N., Ljubljanac, M., Havier, A., and Nierhaus, T. (2020). The multimodal Ganzfeld-induced altered state of consciousness induces decreased thalamo-cortical coupling. *Sci. Rep.* 10:18686. doi: 10.1038/s41598-020-75019-3

Schulte, P. (2023). *Mental content*. Cambridge: Cambridge University Press

Schurger, A., and Graziano, M. (2022). Consciousness explained or described? *Neurosci. Conscious.* 2022:niac001. doi: 10.1093/nc/niac001

Schwarzkopf, D. S., and Rees, G. (2015). "Perceptual organization and consciousness" in *The Oxford handbook of perceptual organization*. ed. J. Wagemans (Oxford, U.K: Oxford University Press), 799–819.

Scrivener, C. L., Malik, A., Lindner, M., and Roesch, E. B. (2021). Sensing and seeing associated with overlapping occipitoparietal activation in simultaneous EEG-fMRI. *Neurosci. Conscious.* 21, 2021:niab008. doi: 10.1093/nc/niab008

Searle, J. R. (1997). *The mistery of consciousness*. New York: The New York Review of Books.

Searle, J. R. (2004). *Mind. A brief introduction*. New York: Oxford University Press.

Seth, A. K., and Bayne, T. (2022). Theories of consciousness. *Nat. Rev. Neurosci.* 23, 439–452. doi: 10.1038/s41583-022-00587-4

Skokowski, P. (2022). Sensing Qualia. *Front. Syst. Neurosci.* 16:795405. doi: 10.3389/fnsys.2022.795405

Smit, S., Moerel, D., Zopf, R., and Rich, A. N. (2023). Vicarious touch: overlapping neural patterns between seeing and feeling touch. *NeuroImage* 278:120269. doi: 10.1016/j.neuroimage.2023.120269

Smith, D. W. (2018). "Phenomenology" in *The Stanford encyclopedia of philosophy (summer 2018 Edn.)*. ed. E. N. Zalta. Stanford, CA: Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/sum2018/entries/phenomenology.

Todorović, D. (2008). Gestalt principles. *Scholarpedia* 3:5345. doi: 10.4249/scholarpedia.5345

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242. doi: 10.2307/25470707

Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846

Tononi, G., and Koch, C. (2008). The neural correlates of consciousness: an update. *Ann. N. Y. Acad. Sci.* 1124, 239–261. doi: 10.1196/annals.1440.004

Tononi, G., and Koch, C. (2015). Consciousness: here, there and everywhere? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 1–18. doi: 10.1098/rstb.2014.0167

Toribio, J. (1993). Why there still has to be a theory of consciousness. *Conscious. Cogn.* 2, 28–47. doi: 10.1006/ccog.1993.1003

Tuszynski, J. A. (2020). From quantum chemistry to quantum biology: a path toward consciousness. *J. Integr. Neurosci.* 19, 687–700. doi: 10.31083/j.jin.2020.04.393

Tyler, C. W. (2020). Ten testable properties of consciousness. *Front. Psychol.* 11:1144. doi: 10.3389/fpsyg.2020.01144

Van Gulick, R. (2022). "Consciousness" in *The Stanford encyclopedia of philosophy (winter 2022 edition)*. eds. E. N. Zalta and U. Nodelman. Stanford, CA: Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/win2022/entries/consciousness.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., et al. (2012). A century of gestalt psychology in visual perception: I. Conceptual and theoretical foundations. *Psychol. Bull.* 138, 1218–1252. doi: 10.1037/a0029334

Walling, P. T. (2019). An update on dimensions of consciousness. *Proc Bayl. Univ. Med. Cent.* 33, 126–130. doi: 10.1080/08998280.2019.1656009

Ward, L., and Guevara, R. (2022). Qualia and phenomenal consciousness Arise from the information structure of an electromagnetic field in the brain. *Front. Hum. Neurosci.* 16:874241. doi: 10.3389/fnhum.2022.874241

Whiting, D. (2016). On the appearance and reality of mind. *J. Mind Behav.* 37, 47–70.

Zeman, A. (2001). Consciousness. *Brain* 124, 1263–1289. doi: 10.1093/brain/124.7.1263

Zhi, G. S., and Xiu, R. L. (2023). Quantum theory of consciousness. *J. Appl. Math. Phys.* 11, 2652–2670. doi: 10.4236/jamp.2023.119174

Zhu, M., Hardstone, R., and He, B. J. (2022). Neural oscillations promoting perceptual stability and perceptual memory during bistable perception. *Sci. Rep.* 12:2760. doi: 10.1038/s41598-022-06570-4

TYPE Original Research
PUBLISHED 10 July 2024
DOI 10.3389/fpsyg.2024.1337589

Check for updates

OPEN ACCESS

EDITED BY
Antonino Raffone,
Sapienza University of Rome, Italy

REVIEWED BY
Ken Mogi,
Sony Computer Science Laboratories, Japan
Charlène Aubinet,
University of Liège, Belgium

*CORRESPONDENCE
Ryoichi Watanabe
✉ watanabe.ryoichi.5w@kyoto-u.ac.jp
Yusuke Moriguchi
✉ moriguchi.yusuke.8s@kyoto-u.ac.jp

RECEIVED 13 November 2023
ACCEPTED 24 June 2024
PUBLISHED 10 July 2024

CITATION
Watanabe R and Moriguchi Y (2024) Development of emergent processes and threshold of consciousness with levels of processing.
Front. Psychol. 15:1337589.
doi: 10.3389/fpsyg.2024.1337589

Development of emergent processes and threshold of consciousness with levels of processing

Ryoichi Watanabe* and Yusuke Moriguchi*

Graduation School of Letters, Kyoto University, Kyoto, Japan

Introduction: The transition of experience from unconscious to conscious, the emergent process, is a crucial topic in consciousness studies. Three frameworks exist to explain the process: (1) consciousness arises in an all-or-none manner; (2) consciousness arises gradually; (3) consciousness arises either all at once or gradually, depending on the level of stimulus processing (low- vs. high-level). However, the development of emergent processes of consciousness remains unclear. This study examines the development of emergent processes of consciousness based on the level of stimulus processing framework.

Methods: Ninety-nine children (5–12 year-olds) and adults participated in two online discrimination tasks. These tasks involved color discrimination as lower-level processing and number magnitude discrimination as higher-level processing, as well as backward masking with stimulus onset asynchronies (SOAs) varying from 16.7 to 266.7 ms. We measured objective discrimination accuracy and used a 4-scale Perceptual Awareness Scale (PAS) to assess subjective awareness. We fit the data to a four-parameter nonlinear function to estimate the center of the slope (threshold) and the range of the slope (gradualness, the measure of emergent process of consciousness) of the model.

Results: The results showed the threshold of objective discrimination was significantly higher in 5–6 year-olds than in 7–12 year-olds, but not of subjective awareness. The emergent process of objective discrimination in the number task was more gradual than in the color task.

Discussion: The findings suggest that the thresholds of subjective awareness in 5–6 year-olds and objective discrimination in 7–9 year-olds are similar to those in adults. Moreover, the emergent processes of subjective awareness and objective discrimination in 5–6 year-olds are also similar to those in adults. Our results support the level of processing hypothesis but suggest that its effects may differ across developmental stages.

KEYWORDS

children, visual consciousness, backward masking, awareness, level of processing

1 Introduction

Visual consciousness means subjective and phenomenal visual experience (e.g., what it is like to see an image) (Koch et al., 2016). Transitioning from unconscious to conscious, the emergence of visual consciousness is a prominent topic in consciousness studies (Baars, 2005; Del Cul et al., 2007; Koch and Preuschoff, 2007; Dehaene and Changeux, 2011; Sandberg et al., 2011; Windey et al., 2013; Koch et al., 2016). Research on visual consciousness's emergent processes and neural mechanisms has focused on studies in human adults and macaques (Koch et al., 2016). However, how they develop is almost unknown. Research on the developmental aspects of consciousness has increased recently, focusing primarily on the origins of consciousness in fetuses and newborns (Bayne et al., 2023). However, research on the developmental changes afterwards is lacking. Many consciousness researchers believe that understanding the developmental aspects of visual consciousness is essential for consciousness theories (Mashour et al., 2020; Seth and Tim, 2022). The present study focused on the developing emergent process and threshold of visual consciousness in 5–12 year-olds and adults.

1.1 Review of literature

The emergent processes and thresholds of consciousness are vital to examining the transformation from unconsciousness to consciousness. Masking methods have been used widely in examining the visual consciousness's emergent process and thresholds (Sandberg et al., 2011; Windey et al., 2013; Thiruvasagam and Srinivasan, 2021). The emergent process and threshold of consciousness have been examined using subjective awareness and objective discrimination of task performance (Sandberg et al., 2011). Subjective awareness is measured by the two choices of awareness or unawareness of the stimulus or by the Perceptual Awareness Scale (PAS), which assesses perceptual awareness in a graded manner (Ramsøy and Overgaard, 2004). By using PAS, we can measure the presence and intensity of awareness in a graded manner. As the duration of the stimulus presentation increases, the subjective awareness rate and intensity increases (Sandberg et al., 2010, 2011; Windey et al., 2013). The task accuracy measures objective discrimination performance and d' using signal detection theory. The d' of the signal detection theory is often used as a measure of stimulus discrimination performance; the larger the d', the greater the discrimination performance (Macmillan and Douglas Creelman, 2004). The signal detection theory involves calculating four indicators: hit, miss, false alarm, and correct rejection, based on the stimulus combinations between presented and responded. The d' value is calculated from the difference between the *z*-scores (or standard deviations) of the hit rate and the false alarm rate. For example, in the case of a stimulus color judgment task, if a red stimulus is presented and the participant responds that they saw a red stimulus, a hit is indicated; if the participant responds that they saw a blue stimulus, the response is a miss. Conversely, if a blue stimulus is presented and the participant responds that they saw a red one, this indicates a false alarm, and if the participant responds that they saw a blue stimulus, the response is a correct rejection. As stimulus presentation duration increases, the objective discrimination performance increases (Sandberg et al., 2010, 2011; Windey et al., 2013).

1.1.1 Theory of consciousness

There are three leading positions regarding the process of transitioning from unconscious to conscious (Jimenez et al., 2020): (1) consciousness arises all-or-none (Sergent and Dehaene, 2004; Del Cul et al., 2007; Sekar et al., 2013; Asplund et al., 2014), supported by Global Neuronal Workspace Theory (GNWT; Dehaene and Naccache, 2001; Dehaene and Changeux, 2011); (2) consciousness arises gradually (Ramsøy and Overgaard, 2004; Overgaard et al., 2006; Seth et al., 2008; Pretorius et al., 2016), supported by Recurrent Process Theory (RPT; Lamme, 2006); (3) consciousness arises either all at once or gradually depending on the level of stimulus processing (Windey et al., 2013; Anzulewicz et al., 2015; Binder et al., 2017; Derda et al., 2019; Jimenez et al., 2019, 2021), supported by Level of Processing Hypothesis (LoPH; Windey et al., 2013; Windey and Cleeremans, 2015). The next section explains each position and theory in more details. The position suggesting that consciousness emerges as all-or-none contradicts the one stating that consciousness emerges gradually. However, a comprehensive position exists which covers both ideas by implying that consciousness emerges differently depending on the level of stimulus processing.

1.1.2 All-or-none emergent process and GNWT

The all-or-nothing position assumes that the stage of consciousness is binary, either aware or unaware. *Global Neuronal Workspace Theory* (GNWT) supports this position. GNWT postulates that when the intensity of a stimulus exceeds a certain threshold, the stimulus reaches the global workspace and can be consciously accessed (all-aware) (Dehaene and Naccache, 2001; Dehaene and Changeux, 2011).

Many empirical studies support this position (Del Cul et al., 2006, 2007; Sekar et al., 2013; Asplund et al., 2014). Global workspace is a concept similar to working memory; when information accesses the global workspace, it can be consciously used for other modalities such as reports and memory (Baars, 2005). Del Cul et al. (2006) showed that the longer the target-mask stimulus onset asynchrony (SOA), the higher the objective performance and subjective awareness rating. Moreover, they showed that the trajectory of the objective performance and subjective visibility rating were sigmoidal curves centered on the threshold, as well as that the emergent process of visual consciousness follows a sigmoid curve, suggesting that visual consciousness emerges as all-or-none. Thus, GNWT suggests that objective discrimination and subjective awareness are all or none.

The GNWT argues that information is accessed by consciousness when transferred through the frontal–parietal network to the frontal lobes and the whole brain (Dehaene and Naccache, 2001; Dehaene and Changeux, 2011). Thus, it is suggested that the development of the frontal–parietal network and frontal lobes is related to the emergent process of consciousness (e.g., the threshold or the precision).

The volume and density of gray and white matter in the frontal and parietal lobes peaks during childhood (Giedd et al., 1999; Nagy et al., 2004; Tanaka et al., 2012). Moreover, the activity and connectivity of the frontal and parietal lobes become stronger during childhood and adolescence (Adleman et al., 2002; Kwon et al., 2002; Gogtay et al., 2004). Finally, activity in frontal–parietal regions begins to function from early childhood (Moriguchi and Hiraki, 2013), with weak activity in frontal–parietal regions during inhibition tasks, reflected by the coupling of the frontal–parietal network (Mehnert et al., 2013). Considering the development of these frontal–parietal networks, the threshold in the emergent process of consciousness is predicted to become smaller and more precise as the networks develop.

1.1.3 Gradual emergent process and RPT

The gradual position assumes that the stage of consciousness is not binary but gradual from non-aware to all-aware. *Recurrent processing theory* (RPT) supports this position. RPT postulates that as a stimulus is processed progressively, it becomes more explicit to consciousness (Lamme, 2006). In other words, as the intensity of the stimulus increases (e.g., SOA or stimulus contrast), the conscious experience of that stimulus becomes clearer. Many empirical studies support this position (Christensen et al., 2006; Overgaard et al., 2006; Sandberg et al., 2010, 2011). Sandberg et al. (2010) showed that using the subjective measure PAS, intermediate awareness responses, such as slightly visible or mostly visible, increased for SOAs around the threshold. Our study defines the threshold as the central point of the model's slope. The gradual emergent process is accompanied by its threshold.

Furthermore, they showed that the subjective measures, PAS, confidence rating, and weighting predicted objective discrimination performance. Sandberg et al. (2011) developed a sigmoid function that fits objective discriminant performance and subjective awareness rating. This function can be used to estimate the threshold and steepness of the slope of the model, indicating whether the model is all-or-none or gradual. Subjective awareness rating and objective discrimination performance were showed to increase gradually as SOA increased, which suggests that visual consciousness emerged gradually.

The RPT argues that the recurrent loop of the visual cortex produces a clearer visual consciousness, implying that the development of the visual cortex and the recurrent loop is related to the emergent process of consciousness.

Other studies indicate a recurrent loop in visual information after 7–8 months (Nakashima et al., 2021). The total number and density of synapses peaks at 1 year of age and decreases to the same level as adults at about 10–11 years of age (Huttenlocher et al., 1982; Huttenlocher, 1990). Although there is not necessarily a relationship between synaptic density and cognitive function, based on the recurrent loop structure, the emergent visual consciousness process is predicted to develop from infancy to childhood and remain similar during adulthood.

1.1.4 All-or-none/gradual emergent process and LoPH

Windey et al. (2013) and Windey and Cleeremans (2015) integrated the contradictions between these two positions by varying the stimulus or task's processing level and presented the level of processing hypothesis (LoPH). According to LoPH, when the task requires higher-order processing of the stimulus, the emergent process of consciousness is either all-or-none, and when the task requires lower-order processing of the stimulus, the emergent process of consciousness is gradual (Windey et al., 2013; Windey and Cleeremans, 2015). The level of processing corresponds to feed-forward brain processing of visual information, with higher processing levels referring to the meaning or category of the visual object and lower processing levels referring to the shape or color of the visual object (Hochstein and Ahissar, 2002; Windey and Cleeremans, 2015). Windey et al. (2013) examined the difference in the slope steepness of objective discrimination performance and subjective awareness rating between color judgments and number magnitude judgments. They used the color judgements as the lower-order processing condition and the number magnitude judgements as higher-order one, with numbers in different colors as stimuli. They showed that the steepness of the slope of the model in the low-order processing condition was more gradual than that in the high-order condition for both objective discrimination performance and subjective awareness rating. Empirical studies that support the all-or-none process used tasks that require judgments about the meaning of stimuli (e.g., the meaning of letters or the magnitude of numbers) (Del Cul et al., 2006, 2007; Sekar et al., 2013; Asplund et al., 2014). In contrast, empirical studies that support a gradual process of consciousness used tasks that require judgments of stimulus characteristics (e.g., the color of letters or the shape of a stimulus) (Christensen et al., 2006; Overgaard et al., 2006; Sandberg et al., 2010, 2011).

Considering the all-or-none, gradual, and LoPH positions, the threshold for visual consciousness is predicted to decrease with age from preschool to school, and similar to adults in late childhood. However, no previous studies examine the developmental differences in the emergent process of consciousness with the level of stimulus processing.

Figure 1A shows the model of the all-or-none emergent process, and Figure 1B shows the gradual emergent process. The level of processing model is drawn in Figure 1A for higher-order processing stimuli and in Figure 1B for lower-order processing stimuli (Figure 1C).

1.1.5 Development of the emergent process and threshold of consciousness

The development of the emergent processes of consciousness is still unclear. Research with children has focused on the objective discrimination performance and thresholds of visual stimuli using the backward masking paradigm. In this paradigm, the target and mask stimuli are presented in order, with manipulated time between the two stimuli (e.g., SOA) (Breitmeyer and Ogmen, 1984). When this time is shorter, for example 20 ms, both objective discrimination and subjective awareness are low. On the other hand, as the time between stimuli increases (e.g., 100 ms), the objective discrimination and subjective awareness of the stimuli also increase (Del Cul et al., 2007; Gelskov and Kouider, 2010; Sandberg et al., 2010, 2011; Kouider et al., 2013; Windey et al., 2013; Anzulewicz et al., 2015; Binder et al., 2017; Jimenez et al., 2019; Thiruvasagam and Srinivasan, 2021). Previous studies have shown that young children have lower objective discrimination performance and larger thresholds for discrimination of letter stimuli than school children and adults (Welsandt et al., 1973; LeBlanc et al., 1992; Macchi et al., 2003). They also showed that the objective threshold of children decreased from 5–16 year-olds and was similar to 22 year-olds, but the performance increased from 5–22 year-olds (Welsandt et al., 1973). Recently, Watanabe and Moriguchi (2023) showed that young children have larger thresholds for objective discrimination and subjective awareness than adults and similar emergent processes of objective discrimination and subjective awareness to adults on a form judgment task (i.e., judgment of the shape of stimuli) categorized in the lower-level processing. Thus, it is consistent with previous studies that objective discrimination performance increases and thresholds decrease from preschool age (Welsandt et al., 1973; LeBlanc et al., 1992; Macchi et al., 2003; Watanabe and Moriguchi, 2023).

However, two issues raised in previous studies should be addressed. The first is that previous research examined only the thresholds and emergent processes of subjective awareness and objective discrimination in lower-order but not higher-order processing stimuli. Watanabe and Moriguchi (2023) showed that the threshold for subjective awareness of form stimuli at ages 5–6 is larger than that of adults, but the emergent process is similar. Although form judgments are categorized as low-order processing, examining low- and high-order processing within the same children is necessary. This study used color judgments as low-order and

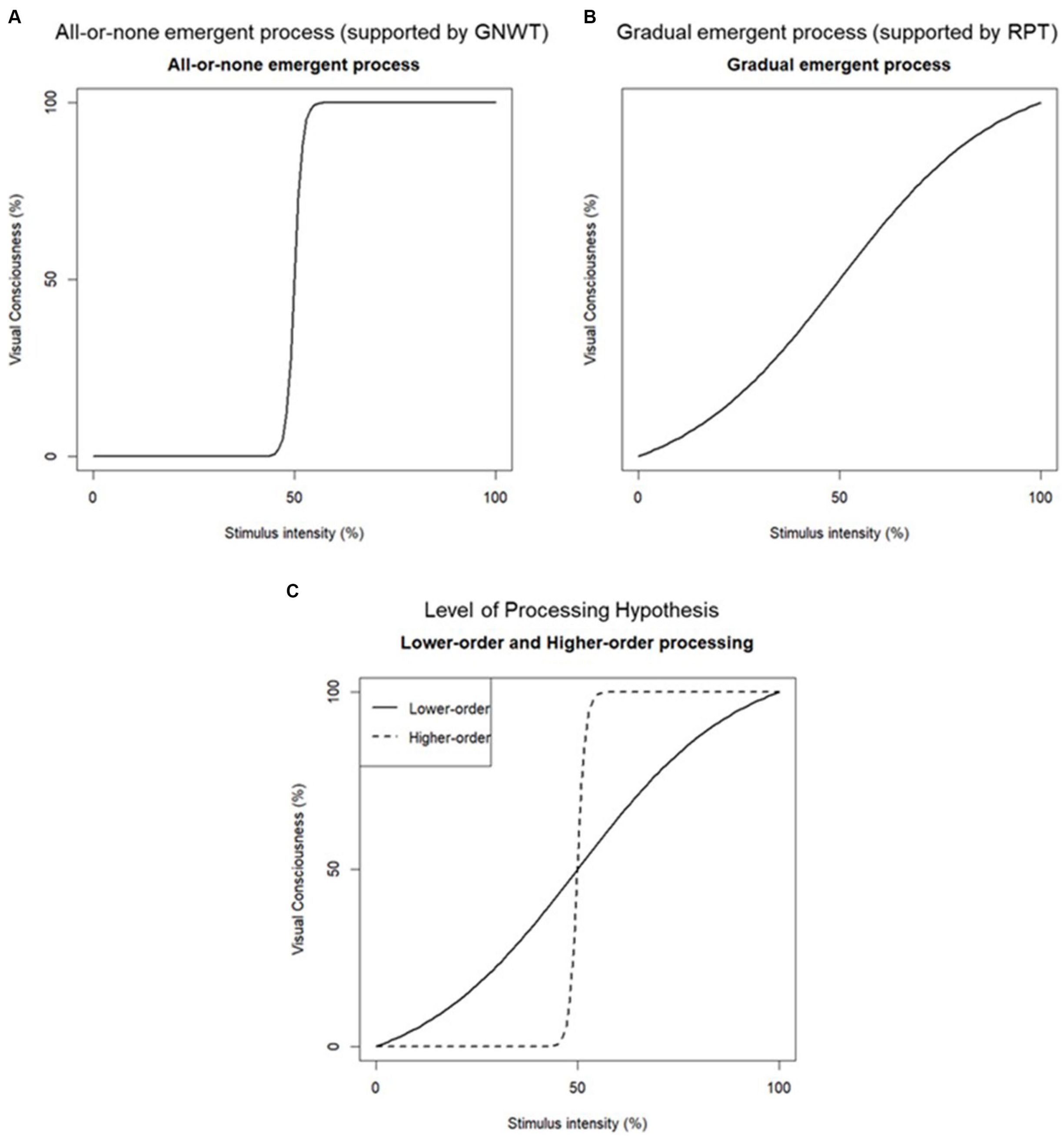

FIGURE 1
Models of the emergent process of visual consciousness. The model of the all-or-none emergent process **(A)**. The model of the gradual emergent process **(B)**. The model of the level of processing hypothesis, with solid line model representing the lower-order processing model and the dotted one showing the higher-order processing model **(C)**. The thresholds are 50%, the stimulus intensity (%; *x*-axis), and the visual consciousness (%; *y*-axis).

number magnitude judgments as high-order processing (Windey et al., 2013). For example, young children with immature executive function may have larger thresholds for number magnitude judgments and lower performance than for color judgments compared to other age groups (Prager et al., 2016), and their emergent processes may follow a different trajectory than those of adults. The second issue that should be addressed is the timing when developmental changes occur. Few studies have assessed the developmental changes of emergent processes, and it is unclear whether the emergent processes develop significantly from early childhood to childhood or later. Developmental trajectories may also differ depending on the level of processing of stimuli or tasks (i.e., low- vs. high-level processing).

1.2 The present study

1.2.1 Purpose of this study

This study examines the developmental aspect of the level of processing hypothesis in visual consciousness. Thus, we analyzed the thresholds and emergent processes of visual consciousness with the level of stimulus processing and age differences. We focused on children aged 5–12 years and adults. Based on previous research, 5 year-olds can respond to objective discrimination and subjective awareness in backward masking tasks (Welsandt et al., 1973; LeBlanc et al., 1992; Macchi et al., 2003; Watanabe and Moriguchi, 2023). Furthermore, the threshold and performance for objective discrimination have increased significantly between the ages of 5–12 years (Welsandt et al., 1973; LeBlanc et al., 1992; Macchi et al., 2003). We adopted the masking task used by Binder et al. (2017). They used colored numbers as the target stimulus, with the color judgment task as the lower-level processing condition and the number magnitude judgment task as the higher-level processing condition.

First, we measured objective discrimination accuracy and subjective awareness of the stimuli using the four-point PAS scale (Overgaard et al., 2006). Children aged 5–6 can also answer PAS (Watanabe and Moriguchi, 2023). Second, we fit the obtained objective discrimination accuracy and subjective awareness ratings for both conditions with a four-parameter nonlinear psychometric function (Sandberg et al., 2011; Windey et al., 2013; Thiruvasagam and Srinivasan, 2021). Finally, we measured the center of the slope of the fitted psychophysical function as the threshold and the steepness of the slope as the gradualness of the emergent process of consciousness. The steeper the slope indicates an all-or-none dichotomous transition.

Based on the level of processing hypothesis (Windey et al., 2013; Anzulewicz et al., 2015), we predicted that the steepness of the slope for the objective discrimination and subjective awareness of the lower-level color judgment task would be larger than that for the higher-level number magnitude judgment task in 5–12 year-olds and adults (Hypothesis 1). Furthermore, based on the previous research about children's thresholds of objective discrimination and subjective awareness (Watanabe and Moriguchi, 2023), we predicted that the thresholds for objective discrimination and subjective awareness for the lower-order color and the higher-order number magnitude judgment tasks decrease with age (Hypothesis 2). We examined the interaction between the level of processing of stimuli and age (Explorative Hypotheses 1 and 2).

1.2.2 Hypotheses

> *Hypothesis 1*. The gradualness of the slope of the objective discrimination and the subjective awareness for lower-level color judgment tasks is larger than that for higher-level number magnitude judgment tasks in 5–12 year-olds and adults.
>
> *Hypothesis 2*. The thresholds of the objective discrimination and the subjective awareness for lower-level color judgment tasks and higher-level number magnitude judgment tasks decrease with age.
>
> *Explorative Hypothesis 1*. There is an interaction in the gradualness of the slope between the level of processing of stimuli and age or not.
>
> *Explorative Hypothesis 2*. There is an interaction in the threshold between the level of processing of stimuli and age or not.

2 Method

2.1 Participants

We recruited 126 participants in four age groups (30 5–6 year-olds, 36 7–9 year-olds, 29 10–12 year-olds, and 31 adults). Twenty-seven participants (seven 5–6 year-olds, ten 7–9 year-olds, three 10–12 year-olds, and seven adults) did not complete the entire trial due to computer errors or omission of participant number. Therefore, our final sample consisted of 99 participants in the four age groups (23 5–6 year-olds, 26 7–9 year-olds, 26 10–12 year-olds, and 24 adults (mean age = 48.13, SD = 6.89)) (Table 1). The sample size estimation using the G*Power 3.1.9.7 (Faul et al., 2007) showed that 19 participants per age group were enough (effect size $f = 0.25$, α error probability = 0.05, Power = 0.95, number of groups = 4, number of measures = 2, correlation among repeated measures = 0.5, and nonsphericity correction $e = 1$) focusing on the main effect of the task.

The Ethics Committee of the Unit for Advanced Studies of the Human Mind, Kyoto University, approved the study procedure (No. 2-P-11). Written informed consent was obtained from the adult participants and the parents of all child participants.

2.2 Stimuli and apparatus

We used four different numbers (1, 3, 7, 9) in four different colors (RGB-values of red = 255, 0, 0; light red = 255, 100, 100; blue = 0, 0, 255; light blue = 100, 100, 255) as target stimuli based on previous research (Sandberg et al., 2011; Windey et al., 2013; Binder et al., 2017; Derda et al., 2019; Thiruvasagam and Srinivasan, 2021). We used random multicolored patches generated from four colored rectangles (size 50 × 50 pixels) as a backward mask stimulus based on previous research (Sandberg et al., 2011; Windey et al., 2013; Binder et al., 2017; Derda et al., 2019; Thiruvasagam and Srinivasan, 2021). We presented the stimuli on a grey background (RGB-values of grey = 125, 125, 125). Participants viewed the stimulus from about 60 cm from their personal computers (PCs) with a 13- to 16-inch, 60 Hz refresh rate and an 800 × 600 pixels resolution. We set the size of the target stimulus at the height of 0.3 (about 5 degrees of visual angle) and the size of the mask stimulus at the height of 0.5 (about 8.5 degrees of visual angle) on the screen. We made the tasks using PsychoPy (Psychology Software Tools, Inc., 2002) and Pavlovia[1].

2.3 Procedure

Before the experiment, participants (or their parents) accessed the task URL through a web browser and downloaded the tasks on the Pavlovia. Participants completed two tasks in approximately 50–60 min. For 5–6 year-old participants, we asked the participants to respond verbally or point to the answer in each trial and their parents to click the appropriate display location for the participants. Our pilot experiment demonstrated that 5–6 year-olds had difficulty responding using a computer mouse. Participants in the other age groups responded by

1 https://pavlovia.org/

TABLE 1 Participant information.

	Total	Computer error	Analyzable data	Catch trials	Model fitting	Final
5–6 year-olds	30	−7	23	−1	−4	18
7–9 year-olds	36	−10	26	0	−2	24
10–12 year-olds	29	−3	26	0	−4	22
Adults	31	−7	24	0	−3	21
All	126	−27	99	−1	−13	85

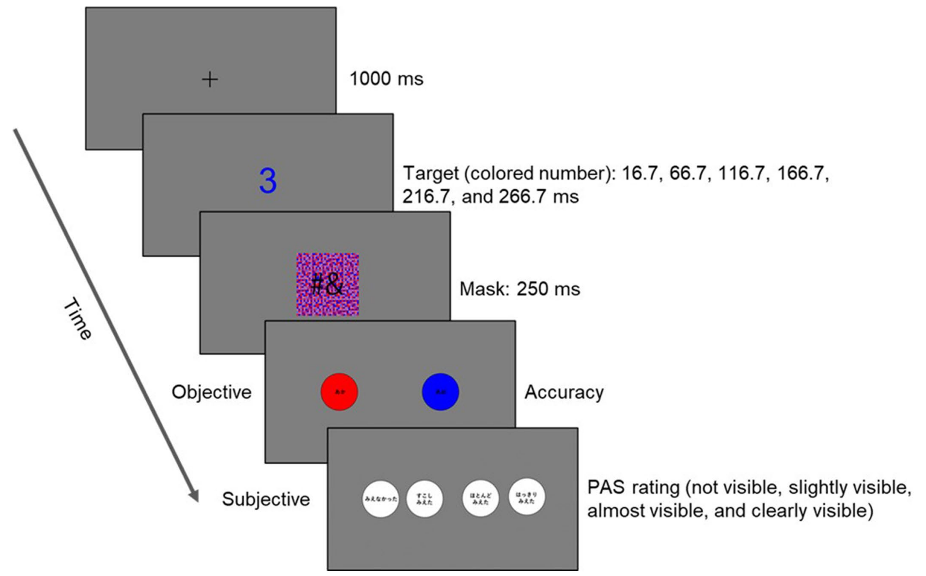

FIGURE 2
Design of the stimuli and procedure of the task. The target stimuli were colored numbers, and the masking stimuli were random multicolored patches. Participants responded to the objective discrimination question, "Was the number red or blue? Was the number larger or smaller than five?" and the subjective awareness question, "How clearly did you see the number?" as the 4-scale PAS.

themselves with a mouse. The experimenter directly instructed all 5–6 year-olds about the task and observed that the parents and children conducted the tasks via the Zoom meeting app to check that they precisely clicked their children's answers.

The design of the procedure, summarized in Figure 2, was similar to previous studies (Sandberg et al., 2011; Windey et al., 2013; Binder et al., 2017; Derda et al., 2019; Thiruvasagam and Srinivasan, 2021). Participants conducted two discrimination tasks (the color and number magnitude judgment tasks) in separate blocks online. Both tasks will involve identical sequences of stimuli. We counterbalanced the order of the tasks among participants. In the color judgment task, participants judged whether the color of the number was red or blue. In the magnitude judgment task, participants judged whether the number was smaller or larger than five. In both tasks, each trial began with a black fixation cross displayed centrally for 1,000 ms. Then, the target stimulus appeared for one of six durations (16.7, 66.7, 116.7, 166.7, 216.7, and 266.7 ms). Then, participants judged whether the number was red/blue or smaller/larger than five by clicking one of the two panels (red and blue or smaller and larger). Then, they evaluated the 4-scale PAS for the target stimulus by clicking one of the four panels (no experience, slight experience, almost experience, and clear experience). Participants conducted 24 trials and one catch trial every four blocks. In the catch trial, the catch instruction, which instructed that participants click one of the four panels (e.g., click "smaller" and "slight experience"), appeared for 2000 ms. Thus, participants conducted 96 and four catch trials in each task.

2.4 Data analysis

2.4.1 Data exclusion

We excluded the participants' data whose performance in the catch trials was lower than 50% (2/4), and the model fitting decision coefficient R^2 was lower than 0.5 after the analysis. The exclusion criterion for catch trials was set at the chance probability of 50%, indicating that the excluded participants were not performing above chance. The exclusion criterion for model fitting was set at 0.5 for R^2 (ranging from 0 to 1.0), representing a midpoint value to ensure that the model fits the data.

2.4.2 Nonlinear models

We fit a four-parameter nonlinear (Eq. 1) to the objective discrimination accuracy and subjective awareness rating (Sandberg et al., 2011).

$$f(x) = a + (b - a)/(1 + \exp.((c - x)/d)) \quad (1)$$

Parameters *a* and *b* reflect the lower and upper boundaries of the psychometric function (i.e., $a = 0$ and $b = 1$ in the objective discrimination accuracy; $a = 1$ and $b = 4$ in the subjective awareness rating. Parameters *c* and *d* reflect the threshold and the steepness of the model slope, respectively. Larger parameter *d* indicates a more gradual model slope.

2.4.3 Statistical analysis

First, we analyzed task, age, and SOA differences in objective discrimination performance and subjective awareness rating. We conducted a three-way ANOVA on discrimination performance "sdt d'" of the signal detection theory as objective discrimination performance and PAS in each SOA. The independent variables were tasks (the lower-level color judgment task and the higher-level number magnitude judgment task) and age (5–6 year-olds, 7–9 year-olds, 10–12 year-olds, and adults), and SOA (16.7, 66.7, 116.7, 166.7, 216.7, and 266.7 ms). If there was a significant difference in the interaction, we conducted a *post hoc* analysis.

2.4.3.1 Hypotheses 1 and explorative hypotheses 1

We conducted a two-way ANOVA on the gradualness of the slope of the objective discrimination and subjective awareness. The independent variables were tasks (the lower-level color judgment task and the higher-level number magnitude judgment task) and age (5–6 year-olds, 7–9 year-olds, 10–12 year-olds, and adults). If there was a significant difference in the interaction, we conducted a *post hoc* analysis.

2.4.3.2 Hypotheses 2 and explorative hypotheses 2

We conducted a two-way ANOVA on the threshold of objective discrimination and subjective awareness. The independent variables were tasks (the lower-level color judgment task and the higher-level number magnitude judgment task) and age (5–6 year-olds, 7–9 year-olds, 10–12 year-olds, and adults). If there was a significant difference in the interaction, we conducted a *post hoc* analysis.

3 Results

3.1 Data exclusion

The final sample for analysis included 99 participants in four age groups (23 5–6 year-olds, 26 7–9 year-olds, 26 10–12 year-olds, and 24 adults (mean age = 48.13, SD = 6.89)).

We excluded one 5–6 year-old child because their catch-trial accuracy was lower than 50% (2/4 trials). We also excluded 13 participants (four 5–6 year-olds, two 7–9 year-olds, four 10–12 year-olds, and three adults) because the model fitting decision coefficient R^2 was lower than 0.5. Then, the data of 85 participants in four age groups (18 5–6 year-olds, 24 7–9 year-olds, 22 10–12 year-olds, and 21 adults) were analyzed for the hypotheses (Table 1). The number of 5–6 year-olds did not reach the initial goal of 19 but was instead held at 18 to meet the recruitment budget limit maintain consistency in recruitment methods.

3.2 Subjective awareness and objective discrimination

3.2.1 Perceptual awareness scale

Figures 3, 4 show PAS within tasks, SOAs, and between-age groups in subjective awareness. A significant main effect was found for SOA, $F_{(5,440)} = 466.2214$, $p < 0.001$, $\eta^2 = 0.516$, partial $\eta^2 = 0.841$.

Tukey's *post hoc* for the main SOA effect analysis showed that the 16.7 ms was significantly lower than the other SOA ($ps < 0.001$), 66.7 ms was significantly lower than the more SOA ($ps < 0.001$), 116.7 ms was significantly lower than 166.7 ms ($p = 0.001$), 216.7, and 266.7 ($ps < 0.001$), 166.7 ms was significantly lower than 216.7 ms ($p = 0.002$) and 266.7 ($p = 0.001$). There was no significant difference between 216.7 and 266.7 ($p = 0.964$).

3.2.2 Signal detection theory d'

Figures 5, 6 show signal detection theory d' within tasks and SOAs and between the age groups in objective discrimination. Significant main effects of task, SOA, and age were found: $F_{(1,88)} = 4.218$, $p = 0.043$, $\eta^2 = 0.003$, partial $\eta^2 = 0.046$, $F_{(5,440)} = 703.018$, $p < 0.001$, $\eta^2 = 0.641$, partial $\eta^2 = 0.889$, $F_{(3,88)} = 4.19$, $p = 0.008$, $\eta^2 = 0.016$, partial $\eta^2 = 0.125$. Moreover, an interaction was detected between task and SOA, $F_{(5,440)} = 4.054$, $p = 0.001$, $\eta^2 = 0.003$, partial $\eta^2 = 0.044$. The other effect was not significant, $ps > 0.05$.

Tukey's *post hoc* for the main effect of task analysis showed that d' in the color task was significantly larger than in the number task ($p = 0.043$). Tukey's *post hoc* for the main SOA effect analysis showed that the 16.7 ms was significantly lower than the other SOA ($ps < 0.001$), and 66.7 ms was significantly lower than the more SOA ($ps < 0.001$). The other difference was not significant. Tukey's *post hoc* for the main age effect analysis showed that 5–6 year-olds were significantly lower than 7–9 year-olds and 10–12 year-olds ($p = 0.017$ and $p = 0.012$), not adults ($p = 0.106$). The other age difference was not significant ($ps > 0.05$).

Tukey's *post hoc* for task and SOA interaction analysis showed that the color 16.7 ms was significantly smaller than the other condition ($ps < 0.001$), not the number 16.7 ms. The color 66.7 ms was significantly smaller than the colors 116.7 ($p = 0.001$), 216.7 ($p = 0.002$), and 266.7 ms ($p = 0.005$) and larger than the number 16.7 ($p < 0.001$) and 66.7 ms ($p = 0.047$). The colors 116.7, 166.7, 216.7, and 266.7 ms were significantly larger than the numbers 16.7 and 66.7 ms ($ps < 0.001$). The number 16.7 ms was significantly smaller than 66.7, 116.7, 166.7ms, 216.7, and 266.7 ms ($ps < 0.001$). The number 66.7 ms was significantly smaller than the number 116.7 ($p = 0.002$), 166.7, 216.7, and 266.7 ms ($ps < 0.001$). The other difference was not significant ($ps > 0.05$).

3.3 Hypothesis 1

3.3.1 Parameter *d* (gradualness of the slope)

3.3.1.1 Subjective awareness

Figure 7 shows parameter *d* within tasks and between the age groups in the subjective awareness. No significant main effect

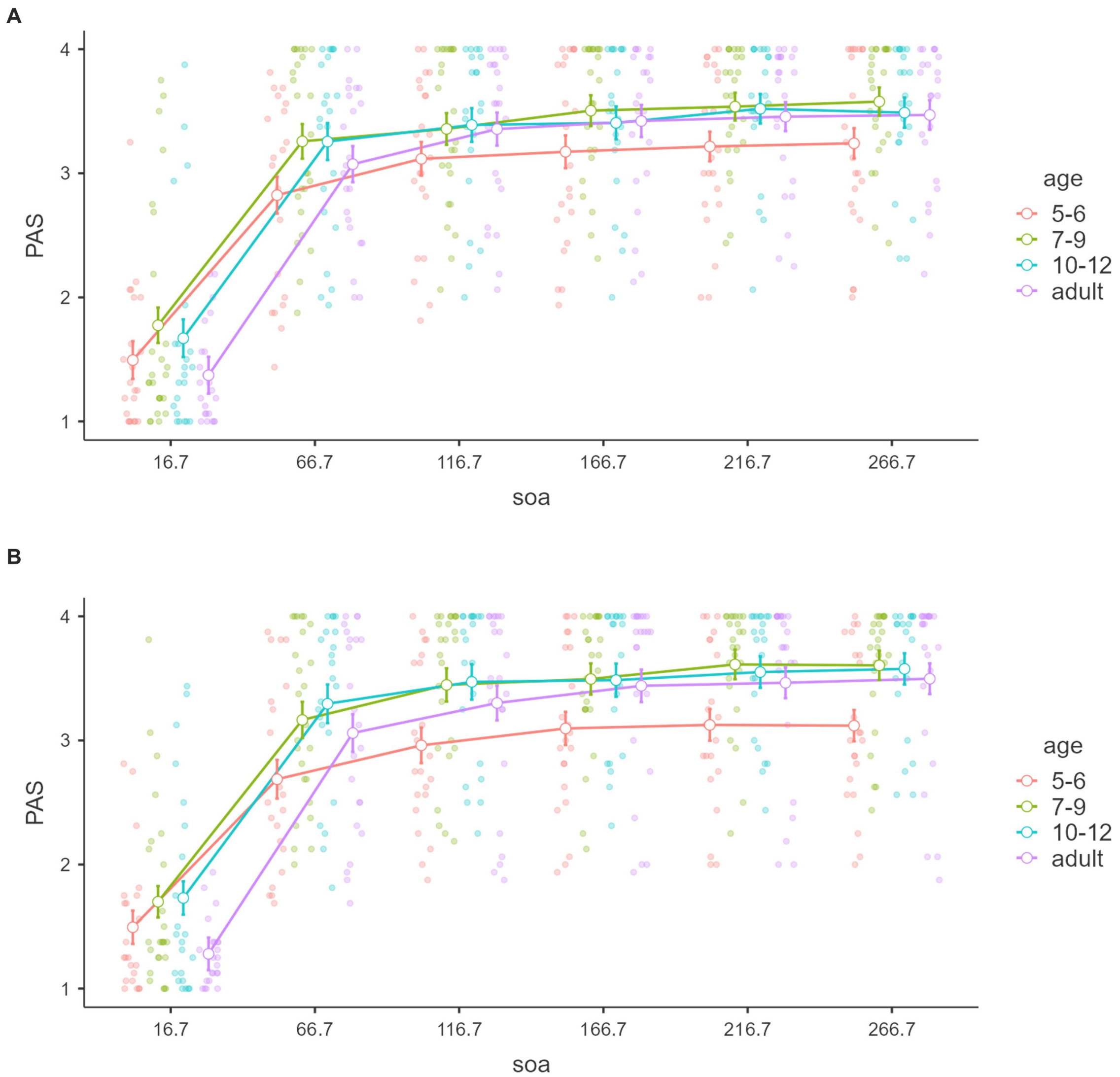

FIGURE 3
Perceptual awareness scale with tasks. Mean "PAS" of the signal detection theory (*y*-axis) of subjective awareness with SOAs (*x*-axis) and age groups (color groups) of the color and number tasks **(A,B)**. Error bars represent standard error.

was found for task and age, and interaction between task and age, $F_{(1,85)} = 0.00396$, $p = 0.950$, $\eta^2 = 0.000$, partial $\eta^2 = 0.000$, $F_{(3,85)} = 1.01$, $p = 0.394$, $\eta^2 = 0.026$, partial $\eta^2 = 0.034$, $F_{(3,85)} = 1.16685$, $p = 0.327$, $\eta^2 = 0.009$, partial $\eta^2 = 0.040$.

3.3.1.2 Objective discrimination

Figure 8 shows parameter *d* within tasks and between the age groups in the objective discrimination. A significant main effect of task was found, showing that the parameter *d* of the number task was significantly larger than that of the color task $F_{(1,81)} = 4.01$ $p = 0.049$, $\eta^2 = 0.021$, partial $\eta^2 = 0.047$, although no main effect of age and interaction between task and age, $F_{(3,81)} = 0.208$, $p = 0.891$, $\eta^2 = 0.004$, partial $\eta^2 = 0.008$, $F_{(3,81)} = 1.16$, $p = 0.332$, $\eta^2 = 0.018$, partial $\eta^2 = 0.041$.

3.4 Hypothesis 2

3.4.1 Parameter *c* (threshold)

3.4.1.1 Subjective awareness

Figure 9 shows parameter *c* within tasks and between the age groups in the subjective awareness. There was no significant main effect of age and task, and interaction between task and age, $F_{(1,85)} = 0.0608$, $p = 0.806$, $\eta^2 = 0.000$, partial $\eta^2 = 0.001$, $F_{(3,85)} = 1.56$, $p = 0.205$, $\eta^2 = 0.035$, partial $\eta^2 = 0.052$, $F_{(3,85)} = 0.1689$, $p = 0.917$, $\eta^2 = 0.002$, partial $\eta^2 = 0.006$.

3.4.1.2 Objective discrimination

Figure 10 shows parameter *c* within tasks and between the age groups in the objective discrimination. A significant main effect of age

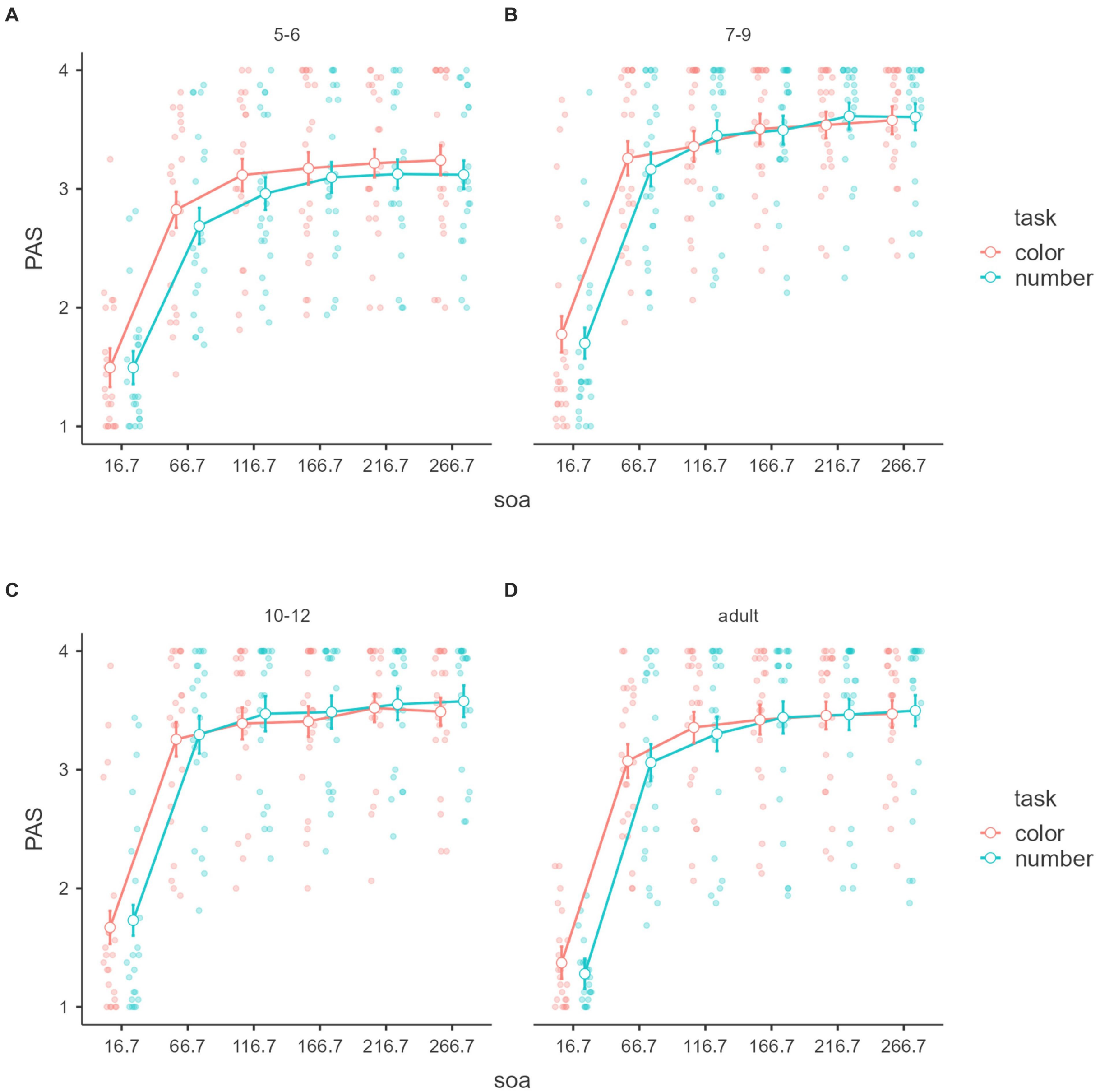

FIGURE 4
Perceptual awareness scale with age groups. Mean "PAS" of the signal detection theory (*y*-axis) of subjective awareness with SOAs (*x*-axis) and tasks (color groups) of the age groups **(A–D)**. Error bars represent standard error.

was found, $F_{(3,81)} = 3.50$, $p = 0.019$, $\eta^2 = 0.068$, partial $\eta^2 = 0.115$, although no main effect of task and interaction between task and age, $F_{(1,81)} = 1.172$, $p = 0.282$, $\eta^2 = 0.006$, partial $\eta^2 = 0.014$, $F_{(3,81)} = 0.696$, $p = 0.557$, $\eta^2 = 0.010$, partial $\eta^2 = 0.025$.

Tukey's *post hoc* analysis showed that the parameter *c* of 5–6 year-olds was significantly larger than that of 7–9 year-olds ($p = 0.046$) and 10–12 year-olds ($p = 0.028$), but not adults ($p = 0.051$). There was no age difference among 7–9 year-olds, 10–12 year-olds, and adults, $ps > 0.05$ (Figure 10).

4 Discussion

This study examined how thresholds and emergent processes of objective discrimination and subjective awareness develop with different stimulus processing levels (high-order number task and lower-order color task) using the backward masking task based on the level of processing hypothesis. We set two hypotheses and two exploratory hypotheses. Hypothesis 1 and Explorative Hypothesis 1 examined the development of the emergent process of visual consciousness. Hypothesis 1 was that the gradualness of the slope of the objective discrimination and the subjective awareness for lower-level color judgment task is larger than that for higher-level number magnitude judgment task in 5–12 year-olds and adults. Hypothesis 2 and Explorative Hypothesis 2 examined the development of the threshold of visual consciousness. Hypothesis 2 was that the thresholds of objective discrimination and subjective awareness for lower-level color and higher-level number magnitude judgment tasks decreased with age.

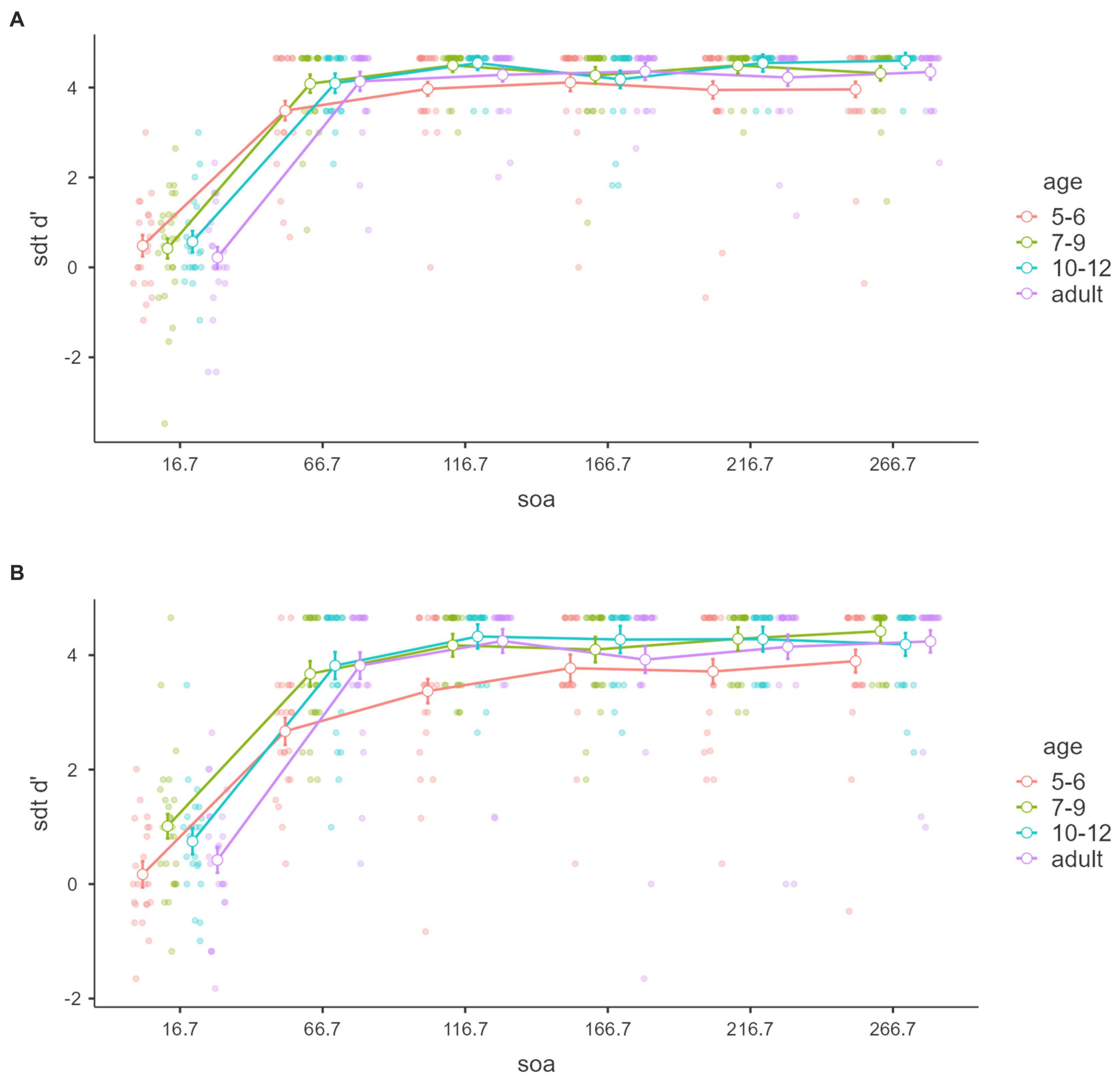

FIGURE 5
Discrimination ability sdt d' with tasks. Mean "sdt d'" of the signal detection theory (*y*-axis) of objective discrimination with SOAs (*x*-axis) and age groups (color groups) of the color and number tasks **(A,B)**. Error bars represent standard error.

Hypothesis 1 was not supported. No significant main effect of task and interaction between task and age in the gradualness of the slope of the subjective awareness was detected. Moreover, a significant main effect of task was found, although no interaction was detected between task and age in the gradualness of the slope of the objective discrimination. The slope of the number task was significantly more gradual than the color task's. The results suggest that there are no task or age differences in the emergent process of subjective awareness but task differences in the emergent process of objective discrimination, depending on the level of processing. Furthermore, the results suggest that number magnitude judgments, which was the higher-order processing, occur more gradually than color judgments, which was the lower-order processing.

The results did not support the level of processing hypothesis and were inconsistent with Windey et al. (2013). Contrary to our results, Windey et al. (2013) showed that the slope of the subjective awareness and objective discrimination in the lower-order color judgment task was more gradual than in the higher-order number magnitude judgment task. However, although there was no significant interaction between age and task differences, only our adult results may be consistent with Windey et al. (2013). The mean slope gradualness of subjective awareness and objective discrimination was larger for the color task than for the number task in only adults.

The differences in SOA and the age of the participants may explain why the results of the present study did not replicate the results of the previous studies. First, the SOA in the present study was set larger than in previous studies to allow 5–6 year-olds to perform the task. In the previous study, SOAs were set at 10 ms intervals at less than 100 ms (Windey et al., 2013). The larger SOAs and SOA intervals may have resulted in lower task difficulty and less difference between tasks. Second, the results for adults showed a similar trend to Windey et al. (2013), but the results for children showed the opposite pattern. The inclusion of children can lead to different results between the previous study and the present study.

Hypothesis 2 was partially supported. No significant main effect of age and task or interaction between task and age in the threshold of

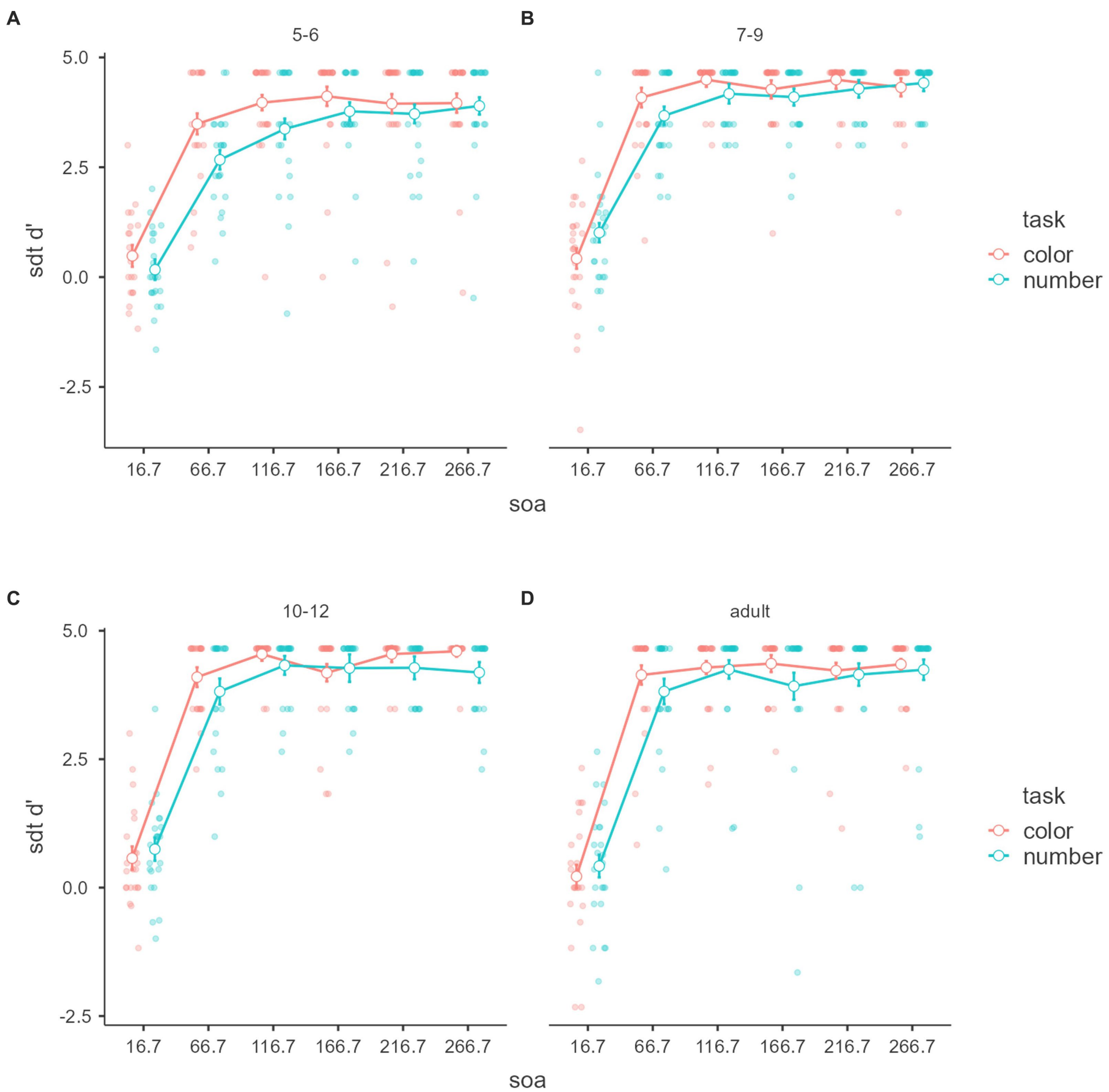

FIGURE 6
Discrimination ability sdt d' with age groups. Mean "sdt d'" of the signal detection theory (*y*-axis) of objective discrimination with SOAs (*x*-axis) and tasks (color groups) of the age groups **(A–D)**. Error bars represent standard error.

subjective awareness was found. Moreover, a significant main effect of age was detected, although no main effect of task and interaction between task and age in the objective discrimination was found. The threshold of 5–6 year-olds was significantly larger than that of 7–9 year-olds and 10–12 year-olds, and the other age difference and interaction between task and age were not. The results suggest that the thresholds of objective discrimination become smaller between ages 5–6 and 7–9, but there are no age differences after that. Surprisingly, there were no significant differences between the 5–6 years-old group and adults, which may be explained through the older age of the adult participants in the study. Compared to the face-to-face experiment, the adults taking part online tended to be older. Therefore, it is possible that the performance of the adult participants was lower than in previous studies. The thresholds of subjective awareness and objective discrimination were inconsistent with Watanabe and Moriguchi (2023), which showed that the threshold of objective discrimination and subjective awareness in young children (5–6 year-olds) was larger than that in adults in figure stimuli.

Moreover, the lack of difference between tasks was consistent with the findings of Windey et al. (2013), which showed no task difference in the threshold. The threshold of subjective awareness is developed at age five and is equivalent to that of adults. In contrast, the threshold of objective discrimination is developed at age nine and is equivalent to that of adults.

A possible reason for the discrepancy between subjective awareness and objective discrimination is that their neural bases may

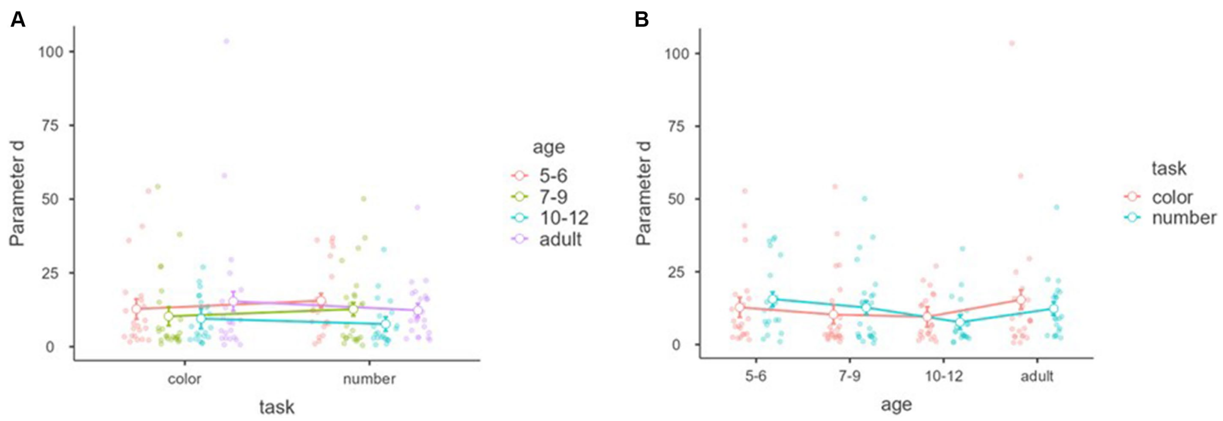

FIGURE 7
Parameter *d* (gradualness of the slope) with tasks and age groups of the subjective awareness. Mean parameters "d (gradualness of the slope)" of the nonlinear model (*y*-axis) of subjective awareness with tasks (*x*-axis) and age groups (color groups) **(A)**. Mean parameter of "d (gradualness of the slope)" of the nonlinear model (*y*-axis) of subjective awareness with age groups (*x*-axis) and tasks (color groups) **(B)**. Error bars represent standard error.

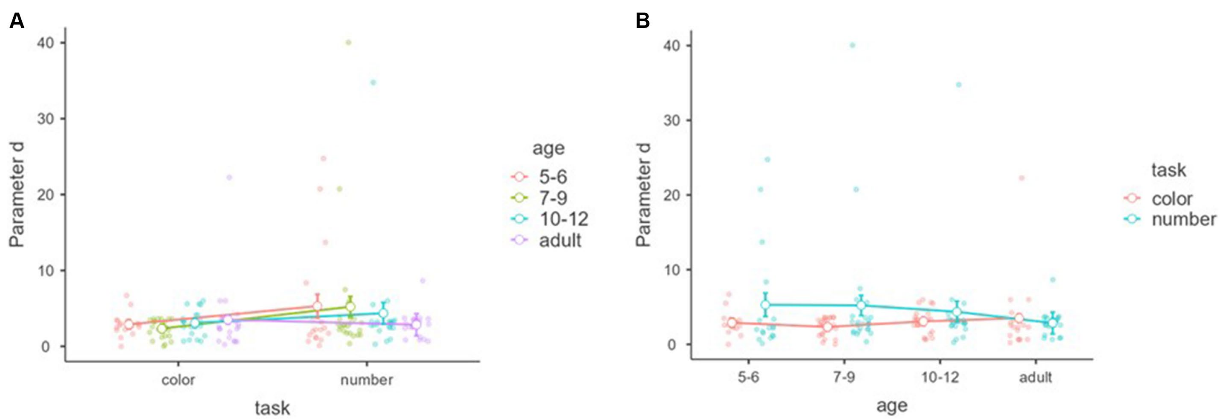

FIGURE 8
Parameter *d* (gradualness of the slope) with tasks and age groups of the objective discrimination. Mean parameters "d (gradualness of the slope)" of the nonlinear model (*y*-axis) of objective discrimination with tasks (*x*-axis) and age groups (color groups) **(A)**. Mean parameter of "d (gradualness of the slope)" of the nonlinear model (*y*-axis) of objective discrimination with age groups (*x*-axis) and tasks (color groups) **(B)**. Error bars represent standard error.

differ. The frontal lobes and frontal–parietal network may be related to objective discrimination more than to subjective awareness. Young children's (aged 5–6 years old) frontal lobes and frontal–parietal networks are more immature than those of adults but the posterior perceptual area are similar to adults (Huttenlocher, 1990; Adleman et al., 2002; Nagy et al., 2004; Tanaka et al., 2012; Mehnert et al., 2013; Moriguchi and Hiraki, 2013), which may result in developmental difference between objective discrimination and subjective awareness. However, subjective awareness is also based on the frontal–parietal network and the frontal lobes (Hatamimajoumerd et al., 2022), with several previous studies showing the relationship of objective discrimination and subjective awareness to the frontal lobes and the frontal–parietal network (Dehaene et al., 2003; Pins and Ffytche, 2003; Lamme, 2006; Del Cul et al., 2007, 2009; Dehaene and Changeux, 2011; van Vugt et al., 2018). Therefore, further research is needed to clarify the differences in the neural basis of objective discrimination and subjective awareness.

Our findings and previous research suggest that level of processing and age have little or no influence on the emergent process of subjective awareness and objective discrimination. That there were no age differences in the emergent process of subjective awareness and objective discrimination in the color and number magnitude tasks was congruent with Watanabe and Moriguchi (2023), which showed no age differences between young children and adults in the figure stimuli, which was the lower-order processing. These findings suggest that the emergent processes of visual consciousness are similar to adults by age 5, regardless of the level of processing.

We examined discrimination ability and response bias. Significant main tasks, SOAs, age differences, and the interaction between tasks and SOAs were found. The *post hoc* analysis of SOA showed a difference between 16.7 ms and 66.7 ms or more and that 66.7 ms did

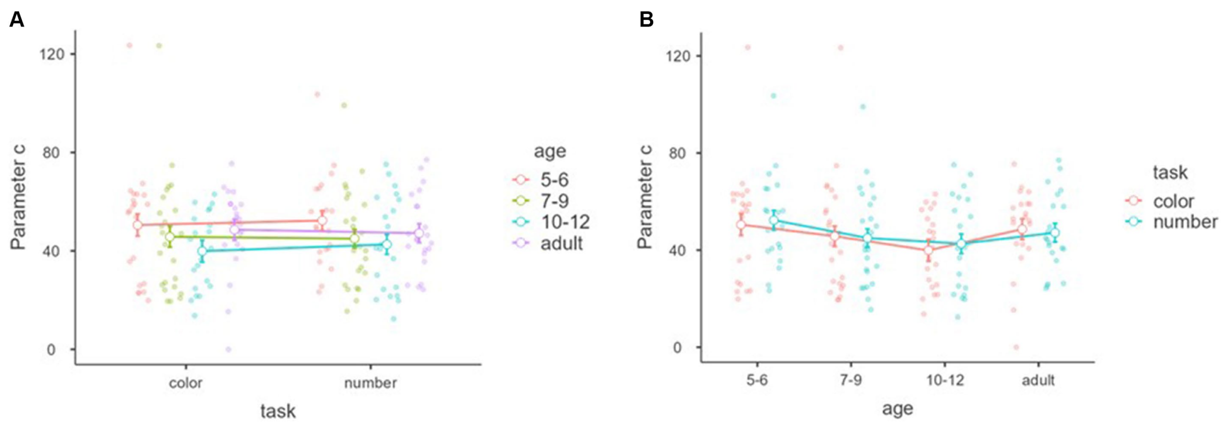

FIGURE 9
Parameter *c* (threshold) with tasks and age groups of the subjective awareness. Mean parameters "c (threshold)" of the nonlinear model (*y*-axis) of subjective awareness with tasks (*x*-axis) and age groups (color groups) **(A)**. Mean parameter of "c (threshold)" of the nonlinear model (*y*-axis) of subjective awareness with age groups (*x*-axis) and tasks (color groups) **(B)**. Error bars represent standard error.

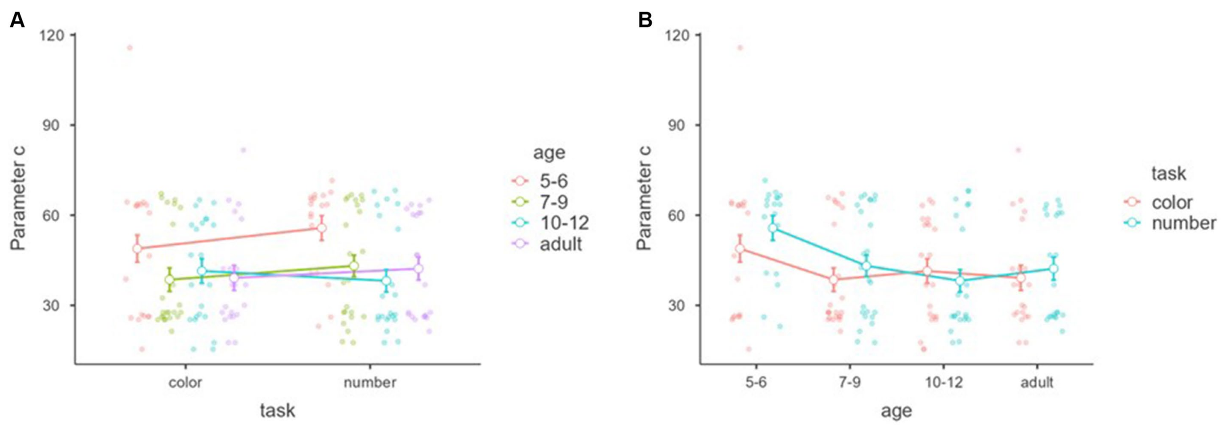

FIGURE 10
Parameter *c* (threshold) with tasks and age groups of the objective discrimination. Mean parameters "c (threshold)" of the nonlinear model (*y*-axis) of objective discrimination with tasks (*x*-axis) and age groups (color groups) **(A)**. Mean parameter of "c (threshold)" of the nonlinear model (*y*-axis) of objective discrimination with age groups (*x*-axis) and tasks (color groups) **(B)**. Error bars represent standard error.

not differ from 116.7 ms or more. The result suggests a threshold between 16.7 ms and 66.7 ms. The *post hoc* analysis of age groups showed that 5–6 year-olds were lower than 7–9 year-olds and 10–12 year-olds, not adults. The result suggests that discrimination performance increases between ages 5 and 9 and maintains between ages 9–12 and adults. These results were consistent with the results of Hypothesis 2.

We showed two new findings. First, no age differences were found in the emergent processes of objective discrimination and subjective awareness. Second, the thresholds of objective discrimination for color and number stimuli decrease developmentally from 5–6 year-olds to 7–12 year-olds. The present results contribute to the knowledge of the development of visual consciousness, which was largely lacking; 5–6 year-olds, like older children and adults, experience subjective awareness, but their objective discrimination of stimuli is undeveloped. Thus, the results suggest that the difference between subjective awareness and objective discrimination decreases with age.

Furthermore, our study demonstrated that the level of processing hypothesis may be consistent for adults, but not for children. Previous studies study, the have indicated a more gradual model for the lower-order processing condition for color judgments compared to the model for the higher-order condition for number magnitude judgments. However, our results showed the opposite for children. These findings can be interpreted in two ways. The first interpretation is that the level of processing hypothesis holds, but its effects may differ across developmental stages. Watanabe and Moriguchi (2023) examined the objective discrimination and subjective awareness of form stimuli in young children and adults by conducting a face-to-face experiment. Their results showed that for both objective discrimination and subjective awareness, the thresholds in young children were higher than those in adults. Thus, children's responses may be easily influenced by stimuli and experimental methods. The second interpretation is that the level of processing hypothesis itself is suspect. However, we choose to support the first interpretation due to the accumulation of studies that confirm the level of processing hypothesis (Windey et al., 2013; Anzulewicz et al., 2015; Windey and Cleeremans, 2015; Binder et al., 2017; Derda et al., 2019; Jimenez et al., 2021; Thiruvasagam and Srinivasan, 2021). We, however, suggest that the level of processing hypothesis needs to be reconsidered, including its developmental aspects.

This study has some limitations. First, participants conducted the color and number tasks as an online experiment. Differences in the experimental environment may have caused inconsistencies between our results and previous research. In this study, we controlled the experimental environment (e.g., bright and quiet room) and computer setting (e.g., 13–16 inch, 60 Hz, 800*600) as much as possible to reduce between-participants differences. Moreover, in the experiment with 5–6 year-olds, we connected Zoom with the participants to confirm that the parents accurately answered the children's responses.

Second, a difference in task difficulty may have existed between the color and number tasks. The stimuli and SOA were set the same for the color and number tasks, and the questions and answers were changed. However, the performance of the color task was greater than that of the number task. This difference in performance may have made it easier to make the slope of the number task more gradual than the color task. In our tasks, we increased the SOAs and the intervals of the SOAs of the tasks compared to previous studies to allow children to respond. Although there was no significant interaction between age and task or age and SOA, the results may have reflected differences in performance on the color and number tasks in children. Future studies should further examine the effects of SOA and age in greater detail.

Third, participants responded to how clearly they could see the numbers in the color and number tasks as a subjective awareness. The response criteria may differ between and within participants. For example, participants may have responded with subjective awareness of the "color" of the number, the "size" of the number, and the number "itself." In the future, direct questions should be asked about the color of the numbers and size, and differences depending on the question should be examined. The PAS is a scale for rating perceptual awareness (e.g., the color of a number) but not for rating cognitive discrimination (e.g., the size of a number). Thus, there may be limitations in directly comparing the two. To examine the level of processing hypothesis, separating lower-order and higher-order processing in perception and lower-order and higher-order processing in cognition may be necessary.

Finally, we considered only the developmental aspects of the level of processing hypothesis, without direct comparisons between the all-or-none position and the gradual position of the emergent consciousness process. One reason is that the parameter d', which we used as a measure of this process, was a relatively comparative measure of higher-order and lower-order processing conditions. It is necessary to determine a value for parameter d' that would support an all-or-none emergent process and a gradual emergent process, respectively.

5 Conclusion

Despite these limitations, this study is the first to examine the development of emergent processes and the threshold of visual consciousness in childhood with the level of processing. We examined the developmental aspects of the level of processing hypothesis, the most recent theory on the emergent process of visual consciousness and thresholds. The results showed that in objective discrimination, thresholds were higher in 5-6 year-olds than in older children in both the higher- and lower-order processing tasks, but regarding subjective awareness, no age differences were shown between the two tasks. Moreover, in objective discrimination, the emergent process in the higher-order task was more gradual than that in the lower-order task, but there was no task difference in subjective awareness. In conclusion, our study supports the level of processing hypothesis, but notes that its effects may differ across developmental stages. This paper contributes new knowledge to developmental and consciousness research by revealing the development of visual consciousness.

Data availability statement

The datasets presented in this study can be found in online repositories (https://osf.io/cqezv/).

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Unit for Advanced Studies of the Human Mind, Kyoto University (No. 2-P-11). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

RW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. YM: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. YM is supported by a Grant-in-Aid for Transformative Research Areas Grant-in-Aid for Transformative Research Areas (20H05710, 23H04830) from the Japan Society for the Promotion of Science. RW is supported by the Japan Society for the Promotion of Science (Grant Number: 22J15748).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adleman, N. E., Menon, V., Blasey, C. M., White, C. D., Warsofsky, I. S., Glover, G. H., et al. (2002). A developmental fMRI study of the Stroop color-word task. *NeuroImage* 16, 61–75. doi: 10.1006/nimg.2001.1046

Anzulewicz, A., Asanowicz, D., Windey, B., Paulewicz, B., Wierzchoń, M., and Cleeremans, A. (2015). Does level of processing affect the transition from unconscious to conscious perception? *Conscious. Cogn.* 36, 1–11. doi: 10.1016/j.concog.2015.05.004

Asplund, C. L., Fougnie, D., Zughni, S., Martin, J. W., and Marois, R. (2014). The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychol. Sci.* 25, 824–831. doi: 10.1177/0956797613513810

Baars, B.J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience, Progress in brain research, *150*, 45–53), (Ed.) S. Laureys, Elsevier.

Bayne, T., Frohlich, J., Cusack, R., Moser, J., and Naci, L. (2023). Consciousness in the cradle: on the emergence of infant experience. *Trends Cogn. Sci.* 27, 1135–1149. doi: 10.1016/j.tics.2023.08.018

Binder, M., Gociewicz, K., Windey, B., Koculak, M., Finc, K., Nikadon, J., et al. (2017). The levels of perceptual processing and the neural correlates of increasing subjective visibility. *Conscious. Cogn.* 55, 106–125. doi: 10.1016/j.concog.2017.07.010

Breitmeyer, Bruno, and Ogmen, Haluk. (1984). Visual masking: time slices through conscious and unconscious vision. Visual Masking. Oxford University Press

Christensen, M. S., Ramsøy, T. Z., Lund, T. E., Madsen, K. H., and Rowe, J. B. (2006). An fMRI study of the neural correlates of graded visual perception. *NeuroImage* 31, 1711–1725. doi: 10.1016/j.neuroimage.2006.02.023

Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018

Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37. doi: 10.1016/s0010-0277(00)00123-2

Dehaene, S., Sergen, C., and Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. USA* 100, 8520–8525. doi: 10.1073/pnas.1332574100

Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* 5:e260. doi: 10.1371/journal.pbio.0050260

Del Cul, A., Dehaene, S., and Leboyer, M. (2006). Preserved subliminal processing and impaired conscious access in schizophrenia. *Arch. Gen. Psychiatry* 63, 1313–1323. doi: 10.1001/archpsyc.63.12.1313

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., and Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 132, 2531–2540. doi: 10.1093/brain/awp111

Derda, M., Koculak, M., Windey, B., Gociewicz, K., Wierzchoń, M., Cleeremans, A., et al. (2019). The role of levels of processing in disentangling the ERP signatures of conscious visual processing. *Conscious. Cogn.* 73:102767. doi: 10.1016/j.concog.2019.102767

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146

Gelskov, S. V., and Sid, K. (2010). "Psychophysical thresholds of face visibility during infancy." *Cogn.* 114, 285–92. doi: 10.1016/j.cognition.2009.09.012

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nat. Neurosci.* 2, 861–863. doi: 10.1038/13158

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Deanna Greenstein, A., Vaituzis, C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. USA* 101, 8174–8179. doi: 10.1073/pnas.0402680101

Hatamimajoumerd, E., Ratan Murty, N. A., Pitts, M., and Cohen, M. A. (2022). Decoding perceptual awareness across the brain with a no-report fMRI masking paradigm. *Curr. Biol.* 32, 4139–49.e4. doi: 10.1016/j.cub.2022.07.068

Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804. doi: 10.1016/s0896-6273(02)01091-7

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia* 28, 517–527. doi: 10.1016/0028-3932(90)90031-I

Huttenlocher, P. R., de Courten, C., Garey, L. J., and Van der Loos, H. (1982). Synaptogenesis in human visual cortex—evidence for synapse elimination during normal development. *Neurosci. Lett.* 33, 247–252. doi: 10.1016/0304-3940(82)90379-2

Jimenez, M., Hinojosa, J. A., and Montoro, P. R. (2020). Visual awareness and the levels of processing hypothesis: a critical review. *Conscious. Cogn.* 85:103022. doi: 10.1016/j.concog.2020.103022

Jimenez, M., Poch, C., Villalba-García, C., Sabater, L., Hinojosa, J. A., Montoro, P. R., et al. (2021). The level of processing modulates visual awareness: evidence from behavioral and electrophysiological measures. *J. Cogn. Neurosci.* 33, 1295–1310. doi: 10.1162/jocn_a_01712

Jimenez, M., Villalba-García, C., Luna, D., Hinojosa, J. A., and Montoro, P. R. (2019). The nature of visual awareness at stimulus energy and feature levels: a backward masking study. *Atten. Percept. Psychophys.* 81, 1926–1943. doi: 10.3758/s13414-019-01732-5

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22

Koch, C., and Preuschoff, K. (2007). Betting the house on consciousness. *Nat. Neurosci.* 10, 140–141. doi: 10.1038/nn0207-140

Kouider, S., Carsten, S., Sofie, V. G., Leonardo, S. B., Michel, D., and Vincent de, G, et al. (2013). "A neural marker of perceptual consciousness in infants." *Science* 340, 376–80. doi: 10.1126/science.1232509

Kwon, H., Reiss, A. L., and Menon, V. (2002). Neural basis of protracted developmental changes in visuo-spatial working memory. *Proc. Natl. Acad. Sci. USA* 99, 13336–13341. doi: 10.1073/pnas.162486399

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. doi: 10.1016/j.tics.2006.09.001

LeBlanc, R. S., Muise, J. G., and Blanchard, L. (1992). Backward visual masking in children and adolescents: sensory transmission, accrual rate and asymptotic performance. *J. Exp. Child Psychol.* 53, 105–114. doi: 10.1016/0022-0965(92)90043-6

Macchi, M., Rossi, L. N., Cortinovis, I., Menegazzo, L., Burri, S. M., Piller, M., et al. (2003). Development of visual perception and attention, assessed by backward masking and application in children with epilepsy. *Dev. Med. Child Neurol.* 45, 562–567. doi: 10.1017/s0012162203001026

Macmillan, N. A., and Douglas Creelman, C. (2004). Detection theory: a User's guide. *2nd* Edn: Psychology Press.

Mashour, G. A., Roelfsema, P., Changeux, J.-P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron* 105, 776–798. doi: 10.1016/j.neuron.2020.01.026

Mehnert, J., Akhrif, A., Telkemeyer, S., Rossi, S., Schmitz, C. H., Steinbrink, J., et al. (2013). Developmental changes in brain activation and functional connectivity during response inhibition in the early childhood brain. *Brain and Development* 35, 894–904. doi: 10.1016/j.braindev.2012.11.006

Moriguchi, Y., and Hiraki, K. (2013). Prefrontal cortex and executive function in young children: a review of NIRS studies. *Front. Hum. Neurosci.* 7:867. doi: 10.3389/fnhum.2013.00867

Nagy, Z., Westerberg, H., and Klingberg, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *J. Cogn. Neurosci.* 16, 1227–1233. doi: 10.1162/0898929041920441

Nakashima, Y., Kanazawa, S., and Yamaguchi, M. K. (2021). Perception of invisible masked objects in early infancy. *Proc. Natl. Acad. Sci. USA* 118:40118. doi: 10.1073/pnas.2103040118

Overgaard, M., Rote, J., Mouridsen, K., and Ramsøy, T. Z. (2006). Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Conscious. Cogn.* 15, 700–708. doi: 10.1016/j.concog.2006.04.002

Prager, E. O., Maria, D. S., and Stephanie, M. C. (2016). "Executive function and magnitude skills in preschool children." *J. Exp. Child Psychol.* 147. doi: 10.1016/j.jecp.2016.01.002

Pins, D., and Ffytche, D. (2003). The neural correlates of conscious vision. *Cereb. Cortex* 13, 461–474. doi: 10.1093/cercor/13.5.461

Pretorius, H., Tredoux, C., and Malcolm-Smith, S. (2016). Subjective awareness scale length influences the prevalence, not the presence, of graded conscious states. *Conscious. Cogn.* 45, 47–59. doi: 10.1016/j.concog.2016.08.007

Ramsøy, T. Z., and Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenol. Cogn. Sci.* 3, 1–23. doi: 10.1023/B:PHEN.0000041900.30172.e8

Sandberg, K., Bibby, B. M., Timmermans, B., Cleeremans, A., and Overgaard, M. (2011). Measuring consciousness: task accuracy and awareness as sigmoid functions of stimulus duration. *Conscious. Cogn.* 20, 1659–1675. doi: 10.1016/j.concog.2011.09.002

Sandberg, K., Timmermans, B., Overgaard, M., and Cleeremans, A. (2010). Measuring consciousness: is one measure better than the other? *Conscious. Cogn.* 19, 1069–1078. doi: 10.1016/j.concog.2009.12.013

Sekar, K., Findley, W. M., Poeppel, D., and Llinás, R. R. (2013). Cortical response tracking the conscious experience of threshold duration visual stimuli indicates visual perception is all or none. *Proc. Natl. Acad. Sci. USA* 110, 5642–5647. doi: 10.1073/pnas.1302229110

Sergent, C., and Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.* 15, 720–728. doi: 10.1111/j.0956-7976.2004.00748.x

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., and Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends Cogn. Sci.* 12, 314–321. doi: 10.1016/j.tics.2008.04.008

Seth, A. K., and Tim, B. (2022). "Theories of consciousness." *Nature reviews. Neurosci.* 23, 439–52. doi: 10.1038/s41583-022-00587-4

Tanaka, C., Matsui, M., Uematsu, A., Noguchi, K., and Miyawaki, T. (2012). Developmental trajectories of the fronto-temporal lobes from infancy to early adulthood in healthy individuals. *Dev. Neurosci.* 34, 477–487. doi: 10.1159/000345152

Thiruvasagam, S., and Srinivasan, N. (2021). Gradedness of visual awareness depends on attentional scope: global perception is more graded than local perception. *Conscious. Cogn.* 94:103174. doi: 10.1016/j.concog.2021.103174

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., et al. (2018). The threshold for conscious report: signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542. doi: 10.1126/science.aar7186

Watanabe, R., and Moriguchi, Y. (2023). Young children's subjective and objective thresholds and emergent processes of visual consciousness using a backward masking task. *Conscious. Cogn.* 116:103605. doi: 10.1016/j.concog.2023.103605

Welsandt, R. F., Zupnick, J. J., and Meyer, P. A. (1973). Age effects in backward visual masking (Crawford paradigm). *J. Exp. Child Psychol.* 15, 454–461. doi: 10.1016/0022-0965(73)90095-7

Windey, B., and Cleeremans, A. (2015). Consciousness as a graded and an all-or-none phenomenon: a conceptual analysis. *Conscious. Cogn.* 35, 185–191. doi: 10.1016/j.concog.2015.03.002

Windey, B., Gevers, W., and Cleeremans, A. (2013). Subjective visibility depends on level of processing. *Cognition* 129, 404–409. doi: 10.1016/j.cognition.2013.07.012

frontiers | Frontiers in Psychology

TYPE Methods
PUBLISHED 18 September 2024
DOI 10.3389/fpsyg.2024.1393913

OPEN ACCESS

EDITED BY
Luca Simione,
UNINT - Università degli studi Internazionali di Roma, Italy

REVIEWED BY
Lampros Perogamvros,
University of Geneva, Switzerland
Javier Villanueva-Valle,
National Autonomous University of Mexico, Mexico

*CORRESPONDENCE
Mariza van Wyk
✉ mariza.v.w@gmail.com

RECEIVED 29 February 2024
ACCEPTED 02 September 2024
PUBLISHED 18 September 2024

CITATION
van Wyk M, Solms M and Lipinska G (2024) A novel method for objectively classifying sequential emotion within dreams: a proof-of-concept pilot study.
Front. Psychol. 15:1393913.
doi: 10.3389/fpsyg.2024.1393913

A novel method for objectively classifying sequential emotion within dreams: a proof-of-concept pilot study

Mariza van Wyk*, Mark Solms and Gosia Lipinska

UCT Sleep Sciences, Department of Psychology, University of Cape Town, Cape Town, South Africa

Traditionally, emotions in dreams have been assessed using subjective ratings by human raters (e.g., external raters or dreamers themselves). These methods have extensive support and utility in dream science, yet they have certain innate limitations due to the subjective nature of the rating methodologies. Attempting to circumvent several of these limitations, we aimed to develop a novel method for objectively classifying and quantifying sequential (word-for-word) emotion within a dream report. We investigated whether sentiment analysis, a branch of natural language processing, could be used to generate continuous positive and negative valence ratings across a dream. In this pilot, proof-of-concept study, we used 14 dream reports collected upon awakening following overnight polysomnography. We also collected pre- and post-sleep affective data and personality metrics. Our objectives included demonstrating that (1) valence ratings derived from sentiment analysis (Valence Aware Dictionary for sEntiment Reasoning [VADER]) could be used to visualize (plot) positive and negative emotion fluctuations within a dream, (2) how the visual properties of emotion fluctuations within a dream (peaks and troughs, area under the curve) can be used to generate novel "emotion indicators" as proxies for emotion regulation throughout a dream, and (3) these emotion indicators correlate with sleep, affective, and personality variables known to be associated with dreaming and emotion regulation. We describe 6 novel, objective dream emotion indicators: Total number of Peaks, total number of Troughs, Positive, Negative, and Overall Emotion Intensity (composites from an "area under the curve" method using the trapezoid rule applied to the peaks and troughs), and the Emotion Gradient (a polynomial trendline fitted to the emotion fluctuations in the dream chart). The latter signifies the overall direction of sequential emotion changes within a dream. Results also showed that 5/6 emotion indicators correlated significantly with at least one existing sleep, affective, or personality variable known to be associated with dreaming and emotion regulation. We propose that the novel emotion indicators potentially serve as proxies for emotion regulation processes unfolding within a dream. These preliminary findings provide a methodological foundation for future studies to test and refine the method in larger and more diverse samples.

KEYWORDS

emotion regulation, emotions in dreams sentiment analysis, dream emotional intensity, objective dream ratings, classifying dream emotions, trapezoid rule, area under the curve

1 Introduction

Traditionally, content and emotion in dreams have been assessed through three main methods with varying degrees of subjectivity associated with each method: (1) self-ratings by the dreamer themselves, (2) ratings by a blind external judge, (3) external ratings based on a classification system developed by Hall and van de Castle (1966; Röver and Schredl, 2017; Zadra and Domhoff, 2011). These methods have widespread support in dream science, yet it has obvious limitations associated with the subjective nature of these ratings (Röver and Schredl, 2017; Schredl and Doll, 1998; Schredl, 2010a; Sikka et al., 2014). Some of the limitations include significant disparities across these methods with regard to the presence and/or intensity of emotions in dreams, over- and/or under-estimation of positive and negative emotions, incongruencies in ratings of explicitly mentioned emotions versus implicit/experienced emotions (the latter being more likely to be identified by self-raters), as well as discrepancies in the emotion/content classification systems/scales used.

Although there are many advantages to obtaining subjective data, objective approaches provide advantages of their own. Sentiment analysis, a semantic analytic technique and a branch of natural language processing (NLP), is a powerful tool used to analyze the emotional properties of texts. It works by classifying and quantifying positive, negative, and neutral emotions detected in texts. Depending on the method used, sentiment analysis assigns a categorical or continuous valence rating to the overall text, sentences in the text, or individual words in the text. Encouragingly, sentiment analysis has recently been successfully employed to analyze emotion in dream reports (Yu, 2022). More specifically, one of the methods used in the study by Yu (2022) is called the Valence Aware Dictionary for sEntiment Reasoning (VADER)—an automated software program that analyses textual data based on an established lexicon, coupled with annotated lexical features. As an objective approach to rating emotions in dreams, the VADER program was described by the authors as a reliable and effective sentiment analysis tool. It is proposed to complement existing methods in the following ways: (1) detecting subtle, indirect affective feelings by considering the tone of every word used, (2) circumventing inter-rater discrepancies due to its automated nature, and (3) generating continuous (as opposed to categorical) classifications of emotion/valence. This yields a metric that captures the emotional intensity of dream experiences in an objective manner. Therefore, the three main properties of the VADER program enable one to objectively track sequential (word-for-word) emotion intensity within a dream by generating continuous positive and negative valence ratings. This method circumvents several of the limitations/biases inherent to strictly subjective approaches to rating emotions in dreams. Additionally, this method also overcomes some of the limitations present in other forms of semantic analysis, such as the dictionary-based approach, which determines the word frequency count for specific categories, such as positive or negative words. This approach is limited in that it does not take context into account (for example, what was said before or after a particular counted word) and does not measure emotion continuously, as it progresses through the dream (Elce et al., 2021).

1.1 Rationale and aims

Based on current and ongoing limitations evident in dream rating, we had the following main aim: to develop an objective method to classify and quantify emotion in a sequential (word-for-word) manner within individual dream reports. With this exploratory approach we aimed to provide preliminary evidence for the proposition that sequential emotion fluctuations could serve as objective indicators of emotional processes unfolding within a dream. We set out to achieve this by using sentiment analysis to generate sequential emotional valence ratings throughout a dream. We then plotted and visualized these valence ratings and applied specific statistical and mathematical techniques to the peak and trough characteristics of the chart to create a range of "emotion indicators." To validate whether these indicators are effective in quantifying dream emotion, we correlated them with external variables that are known to be related to dreaming and/or emotion regulation. These external variables fell into a number of categories that included polysomnography-derived parameters (e.g., number of awakenings and sleep stage distribution, classified as a "sleep" variable), along with pre- and post-sleep affective data (which we classified as "state" variables), and, finally, personality characteristics (which we classified as "trait" variables). We aimed to link, in a preliminary fashion, these external variables with the dream emotion indicators. In summary, we had the following objectives:

1) We aimed to show that sentiment analysis could be used to classify, quantify, and visualize positive and negative emotion fluctuations within a dream;
2) The visual properties of emotion fluctuations within a dream (peaks and troughs) can be used to generate novel "emotion indicators"; and:
3) The various emotion indicators are correlated with sleep, state, and trait variables.

Here, we present our new method, and accompanying preliminary findings, as a proof-of-concept (POC) study—a first step in developing a novel method for objectively classifying sequential emotion within a dream. We propose that the emotion indicators could serve as potential proxies for emotion regulatory processes during dreaming. With these preliminary results we aim to provide a methodological foundation for future studies to test and refine the method employed in the current study in larger and more diverse samples.

2 Methods

2.1 Participants

This pilot study contained dream reports from 14 healthy university students. The sample used in this study is a subset of a sample recruited for a larger study which investigated sleep architecture, memory, and affect in relation to dream recall frequency (DRF). Screening for the larger study occurred in two phases: the first phase consisted of an online survey, and the second phase consisted of a face-to-face clinical interview. During the online screening phase, participants completed questions related to demographics, medical and psychiatric history, sleep quality, unusual sleep experiences (e.g., sleep paralysis), medication use, and DRF.

Participants were excluded from participation if they met any of the following criteria: (a) were below the age of 20 or over the age of 40, (b) had any medical or neurological condition that could affect the outcomes of the study, (c) had a current and/or past history of any sleep disorder, (d) had a current and/or past history of any psychiatric disorder, (e) used sleeping pills, sedative medication, psychoactive medication or any other medication that might affect sleep or dreaming, (f) had a past and/or current history of alcohol or substance abuse or dependence, (g) were currently pregnant, or (h) had below-average cognitive ability. Literature shows that these factors have an independent relationship with sleep and/or dreaming and can affect study outcomes (Lee, 1998; Blackman, 2000; Irwin et al., 2000; Nielsen and Stenstrom, 2005; Schredl, 2010b; Pagel, 2010; Schredl et al., 2013; Skancke et al., 2014).

All eligible participants underwent overnight polysomnography (PSG) on two non-consecutive nights. The qualifying/inclusion criterion for the current study was participants reporting a dream upon awakening following the second (testing) night at the sleep laboratory. This was to ensure that we would be able to correlate dream emotion indicators with the external, validated (1) sleep (PSG-derived) variables, (2) affective (state), and (3) personality (trait) variables.

With regard to the demographic characteristics of the final sample for the current study, 57% of participants identified as women and 43% of the sample identified as men. The average age was 21.93 with a range of 20–27.

2.2 Study procedure

All study procedures complied with the Declaration of Helsinki. Ethical clearance from the relevant ethical review boards of both the Department of Psychology and the Faculty of Humanities was obtained prior to recruitment commencing. All participants included in this study completed an in-person informed consent form.

2.2.1 Online screening

The (a) *Michigan Alcoholism Screening Test* (MAST) (Selzer, 1971) was used to exclude participants with alcohol dependence (MAST >4), (b) the *Pittsburgh Sleep Quality Index* (PSQI) (Buysse et al., 1989) was used to exclude participants with poor sleep quality (PSQI >5), and (c) the *Beck Depression Inventory*, 2nd Edition (BDI-II) (Beck et al., 1996) was used to exclude participants presenting with depressive symptoms (BDI-II > 13).

2.2.2 Face-to-face screening

The *Mini International Neuropsychiatric Interview* (Version 5.0.0; MINI) (Sheehan et al., 1998) was used to exclude participants with any major psychiatric disorders contained in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) (American Psychiatric Association, 2013). These disorders include depression, bipolar disorder, posttraumatic stress disorder, alcohol abuse/dependence, substance abuse/dependence, and obsessive-compulsive disorder. This measure was also used to cross-validate results obtained via the MAST and BDI-II. More specifically, participants who did not meet the criteria for alcohol abuse/dependence and/or depression on the online screening measures but did meet the criteria for these conditions on the MINI were excluded from participation.

2.3 External experimental measures

2.3.1 Most recent dream form

Participants were asked to complete a Most Recent Dream Form (Domhoff and Schneider, 1998) upon awakening (conclusion of the sleep study). This form asks participants to write down the last dream they can remember (irrespective of when it occurred) in as much detail as possible. This form was used as a criterion for inclusion in the current study—only participants who recalled a dream from the night before upon awakening in the sleep lab, were included in the final sample.

2.3.2 Polysomnography

Two 16-channel Nihon Kohden Neurofax EEG900 electroencephalographs that were adapted for research recorded objective measures of sleep. The PSG included electroencephalography (EEG) which measured brain activity, electrooculography (EOG) which monitored eye movements, chin electromyography (EMG) which monitored muscle tone, and electrocardiography (ECG) which measured heart rate.

A bipolar montage was used with the following bipolar derivations: F3-C3, C3-P3, P3-O1, and F4-C4, C4-P4, P4-O2. This was combined with a referential montage utilizing F3-A2, C3-A2, O1-A2, and F4-A1, C4-A1, O2-A1 derivations. A combination approach was chosen in order to ensure the integrity of the records. Standard filters for sleep recordings were used for the EEG and EOG (0.5–35 Hz), EMG (10–70 Hz), and ECG (1–70 Hz) as recommended by AASM guidelines (American Academy of Sleep Medicine, 2015).

The following external experimental variables were derived from PSG: (1) the proportion of stage 1 non-REM sleep (N1%), (2) the proportion of stage 2 non-REM sleep (N2%), (3) the proportion of stage 3 non-rapid eye movement sleep (N3%), (4) the proportion of REM sleep (REM%), total number of awakenings across the night, the number of awakenings from N1, N2, N3, and REM sleep, as well as the proportion of time spent awake after sleep onset (WASO%).

2.4 Affective/state measures

The measure we used to assess parameters related to positive and negative affect is the *Positive and Negative Affect Schedule—Expanded Version* (PANAS-X) (Watson and Clark, 1994). This questionnaire consists of 60 items measuring 11 specific affects: Fear, Sadness, Guilt, Hostility, Shyness, Fatigue, Surprise, Joviality, Self-Assurance, Attentiveness, and Serenity in the broad context of positive and negative affect. Participants completed the questionnaire in its entirety on the evening before going to sleep on the second (testing night) at the sleep laboratory, as well as the following morning.

We used the "general positive affect" and "general negative affect" composites derived from the evening and morning PANAS to generate the following affective (state) variables to include in our study: (1) Night-time Positive Affect, (2) Night-time Negative Affect, (3) Morning Positive Affect, (4) Morning Negative Affect, (5) overnight percentage change in positive affect, and (6) overnight percentage change in negative affect. A higher score for variable 5 indicates a larger *positive* increase in overnight positive affect (more positive affect the following morning), while a higher score for variable 6 indicates a larger increase in overnight *negative* affect (more negative affect the following morning).

2.5 Personality/trait measures

The first personality/trait measure, the *Ten-Item Personality Inventory* (TIPI) (Gosling et al., 2003), is a short measure of personality based on the "Big Five" personality dimensions. These dimensions include "extroversion," "agreeableness," "openness to experience," "conscientiousness," and "neuroticism" (its converse being "emotional stability"). This short measure was chosen in order to limit the testing burden on participants in the larger study.

The second personality/trait measure we used was the *Boundary Questionnaire—Shortform* (BQ-Sh) (Rawlings, 2001). This measure is based on the original Boundary Questionnaire developed by Hartmann (1989). By means of factor analysis, the BQ-Sh is a short version (46 items) empirically derived from the full (145 items) version. The items in the questionnaire relate to how "permeable" a person's boundary is to external influences. In other words, people with thin boundaries often demonstrate a fluid sense of self, they tend to be sensitive, vulnerable, and sometimes fail to distinguish between fantasy and reality. Conversely, individuals with "thick" boundaries have a distinct sense of self separate from others, can clearly distinguish between fantasy and reality, and are often guarded, and meticulously careful in their actions (Hartmann, 1989; Rawlings, 2001). A higher score on the Boundary Questionnaire is indicative of "thinner" boundaries, i.e., having a more permeable sense of self in relation to the external world.

2.6 Emotion indicators

We developed several novel emotion indicators using the following procedure: generating word-for-word emotional valence ratings throughout the dream via sentiment analysis, (2) applying a "sliding window method" to these valence ratings to control for potential confounds, (3) visualizing the word-for-word (sequential) valence ratings for each dream by plotting the ratings, (4) applying statistical and mathematical techniques to the peak (representing positive emotion) and trough (representing negative emotion) characteristics of the chart to generate the different emotion indicators, (5) validating the emotion indicator variables by correlating them with existing measures/variables known to be associated with dreaming and emotion regulation. Below we describe each step in more detail.

2.7 Sentiment analysis

We employed sentiment analysis, a branch of natural language processing (NLP), that classifies the components of a text as negative, neutral, or positive based on an established lexicon. More specifically, we tested our method by using the VADER package in R statistical software (version 2023.09.0 + 463). The VADER package was recently applied to dream emotion analysis, with some promising results (Yu, 2022). Regarding the analytic process, the software assigns emotion (valence) ratings to each word in the dream report. A negative value is indicative of negative emotion, zero is considered neutral, and a positive value indicates positive emotion. All emotional valence values were continuous in nature. Importantly, our aim was not to replicate the results obtained by Yu (2022), but rather to use the VADER package as a first step to test our novel method for classifying sequential emotion within a dream via specific emotion indicators, given that this package has been used successfully in a similar (dream emotion research) context by Yu (2022).

2.7.1 Dream sentiment analysis procedure

We calculated the average word count across dreams, which was 123.289 with a range of 46–202. Regarding the sentiment analysis procedure, firstly, we pre-processed all dream reports to ensure that only *dream descriptions* (reflecting the actual dream experience) as opposed to *dream commentary* (reflecting thoughts about/interpretations of the dream) were included in each dream report. Following this, we ran each of the 14 pilot dream reports through the VADER package in R. This produced continuous valence ratings for each word in the dream that was either negative, neutral (i.e., 0), or positive in nature. We exported the word-for-word sequential valence ratings for each dream. Next, we applied a "sliding window method" (Martin et al., 2020) as a standardization technique to the dream reports in order to control for two important potential confounds: (1) varying word counts across different dream reports, and (2) difficulties in comparing the degree of emotion across different dreams.

With regard to the specifics of this method, it involves dividing each dream report into fixed-length windows, each containing 30 words and overlapping by 29 words. Following this, the emotion/valence ratings of each 30-word window is summed and then divided by the number of windows in the respective dream reports. This method controls for varying word counts across dreams, and importantly, allows us to "extract" the emotion from the dream when plotting the data points. This is because it eliminates the "flattening" effect that many zero values (neutral/absent emotion) in a dream could have on, for example, the trendline of the chart, as well as the "surface area" of the peaks and troughs containing the plotted valence ratings. See Figure 1 (chart without the sliding window method), and 2 (chart with the sliding window method) for comparison.

2.8 Generating the emotion indicators

First, we visualized sequential emotion across the dream by plotting the valence ratings in what we call a "dream chart." Next, we applied statistical and mathematical techniques to the "micro" properties of the dream chart (e.g., the peaks and troughs in the chart), and secondly to the "macro" properties of the dream chart (e.g., trendline gradient) in order to generate the set of emotion indicators that signify the respective micro and macro changes in emotion across a dream.

2.8.1 Peaks and troughs

Positive emotion is represented by peaks in the dream chart and negative emotion is represented by troughs in the dream chart, with the midline being zero (neutral/absent emotion). The first step was to classify/demarcate the peaks and/or troughs contained in each dream chart. Each chart represents the sequential emotion/valence ratings contained in one dream report. The starting point of a peak was defined as the first positive valence value identified above the horizontal line representing 0 (the absence of emotion, see Figure 2), while its end was defined as the first value of 0 following one, or a sequence of positive values. There could be no, a single, or several peaks present in one dream chart. Similarly, a trough was defined as the first negative valence value identified below the horizontal line representing 0 (the absence of emotion), while its end was defined as the first value of 0 following one,

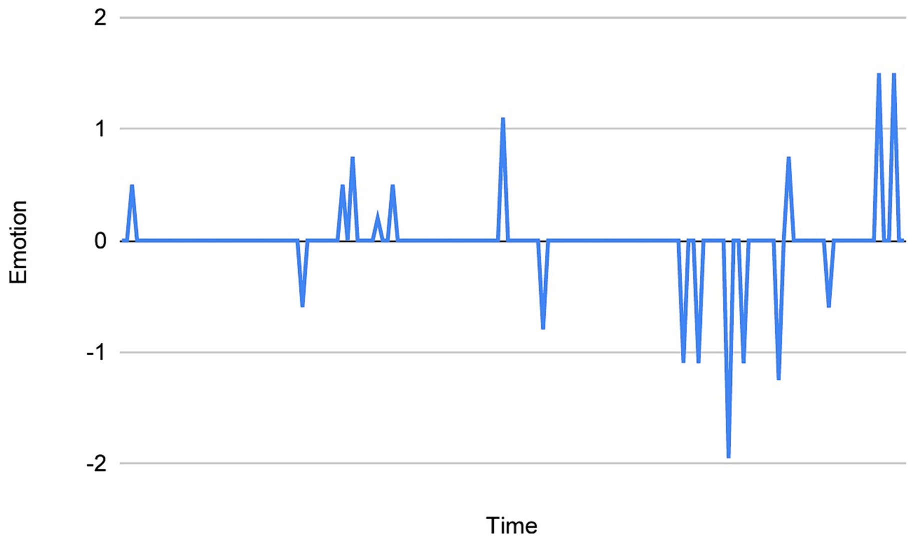

FIGURE 1
An example of a chart *without* the sliding-window method applied.

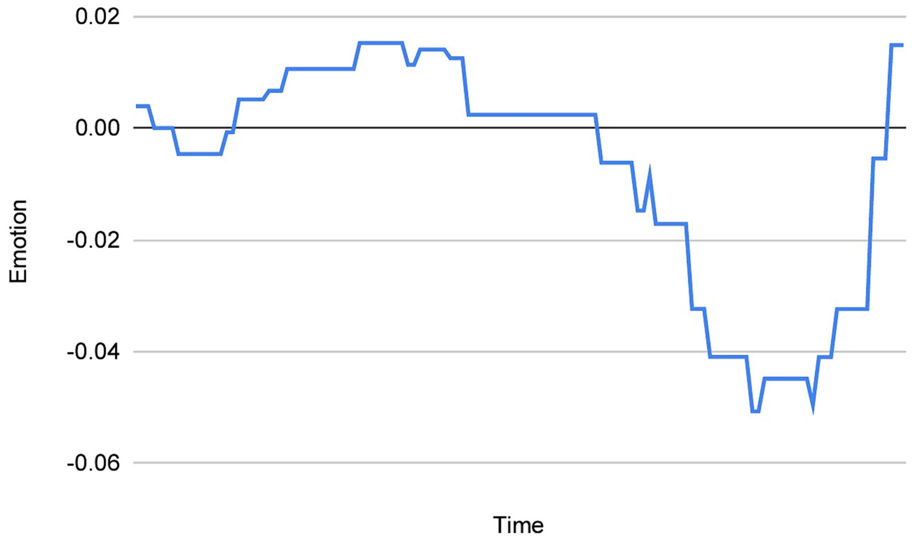

FIGURE 2
An example of a chart *with* the sliding window method applied.

or a sequence of negative values. There could be no, a single, or several troughs present in one dream chart. It is theoretically possible for a peak or a trough to consist of a single positive value in the case of the former, and a single negative value in the case of the latter.

2.8.1.1 Emotion indicators based on dream peaks and troughs

We developed a series of primary dream emotion indicators based on the microstructure (peak and trough properties) of the dream charts. Below we discuss each indicator in more detail and provide formulas where applicable.

2.8.1.2 Total peaks and troughs

The first two emotion indicators based on the peak and trough properties in the dream charts include the total number of peaks and troughs present in a single dream chart:

$$\sum_{\text{Peak}} = TotalPeaks$$

$$\sum_{\text{Trough}} = TotalTroughs$$

2.8.2 Emotion intensity and area under the curve

We also wanted to determine whether the height of the peaks, or, conversely, the depth of the troughs, are influential parameters when objectively quantifying sequential emotion within a dream. Based on this, we refer to the emotion indicators discussed in this section as the "intensity" of positive and negative emotion within a dream, as well as an "overall emotional intensity" indicator.

Given the small number of dream reports of this pilot study, we have a limited range of chart and curve characteristics to work with when devising and conducting analysis. This coupled with the exploratory, POC nature of this study, as well as the aim of future studies replicating and refining the methods, we decided to implement a simple preliminary method as a starting point to approximate the area under the curve for each peak and trough in the dream charts. Put differently, the aim was to use this method to provide us with an initial approximation of the "emotional real estate" (positive or negative) that each peak and trough occupy on the surface area of the dream chart.

2.8.2.1 The trapezoid rule for approximating peak and trough AUCs

The trapezoid rule is an integration rule, in calculus, that evaluates the area under the curve by segmenting the total area into mini trapezoids and summing their areas. We used this method to estimate the area under the curve (AUC) of the respective peaks and troughs contained in our dream charts. In our dream charts, the x-axis represents a "time variable" indicative of the chronological sequence of the dream on a uniformly spaced scale. The y-axis represents the sequential word-for-word emotion ratings derived from sentiment analysis. We used the trapezoidal rule to compute the AUC by approximating each segment between consecutive data points as a trapezoid, and consequently summing their areas (see Figure 3 for a visualization). This computation generates an approximation of the integral of the curve, which represents the cumulative emotionality contained in each peak and trough across the dream chart in a chronological fashion. Put differently, the AUC in this context represents an objective, quantitative indicator of the cumulative positive emotional intensity for each peak, and cumulative negative emotional intensity for each trough. It assists us in uncovering the distribution and intensity of positive and negative emotional intensity within and across a dream.

2.8.2.1.1 Emotion intensity indicators

We derived three composite emotion indicators by applying the trapezoidal rule to calculate the approximated AUC of the peaks and/or troughs contained in our dream charts: (1) positive emotion intensity, (2) negative emotion intensity, and (3) overall emotion intensity.

2.8.2.1.1.1 Positive emotion intensity

This emotion indicator was calculated by summing the approximated AUCs of all the peaks contained in a dream chart (visually represented by all the purple mini trapezoids in Figure 3), which equates to the Positive Emotion Intensity (PEI) indicator. Larger positive values of this indicator denote a higher positive emotion intensity detected in dream charts. In the event where there were no peaks present in a dream chart, the PEI value = 0. In the event where there was only one peak per dream chart, the approximated AUC of the relative peak = PEI. See formula below for calculating the PEI for a dream chart containing >1 peak, where:

$$\textit{Positive Emotion Intensity} = P$$

Peak = a peak defined by the method outlined in the preceding section

aAUC = approximated area under the curve for each peak per dream chart

$$P = \sum \text{aAUC}_{\text{Peak}}$$

2.8.2.1.1.2 Negative emotion intensity

This emotion indicator was calculated by summing the approximated AUCs of all the troughs contained in a dream chart (visually represented by all the orange mini trapezoids in Figure 3), which equates to the Negative Emotion Intensity (NEI) indicator. Larger negative values of this indicator denote a higher negative emotion intensity detected in dream charts. In the event where there were no troughs present in a dream chart, the NEI value = 0. In the event where there was only one trough per dream chart, the approximated AUC of the trough = NEI. See formula below for calculating the NEI for a dream chart containing >1 trough, where:

$$\textit{Negative Emotion Intensity} = N$$

Trough = a trough defined by the method outlined in the preceding section

aAUC = approximated area under the curve for each trough per dream chart

$$N = \sum \text{aAUC}_{\text{Trough}}$$

2.8.2.1.1.3 Overall emotion intensity

This emotion indicator was calculated by summing the following: the aAUCs of all the peaks in a dream chart along with the absolute aAUCs of all the troughs in a dream chart (visually represented by all the orange and purple mini trapezoids in Figure 3). Absolute values were used for troughs in order to capture the full range of emotion intensity of the summed peak and trough aAUCs. Put differently, and conversely, retaining the negative sign of trough aAUC values would have subtracted, as opposed to added, to the range of positive and negative emotional intensity represented by the peaks and troughs in the dream charts. See formula below for calculating the Overall Emotion Intensity, where:

$$E = \textit{Overall Emotion Intensity}$$

$$p = \textit{Positive Emotion Intensity}$$

$$n = \textit{Negative Emotion Intensity}$$

$$|n| = \textit{Absolute value of Negative Emotion Intensity}$$

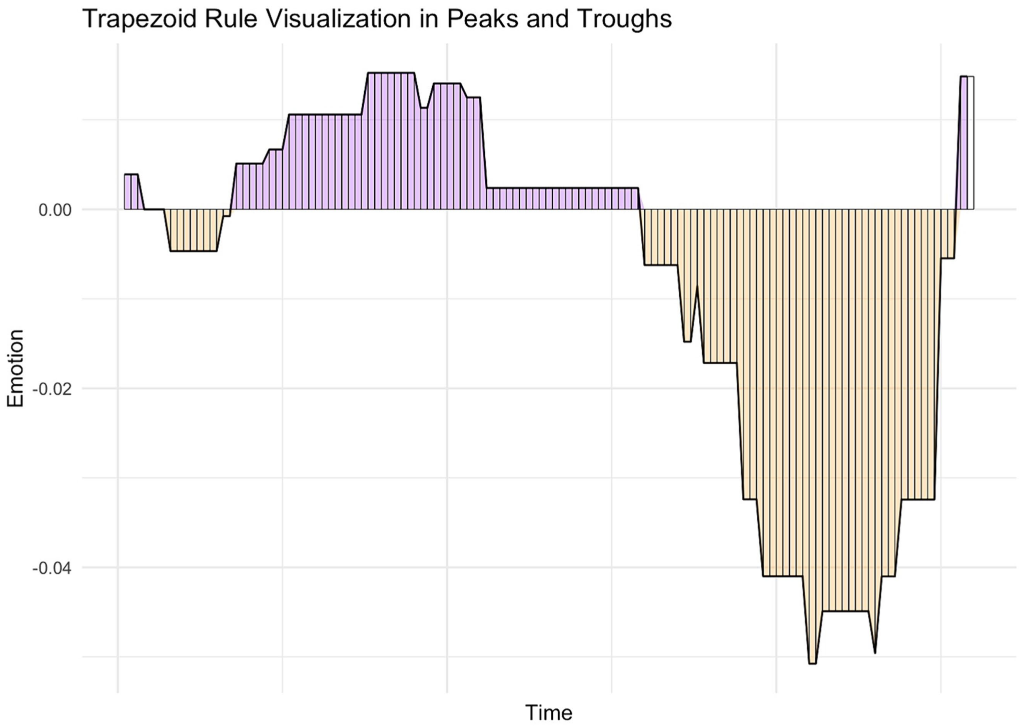

FIGURE 3
Trapezoid rule visualization in peaks and troughs. Peaks are shaded in purple, and troughs are shaded in orange. Peaks and troughs are segmented into mini trapezoids/rectangles using the Trapezoid Rule visualized by the vertical lines within each peak and trough. The summed segments represent an approximation of the area under the curve for each peak and trough.

$$E = p + |n|$$

2.8.3 Emotion gradient: dream chart trendlines

In this method, we aimed to use a "macro" emotion indicator, called "emotion gradient," as a signifier of the overall direction of sequential emotion changes within a dream. In order to achieve this, we used the trendline gradient of the dream charts as the primary macro emotion indicator. This macro indicator stands in contrast to examining the properties of peaks and troughs (and the composites derived from these properties), as micro-level indicators of the direction of emotional changes within a dream. This is consistent with determining, in a preliminary fashion, whether the dream chart trendline gradients could serve as proxies for overall emotion regulation unfolding within the dream.

Given the predominantly fluctuating nature of the data (typically oscillating between peaks and troughs, often with multiple oscillations), preliminary analyses showed that, overall, the most appropriate fit for the data was a polynomial trendline (as opposed to a linear trendline). We decided to use a polynomial trendline as an approximation of the overall/macro direction of emotion changes across a dream. Twelve out of the 14 pilot dream charts exhibited properties where a polynomial trendline would constitute an appropriate fit. Since a polynomial trendline is an appropriate fit for the vast majority of dream charts, we decided to proceed with fitting this type of trendline to the eligible dream charts. Following this, we also examined goodness-of-fit statistics to evaluate the suitability of using a polynomial trendline for all dream charts included in analysis. All analyses were conducted using R statistical software, Version 2023.09.0+463. See Figure 4 for a visual example of a polynomial trendline fitted to a dream chart.

The emotion indicator derived from this method is based on the polynomial trendline gradient and is called the "emotion gradient" of dreams. Importantly, again, we propose this method as a theoretical starting point for future studies to replicate and refine.

2.9 Statistical analysis

All data was analyzed using R statistical software, Version 2023.09.0+463. We included two sets of variables in the correlational analyses: "emotion indicators" (the novel variables described in the preceding sections), and "external variables" (existing sleep, state, and trait variables recognized to have a relationship with dreaming and/or emotion regulation). Descriptive statistics indicated that the variables typically exhibited non-linear properties. This is likely, at least partially, due to the small sample size included in this pilot study. Due to the distribution properties of the data, we decide to run Spearman's Rho, as a non-parametric alternative, provided there were no ties present in the data (Spearman's rho ranks the data, and ties can distort true results). In the event where there were ties present in the data for a particular variable, Pearson's R was used as an alternative. We readily acknowledge that this is not an ideal solution; however, due to the exploratory and POC nature of this pilot study, we decided to proceed and report any significant correlations found, especially if there is existing empirical evidence to support the relationship between the emotion indicators and the external variables. Furthermore, because of the small sample and our intention to provide pilot data, we did not correct for multiple correlations.

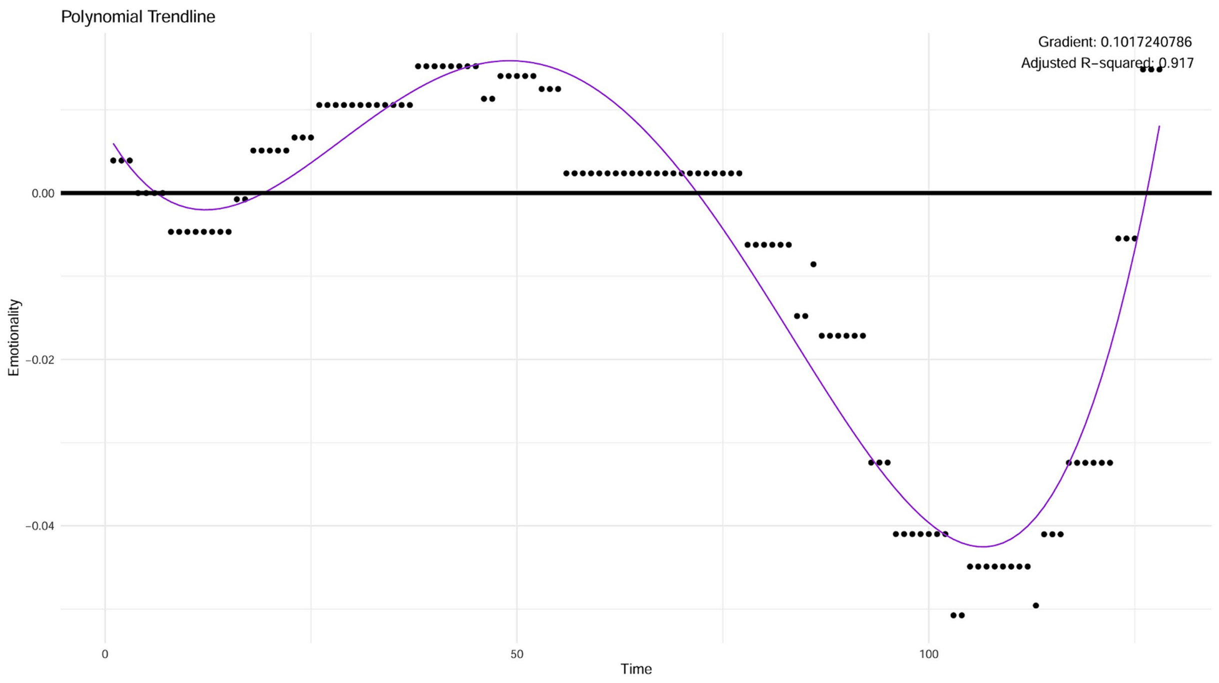

FIGURE 4
Example of a polynomial trendline fitted to a dream chart.

TABLE 1 Descriptive statistics for emotion indicators derived from sentiment analysis.

Emotion indicator	Mean (SD)	Range
Peaks	1.57 (1.16)	4.00
Troughs	1.14 (1.10)	3.00
Positive emotion intensity	1.04 (1.12)	3.76
Negative emotion intensity	−0.60 (0.85)	2.16
Overall emotion intensity	1.59 (1.14)	4.20
Emotion gradient	0.012 (0.037)	0.14
Word count*	123.289 (59.689)	46–202

Peaks = average number of Peaks present per dream report across all dreams (N = 14). Troughs = average number of Troughs present per dream report across all dreams (N = 14). Positive Emotion Intensity = the average approximated area under the curve for all Peaks in a dream across all dreams (N = 14). Negative Emotion Intensity = the average approximated area under the curve for all Troughs in a dream across all dreams (N = 14). Overall Emotion Intensity = the summed Positive Emotion Intensity and the absolute Negative Emotion Intensity for each dream across all dreams (N = 14). Emotional Gradient = the average of the overall direction of sequential emotion changes within a dream across all eligible dreams (N = 11). *We have used the sliding window method to control for discrepancies in word counts across dreams.

3 Results

3.1 Emotion indicators and external variables

The final emotion indicators included the following: Total Peaks (Peaks), Total Troughs (Troughs), Positive Emotion Intensity (PEI), Negative Emotion Intensity (NEI), Overall Emotion Intensity (OEI), and Emotion Gradient (Emo_Grad). The external variables included in analyses can be divided into three broad classes: (1) PSG-derived sleep variables (classified as "sleep" variables), which include variables like sleep stage distribution and number of awakenings, (2) affective variables (classified as "state" variables), which are variables derived from the Positive and Negative Affect Scale—Extended Version (PANAS), and (3) personality variables (classified as "trait" variables) derived from the Ten-Item Personality Inventory, and the Boundary Questionnaire.

In relation to our first and second objectives, generating emotion indicators from quantifying and visualizing emotion fluctuations using sentiment analysis, we ran descriptive statistics on all the micro- and macro-level emotion indicators.

These results show that we were able to successfully classify and quantify the fluctuations between positive and negative emotions within a dream using the sentiment analysis package, VADER, in R (see Table 1). The directions of these fluctuations were largely consistent with the emotional tone of the dream as subjectively evaluated by the researchers.

With regard to our second objective, correlating the above-mentioned emotion indicators with sleep, state, and trait variables, results showed the following: Statistically significant correlations between 5/6 of the novel objective emotion indicators and at least one external variable (sleep, state, or trait variables). The network diagram depicted in Figure 5 visualizes the statistically significant relationships between two sets of nodes: emotion indicators and the different classes of external variables. The direction of the relationships between the two sets of nodes (significant positive or significant negative correlation) are presented as edges in the diagram. The statistical parameters of the results can be found in Table 2.

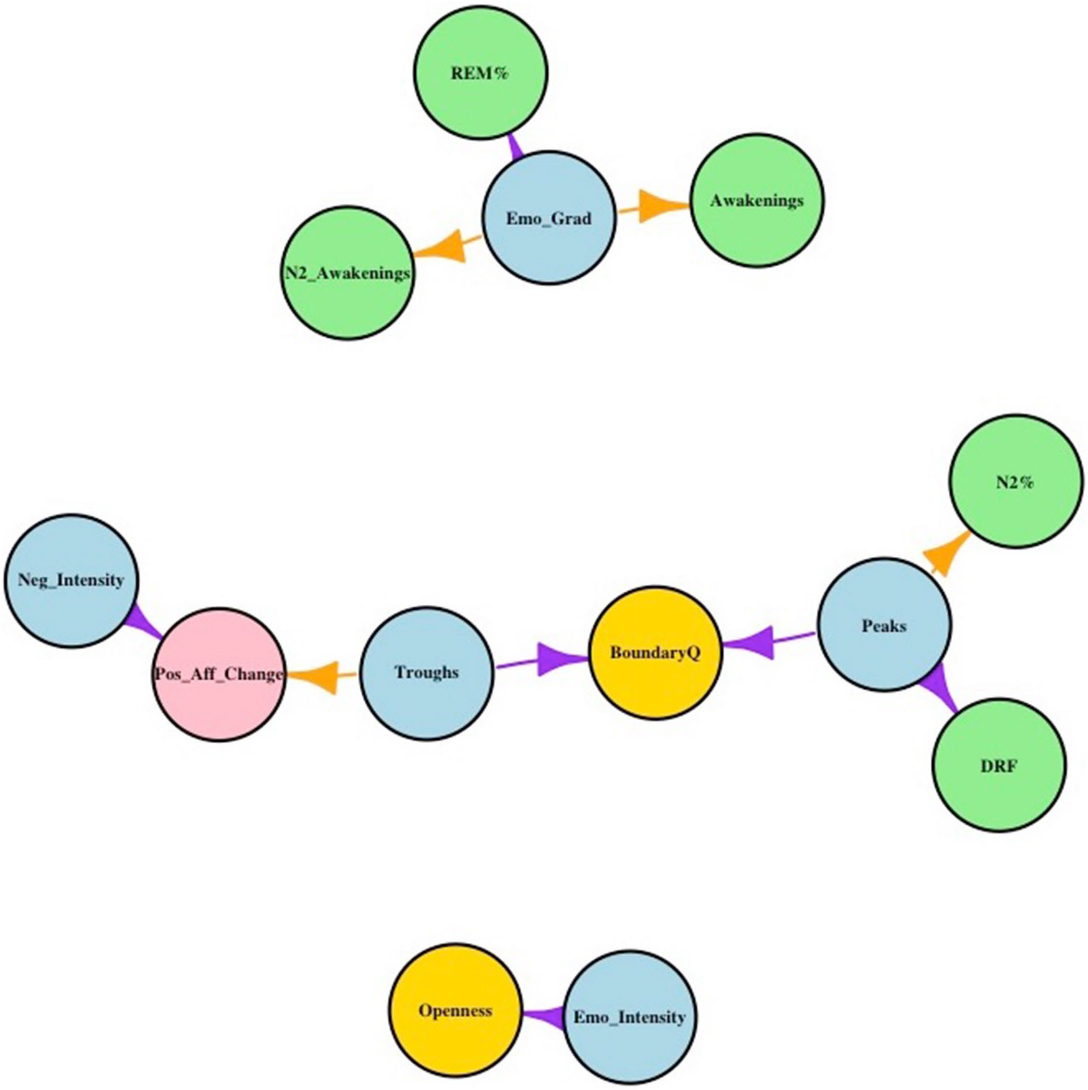

FIGURE 5
All emotion indicators are presented in light blue: Peaks = Total Peaks; Troughs = Total Troughs; Emo_Fluct = Emotion Fluctuations; Neg_Intensity = Negative Emotion Intensity; Emo_Intensity = Overall Emotion Intensity (OEI); Emo_Grad = Emotion Gradient. The state variable is presented in pink: Pos_Aff_Change = percentage overnight change in positive affect. Trait variables are presented in yellow: Openness = TIPI-Openness Big 5 Personality dimensions; BoundaryQ = Boundary Questionnaire—Shortform (higher score = "thinner" boundaries). Sleep variables are presented in light green: Awakenings = The total number of awakenings across the night; N2_Awakenings = Awakenings from stage 2 non-REM sleep across the night; REM% = percentage of time spent in REM sleep across the night; N2% = percentage of time spent in non-REM stage 2 sleep across the night; DRF = dream recall frequency. Purple arrows denote a statistically significant *positive* relationship between an emotion indicator and external variable(s), while orange arrows denote a statistically significant *negative* relationship between an emotion indicator and external variable(s).

TABLE 2 Statistical parameters of significant correlations depicted in the network diagram.

	REM%	N2%	Awakenings	N2 awakenings	DRF	Positive affective change	Boundary Q	Openness
Peaks	0.238	−0.548*	−0.269	−0.320	0.608*	−0.307	0.642*	−0.118
Troughs	0.205	−0.529	−0.076	−0.138	0.358	−0.590*	0.702**	−0.097
Negative emotion intensity$^{\rho}$	0.043	0.074	0.052	−0.043	0.103	−0.666**	0.482	−0.047
Overall emotion intensity	0.088	−0.042	0.227	0.141	0.336	−0.179	0.377	0.548*
Emotion gradient$^{\rho}$	0.627*	−0.555	−0.743**	−0.749**	−0.176	−0.536	−0.141	−0.397

Peaks = Total Peaks; Troughs = Total Troughs; REM% = percentage of time spent in REM sleep across the night; Awakenings = The total number of awakenings across the night; N2_Awakenings = Awakenings from stage 2 non-REM sleep across the night; N2% = percentage of time spent in non-REM stage 2 sleep across the night; DRF = dream recall frequency; Positive Affective Change = percentage overnight change in positive affect; Openness = TIPI-Openness Big 5 Personality dimensions; Boundary Q = Boundary Questionnaire—Shortform (higher score = "thinner" boundaries). *p = <0.05; **p = <0.01. $^{\rho}$ = Spearman's rho is reported where there were no ties present in the dataset, else Pearson's r is reported.

With regard to the first emotion indicator, the emotion gradient, the results show, firstly, that a *steeper, positive* emotion gradient in a dream chart has a significant *positive* relationship with the proportion of REM sleep a person had across the night. Put differently, a *steeper upward* trend of positive emotion across a dream is associated with an *increased* amount of REM sleep across the night. Secondly, results show that a *steeper, upward* trend in positive emotion across a dream is significantly associated with *fewer* overall awakenings, and awakenings from N2, across the night.

The next observed trend involves two emotion indicators based on the microstructure of the dream chart—trough characteristics: the results show that (a) fewer troughs across the dream chart, and (b) troughs with a larger surface area (AUC), are significantly associated with a *higher percentage increase* in overnight positive affect. Therefore, the properties (number and size) of dream troughs appear to be important in affective processes (state variable) in this sample. In addition, the number of troughs in a dream chart is also significantly associated with having "thinner" boundaries ("trait" variable) as measured by the Boundary Questionnaire. Interestingly, the number of peaks, another indicator of the microstructure of the dream chart, is also significantly associated with having "thinner" boundaries. However, where the number of throughs showed a significant association with affective processes (state variable), the number of peaks shows an additional significant association with two sleep variables: N2% and DRF. More specifically, a *higher* number of peaks present in a dream chart is significantly associated with (a) having *less* overall N2 sleep across the night, and (b), having *higher* dream recall rates.

The third and final trend derived from the network diagram relates to another indicator of the microstructure of the dream charts—the overall emotion intensity (OEI) indicator. The OEI is a composite measure that is based on the summed approximated AUCs of all the peaks and troughs present in the dream chart. Therefore, the overall emotion intensity of a dream is influenced by both the number of peaks and troughs, as well as the height and depth (intensity) of the respective peaks and troughs. In summary, it is a composite of the approximated positive and negative intensity of emotion contained in each dream chart. The results show that there is a significant association between higher emotion intensity and scoring high on the "Openness" dimension of the Big Five personality dimensions as measured by the TIPI. Put differently, participants with a higher receptivity to new ideas and experiences tended to have higher overall emotion intensity in their dreams in this sample.

4 Discussion

In this study, we aimed to develop a novel method for objectively classifying and quantifying sequential emotion fluctuations within a dream. We propose that these fluctuations could serve as a proxy for emotion regulation processes unfolding within a dream. We used sentiment analysis, a branch of Natural Language Processing (NLP), as the objective measure for assigning emotion valence ratings to each word in a dream report. Next, we implemented a sliding window method (Martin et al., 2020) in order to (a) control for varying word counts across dream reports, and (b) for "extracting" emotion from the dream reports by eliminating the flattening effect that many 0 s (neutral emotion) could have on the peak and trough properties, as well as the trendline gradient of the dream charts. Based on these results we generated 6 "emotion indicators" derived from both the micro- and macro-level properties of the dream charts. Finally, we correlated the emotion indicators with three classes of external variables: sleep, state, and trait variables.

We were able to successfully complete all of the steps summarized above, while results also confirm that we have met our three operationalized objectives: (1) using sentiment analysis to objectively classify, quantify, and visualize sequential fluctuations between positive and negative emotions across a dream, (2) using these visual properties along with the valence ratings derived from sentiment analysis to generate "emotion indicators," and (3) preliminarily validating the novel emotion indicators by showing a statistically significant correlation with external variables (sleep, state, and trait) known to be associated with dreaming and/or emotion regulation.

4.1 Limitations and directions for future research

The first limitation of this pilot study relates to not including the ratings of several independent, expert (human) raters of dream reports. We suggest that future studies incorporate this into their methodology as an external form of validation with regard to the valence ratings generated by sentiment analysis. The second limitation of this pilot study relates to the small sample size and the limitations this imposes on the results. Therefore, we present our findings in this methodological paper as preliminary and as a proof-of-concept for future studies to replicate and refine. More specifically, we propose that future studies recruit (a) larger, and (b) more diverse (e.g., clinical and-nonclinical) samples. Next, we propose that future studies test the trapezoid rule as a method for approximating the area under the curve against other, potentially more precise and refined methods. In addition, we suggest future studies explore additional types of trendlines to be fitted based on the specific properties of their dream charts so that different types of dream charts can be accommodated in the sample in its entirety. Generating additional types of emotion indicators is another avenue for future studies to pursue. Finally, although beyond the scope of this methodological article, future studies should aim to build on our results from correlational analyses in order to elucidate the nature of the relationship between the emotion indicators and the sleep (e.g., REM%), state (e.g., overnight affective change), and trait (personality) variables. This could be done, for example, by experimentally testing the proposition that these (or other) emotion indicators serve as proxies for emotion regulation processes unfolding within a dream. Investigating dream-related emotion regulation processes has important implications, not only for garnering insights into the course of healthy dreaming, but for understanding psychiatric disorders, where emotion dysregulation and alterations in dream activity coincide.

5 Conclusion

In this proof-of-concept paper we have described a novel method for objectively quantifying sequential emotion within a dream using sentiment analysis. We also present preliminary evidence from descriptive statistics and correlational analyses that support the rationale for using sentiment analysis in this manner, as well as in support of the (preliminary) validity of the emotion indicators derived from our method. We see this method along with the accompanying results as a first step in developing a new, objective method for rating and visualizing emotion in dreams. We also propose that the emotion indicators we have developed could serve as potential proxies for emotion regulation processes unfolding during a dream. With these preliminary findings we aim to provide a methodological foundation for future studies to test and refine the method employed in the current study in larger and more diverse samples.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethical Review Board of the Department of Psychology, University of Cape Town and the Ethical Review Board of the Faculty of Humanities, University of Cape Town. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MW: Writing – original draft, Writing – review & editing. MS: Writing – original draft, Writing – review & editing. GL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding from the National Research Foundation (Grant No. 83345) and the Oppenheimer Memorial Trust (Grant No. 20312/02) supported the research (data collection) phase of this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

American Academy of Sleep Medicine. (2015). *The AASM manual for the scoring of sleep and associated events in 2015 (ed.)*: American Academy of Sleep Medicine.

American Psychiatric Association, D. S. M. T. F., and American Psychiatric Association, D. S. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (Vol. 5, No. 5). Washington, DC: American psychiatric association.

Beck, A. T., Steer, R. A., and Brown, G. K. (1996). *Manual for the beck depression inventory II*. San Antonio: Psychological Corporation.

Blackman, M. R. (2000). Age-related alterations in sleep quality and neuroendocrine function: interrelationships and implications. *J. Am. Med. Assoc.* 284, 879–881. doi: 10.1001/jama.284.7.879

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatr. Res.* 28, 193–213.

Domhoff, G. W., and Schneider, A. (1998). New rationales and methods for quantitative dream research outside the laboratory. *Sleep* 21, 398–404. doi: 10.1093/sleep/21.4.398

Elce, V., Handjaras, G., and Bernardi, G. (2021). The language of dreams: application of linguistics-based approaches for the automated analysis of dream experiences. *Clocks Sleep.* 3, 495–514. doi: 10.3390/clockssleep3030035

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the big-five personality domains. *J. Res. Pers.* 37, 504–528. doi: 10.1016/S0092-6566(03)00046-1

Hall, C., and van de Castle, R. L. (1966). *The content analysis of dreams*. East Norwalk, CT: Appleton-Century-Crofts.

Hartmann, E. (1989). Boundaries of dreams, boundaries of dreamers: Thin and thick boundaries as a new personality measure. *Psychiatr. J. Univ. Ott*.

Irwin, M., Miller, C., Gillin, J. C., Demodena, A., and Ehlers, C. L. (2000). Polysomnographic and spectral sleep EEG in primary alcoholics: an interaction between alcohol dependence and African-American ethnicity. *Alcohol Clin. Exp. Res.* 24, 1376–1384. doi: 10.1111/j.1530-0277.2000.tb02106.x

Lee, K. A. (1998). Alterations in sleep during pregnancy and postpartum: a review of 30 years of research. *Sleep Med. Rev.* 2, 231–242. doi: 10.1016/S1087-0792(98)90010-7

Martin, J. M., Andriano, D. W., Mota, N. B., Mota-Rolim, S. A., Araújo, J. F., Solms, M., et al. (2020). Structural differences between REM and non-REM dream reports assessed by graph analysis. *PLoS One* 15:e0228903. doi: 10.1371/journal.pone.0228903

Nielsen, T. A., and Stenstrom, P. (2005). What are the memory sources of dreaming? *Nature* 437, 1286–1289. doi: 10.1038/nature04288

Pagel, J. F. (2010). Drugs, dreams, and nightmares. *Sleep Med. Clin.* 5, 277–287. doi: 10.1016/j.jsmc.2010.01.007

Rawlings, D. (2001). An exploratory factor analysis of Hartmann's boundary questionnaire and an empirically-derived short version. *Imagin. Cogn. Pers.* 21, 131–144. doi: 10.2190/3XM9-LGA6-MJ76-X658

Röver, S. A., and Schredl, M. (2017). Measuring emotions in dreams: effects of dream length and personality. *Int. J. Dream Res.* 10, 65–68.

Schredl, M. (2010a). Dream content analysis: basic principles. *Int. J. Dream Res.* 3, 65–73.

Schredl, M. (2010b). Do sleep disorders affect the dreaming process? Dream recall and dream content in patients with sleep disorders. *Sleep Med. Clin.* 5, 193–202. doi: 10.1016/j.jsmc.2010.01.008

Schredl, M., Beaton, A. A., Henley-Einion, J., and Blagrove, M. (2013). Handedness and dream-recall frequency. *Dreaming* 23, 156–162. doi: 10.1037/a0032210

Schredl, M., and Doll, E. (1998). Emotions in diary dreams. *Conscious. Cogn.* 7, 634–646. doi: 10.1006/ccog.1998.0356

Selzer, M. L. (1971). The Michigan alcoholism screening test: the quest for a new diagnostic instrument. *Am. J. Psychiatry* 127, 1653–1658. doi: 10.1176/ajp.127.12.1653

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 Suppl 20, 22–33

Sikka, P., Valli, K., Virta, T., and Revonsuo, A. (2014). I know how you felt last night, or do I? Self-and external ratings of emotions in REM sleep dreams. *Conscious. Cogn.* 25, 51–66. doi: 10.1016/j.concog.2014.01.011

Skancke, J. F., Holsen, I., and Schredl, M. (2014). Continuity between waking life and dreams of psychiatric patients: a review and discussion of the implications for dream research. *Int. J. Dream Res.* 7, 39–53.

Watson, D., and Clark, L. A. (1994). The PANAS-X: manual for the positive and negative affect schedule-expanded form. Ames: The University of Iowa Available at: https://iro.uiowa.edu/esploro/outputs/other/The-PANAS-X-Manual-for-the-Positive/9983557488402771/filesAndLinks?index=0.

Yu, C. K. C. (2022). Automated analysis of dream sentiment—the royal road to dream dynamics? *Dreaming* 32, 33–51. doi: 10.1037/drm0000189

Zadra, A., and Domhoff, G. W. (2011). Dream content: quantitative findings. *Princ. Pract. Sleep Med. 5, 585–594. doi: 10.1016/B978-1-4160-6645-3.00050-5*

10.3389/fpsyg.2024.1393913

Appendix

Dream report example 1:

My boyfriend's mother was upset with me. We were at his house, in his kitchen. We were making samoosas and rotis and I wasn't doing very well. I ended up crying because she said that we could not get married because my rotis were not round enough and my samoosas are bad. I cried for most of my dream.

See Figure A1 for visualized valence ratings from sentiment analysis using the VADER package *prior to* applying the sliding window method.

FIGURE A1
Visualization of untransformed dream emotion prior to applying the sliding window method in Dream 1.

See Figure A2 for visualized valence ratings from sentiment analysis using the VADER package *after* applying the sliding window method.

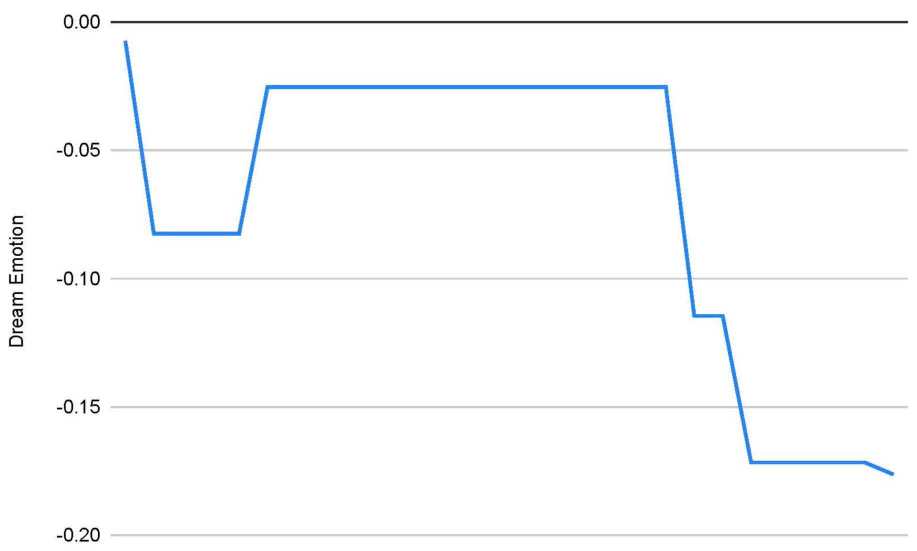

FIGURE A2
Visualization of dream emotion after applying the sliding window method in Dream 1.

Dream report example 2:

I was at Comic Con and I was leaving at one of the gates and Taylor Swift, Selena Gomez and Taylor's girlfriend were standing outside it. I was waiting for someone to come fetch me so I was on my phone. Taylor leaned forward and said "hello." I was shocked and said "hi" and introduced myself and we spoke about where I came from etc. *and then Taylor invited me on her next tour with her. When I went home, my friend Matthew (who is Taylor's number 1 fan) was there. I started making pasta while telling him about my experience but I was too excited, so he made my pasta for me.*

See Figure A3 for visualized valence ratings from sentiment analysis using the VADER package *prior to* applying the sliding window method.

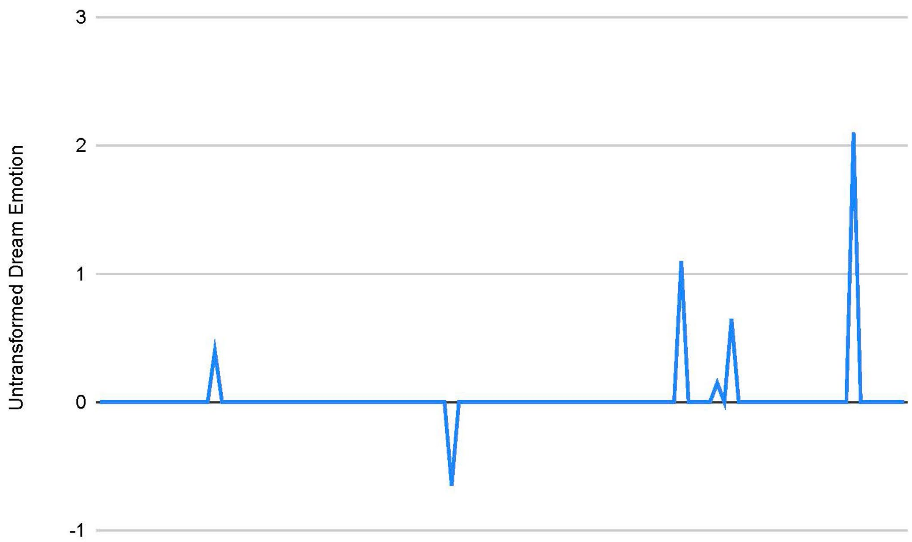

FIGURE A3
Visualization of untransformed dream emotion prior to applying the sliding window method in Dream 2.

See Figure A4 for visualized valence ratings from sentiment analysis using the VADER package *after* applying the sliding window method.

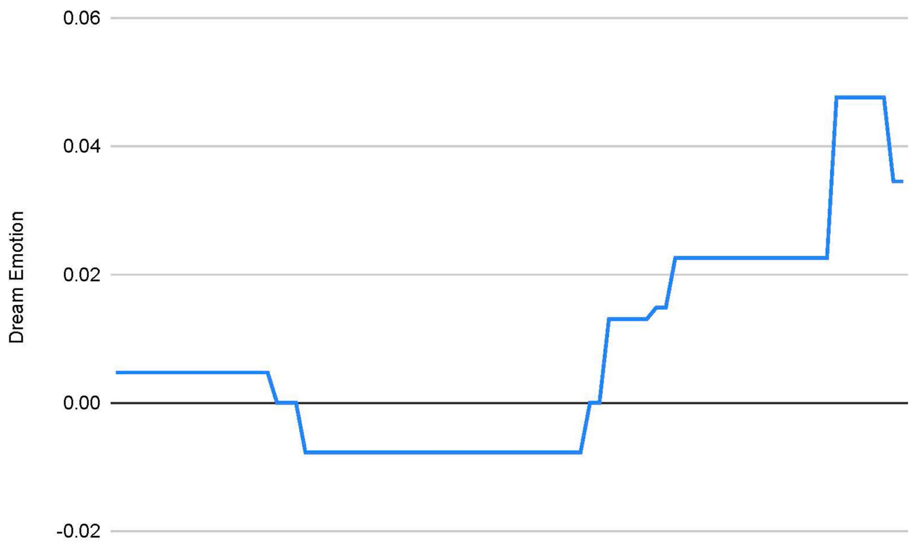

FIGURE A4
Visualization of dream emotion after applying the sliding window method in Dream 2.