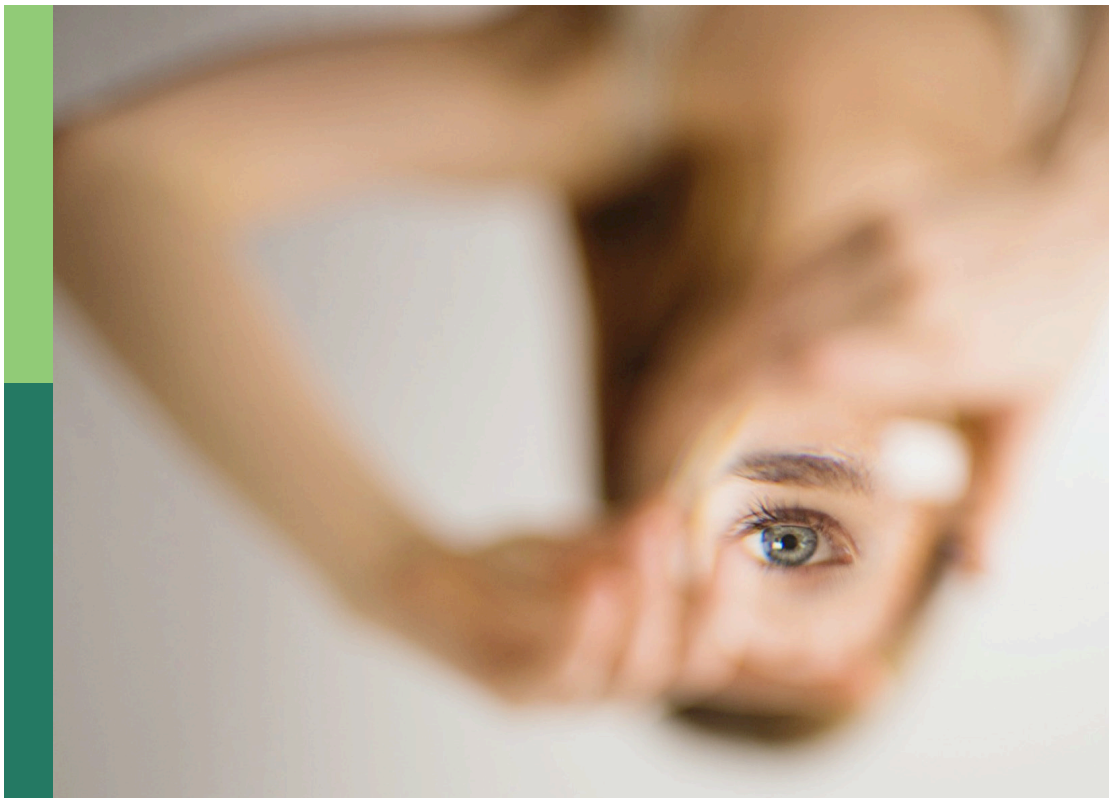
frontiers | Research Topics

Embodied bounded rationality

Edited by
Shaun Gallagher, Riccardo Viale and Vittorio Gallese

Published in
Frontiers in Psychology



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-3343-7
DOI 10.3389/978-2-8325-3343-7

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

Embodied bounded rationality

Topic editors

Shaun Gallagher — University of Memphis, United States

Riccardo Viale — University of Milano-Bicocca, Italy

Vittorio Gallese — University of Parma, Italy

Citation

Gallagher, S., Viale, R., Gallese, V., eds. (2023). *Embodied bounded rationality*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3343-7

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 10 August 2023
DOI 10.3389/fpsyg.2023.1235087

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Bernhard Hommel,
Shandong Normal University, China

*CORRESPONDENCE
Riccardo Viale
✉ viale.riccardo2@gmail.com

RECEIVED 05 June 2023
ACCEPTED 04 July 2023
PUBLISHED 10 August 2023

CITATION
Viale R, Gallagher S and Gallese V (2023) Editorial: Embodied bounded rationality. *Front. Psychol.* 14:1235087. doi: 10.3389/fpsyg.2023.1235087

COPYRIGHT
© 2023 Viale, Gallagher and Gallese. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Embodied bounded rationality

Riccardo Viale[1]*, Shaun Gallagher[2] and Vittorio Gallese[3]

[1]Department of Economics, University of Milano-Bicocca, Milan, Italy, [2]University of Memphis, Memphis, TN, United States, [3]Department of Medicine and Surgery-Unit of Neuroscience, University of Parma, Parma, Italy

KEYWORDS

bounded rationality, embodied cognition, enactivism, problem-solving, decision-making

Editorial on the Research Topic
Embodied bounded rationality

In the last 25 years, a new foundational perspective has emerged in the cognitive sciences under the title of embodied cognition. The core of embodied cognition can be expressed by the general hypothesis that cognitive processes are fundamentally rooted in the morphological traits and sensorimotor and affective systems of the human body. Thinking is based primarily on modal embodied processes rather than amodal ones. These lines of research more or less explicitly recognize the centrality of the embodied variables in economic psychology. This Research Topic aims to demonstrate that the adaptive and ecological dimensions of bounded rationality can be better analyzed by assuming an embodied cognition perspective. Several of the articles in this Research Topic consider how embodied-enactive models of cognition, and the notion of embodied rationality, compare with Herbert Simon's bounded rationality.

Viale et al., in their article "Bounded rationality, enactive problem-solving and the neuroscience of social interaction" aim to show that there is an alternative way to explain human action with respect to the bottlenecks of decision-making psychology. This topic shows that the alternative route recovers the tradition of bounded rationality and problem-solving of Newell and Simon and inserts it into the new research agenda of embodied cognition. According to Simon, the center of gravity of the rationality of the action lies in the ability to adapt. Using the language of embodied cognition, this adaptivity is concerned with the possible solutions implemented to address environmental tasks and problems. From this point of view, the new term, enactive problem-solving, summarizes this fusion between the two moments and could well represent the phenomenon. Problem-solving takes place in a dynamic relationship of an enactive type in a problem space. Within it, repeated feedback allows you to gradually shape the solution. Enactive problem-solving is achieved through the bodily and neural mechanisms typical of embodied cognition, such as the mirror neuron system. Its adaptive function seems effective both in practical and motor tasks and in abstract and symbolic ones. Enactive problem-solving also seems to be able to explain the underlying mechanisms of embodied bounded rationality. Petracca in his article, "Embodying bounded rationality: From embodied bounded rationality to embodied rationality," considers that embodied rationality is associated with the more radical forms of enactive-embodied cognition, which suggests a genuine transformation in our concept of the rational. He considers the relationship between bounded rationality and the concept of embodied rationality drawn from the multiple views found in embodied cognition literature.

He argues that a range of such embodied views, from moderate to radical versions, can inform a new understanding of bounded rationality, which, in Simon's traditional conception, tends to be disembodied. Taking Simon's concept as a "conceptual yardstick," beginning at zero, Petracca sets out to measure how embodied bounded rationality can get. The concepts of embodied cognition are also fundamental for explaining the mechanisms underlying the adaptive heuristics of rational ecology. This also seems to be confirmed by Gigerenzer in his article "Embodied Heuristics." He introduces the concept of embodied heuristics, i.e., innate or learned rules of thumb, that exploit evolved sensory and motor skills to facilitate superior decisions. For example, the Gaze Heuristic solves coordination problems from catching prey in flight to catching a frisbee. Several species have adapted this heuristic to their specific sensorimotor abilities, such as vision, echolocation, running, and flying. Exaptation may explain the evolutionary mechanism that led humans to use gaze heuristics to solve tasks beyond their original purpose, for example, in rocket technology. In addition, Mastrogiorgio et al. in their article "More Thumbs than Rules: Is Rationality an Exaptation?" argue that the adaptive mechanisms of evolution are not sufficient for explaining human rationality and positing that human rationality presents exaptive origins, where exaptations are traits evolved for other functions or no function at all, and later co-opted for new uses. They propose an embodied reconceptualization of rationality—embodied rationality—based on the reuse of the perception–action system, where many neural processes involved in the control of the sensory-motor system, salient in ancestral environments, have been later co-opted to create—by tinkering—high-level reasoning processes, employed in civilized niches. They conclude by claiming the non-neutrality of biological endowment for the specification of cognitive processes.

According to Gigerenzer, the deepening of the embodied characteristic of the gaze heuristic is paradigmatic for the study of embodied cognition in relation to ecological rationality. This concept is also reaffirmed by Nordli and Todd in their article "Embodied and embodied ecological rationality: A common vertebrate mechanism for action selection underlies cognition and heuristic decision-making in humans." They argue that evolution by natural selection has produced an impressive diversity, from fish to birds to elephants, of vertebrate morphology; yet, despite the large species-level differences that otherwise exist in the brains of many animals, the neural circuits that underlie motor control exhibit a functional architecture that is virtually unchanged in every living vertebrate species. The cortico-basal-ganglia-thalamo-cortical (CBGTC) circuitry or loop regulates the embodied pursuit of goals and the learning of embedded goal-pursuit protocols that are custom-molded to fit and exploit structural regularity in the environment. It appears to facilitate motor control, trial-and-error procedural learning, and habit formation. There is evidence to suggest that this same functional circuit has been further adapted to regulate cognitive control in humans and motor control. CBGTC loop may be applied to elimination by aspect and to recognition-based heuristics and can explain the adaptive aspects in the concept of ecological rationality.

The embodied dimension of cognition is developed by Felin and Koenderink, in their article "A generative view of rationality and growing awareness." They propose the concept of generative rationality as an alternative to bounded and ecological rationality. Generative rationality steers away from conceiving rational agents as "intuitive statisticians" in favor of understanding them as "probing organisms." They argue that the statistical and cue-based logic of ecological rationality originates in a misapplication of concepts from psychophysics, such as signal detection or just-noticeable differences. They demonstrate this by considering the city-size task. Generative rationality, rather than building on statistics, builds on biology and the concepts of salience and relevance that are characteristic of the pragmatic intentionality (cues-for-something) intrinsic to perception. In addition, this has implications for understanding the emergence of novelty in economic settings. This leads them to offer a modification of Simon's "scissors" metaphor for bounded rationality. This critique of Simon's bounded rationality is also connected with that of Lee in his article "What can deep neural networks teach us about embodied bounded rationality." He argues that Simon's "bounded rationality" is the principle that humans make decisions based on step-by-step (algorithmic) reasoning using systematic rules of logic to maximize utility. This algorithmic dimension which can be equated to the Turing-Church calculus seems to provide no basis for the interactive and feedback dimension of human cognition, especially at the social level. Instead, the principle of embodied cognition suggests that human decision-makers make use of feedback mechanisms for many of their cognitive functions, including rational decision-making. In this respect, deep neural networks, which have led to a revolution in artificial intelligence, are both interactive and fundamentally non-algorithmic. Their ability to mimic some cognitive abilities much better than previous algorithmic techniques based on symbol manipulation provides empirical evidence of the power of embodied bounded rationality.

A classic way to study social interaction is the experimental use of game theory. In general, the behavioral approach to the study of games is distant from the embodied dimension. Lerique, on the other hand, in his article "Embodied rationality through the glasses of game theory: an empirical touchpoint," asks the question of how to understand embodied rationality with respect to game theory and bounded rationality. He develops a game-theoretic description of an enactive interaction arrangement (the Perceptual Crossing Paradigm—PCP) to compare with more traditional game-theoretic approaches. In this regard, he considers experimental PCP as a characterization of minimal interaction in which agents coordinate their movements without predetermined instructions. In game theory terms, this is a game of assurance, which is solved *via* the sensorimotor interactions of the agents. This allows game-theoretical approaches to be compared with enactive approaches involving participatory sense-making and embodied interaction. From this point of view, his proposal is linked to that of enactive problem-solving by Viale et al.. The sensorimotor dimension of social interaction is more explanatory than decision-making models based on information and symbolic processing psychology.

Embodied cognition has manifested its explanatory ability not only in practical problem-solving or social interaction but also in abstract thinking and reasoning. Some authors claim that conceptual features of higher-order thinking are grounded

on embodied-environmental content. There are many examples: Notions of a set seem to derive from the perception of a collection of objects in a spatial area; recursion builds upon repeated action; derivatives (in calculus) make use of concepts of motion, boundary, etc. Some authors provide a number of examples of advances in mathematics inspired by bodily and socially embedded practices: counting leading to arithmetic and number theory; measuring to calculus; shaping to geometry; architectural formation to symmetry; estimating to probability; moving to mechanics and dynamics; and grouping to set theory and combinatorics. The article of Michirev et al. "A Developmental Embodied Choice Perspective Explains the Development of Numerical Choices" is on the same wavelength. It addresses the topic of the embodiment of decision-making from a developmental perspective, where the body provides cues used in abstract choices. In particular, they consider choices in numerical settings in which the body is not necessarily needed for the solution, like the magnitude-judgment task. They propose a developmental trajectory for developmental turning points at which fingers and hands become cues. Cue validity increases through frequent and successful use over the course of development. The authors conclude that when the base-10 system is introduced, it builds upon our sensorimotor system and its cues.

Embodied cognition theories are generally opposed to dualistic models of the mind. The 4E cognition approach, i.e., embodied, enactive, extended, and embedded cognition, is holistic about the mental dimension. The embodied dimension of bounded rationality excludes the possibility of a separation between Type 1 and Type 2 processes of the mind. On the contrary, Bellini-Leite argues in his article "Dual Process Theory: Embodied and Predictive; Symbolic and Classical" that dual process theory is currently a popular theory for explaining why we show bounded rationality in reasoning and decision-making tasks. According to him, a problem for this theory is identifying a common principle that ties the features T1 and T2 together, explaining how they coordinate to express a common output. Taken together, various reasons have been given to hold this hypothesis in relation to representational format, automaticity, working memory, and speed. Psychological research must verify whether the hypotheses that the T1 responses derive from predictive processing and the T2 responses follow a classical analytic architecture are valid. Experiments with artificial intelligence can test whether this hybrid is useful and feasible. Neuroscience should be able to detect what kind of mechanisms are interrelated in classical reasoning, judgment, and decision-making tasks. Bellini-Leite hypothesis that connects to embodied cognition is that it seems likely that these mechanisms will be found not so much in the brain region but most likely in the action potentials of motor activity.

Another consideration of an epistemic type is proposed by Arfini and Magnani, in their article "Embodied irrationality? Knowledge avoidance, willful ignorance and the paradox of autonomy." They argue that knowledge avoidance and willful ignorance, although often treated as identical, should be distinguished as falling into different categories of the epistemic spectrum. They adopt an epistemic and embodied perspective to clarify the difference between these concepts. Specifically, they define willful ignorance as an irrational pattern of reasoning and, in contrast, knowledge avoidance as epistemically rational in some circumstances. They consider a variety of phenomena, such as wishful thinking, self-deception, and akrasia, and the impact of epistemic feelings, to show how knowledge avoidance can be considered a rational, autonomy-increasing strategy.

How does embodied cognition and its explanatory role in decision-making fit into the ontological representation of reality? Mousavi and Sunder in their article "Emergence and Embodiment in Economic Modeling" introduce a three-tier structure with physics at the bottom, biology at the center, and socio-psychology at the top level. Their structure characterizes the familiar modeling method of economics by specifying social-psychological preferences and goals to construct an objective function, specifying the opportunity set by constraints, and then seeking the optimal choice of action from the set. It is represented as an approach that originates in the outer part of reality with the possibility of proceeding to the biological center, which uses the principles of the physical core to derive its formalization. The three-level structure organizes principles from the physical, biological, and social sciences, proposing a new, broader, non-reductionist perspective on human behavior. The objectives of embodied cognition correspond to a method of investigation from the center to the outside.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Acknowledgments

Thanks to Herbert Simon Society for its contribution to the realization of the Research Topic.

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.

frontiers in Psychology

PERSPECTIVE
published: 20 August 2021
doi: 10.3389/fpsyg.2021.694750

A Developmental Embodied Choice Perspective Explains the Development of Numerical Choices

Alexej Michirev[1], Lisa Musculus[1] and Markus Raab[1,2]*

[1]*Institute of Psychology, German Sport University Cologne, Cologne, Germany,* [2]*School of Applied Sciences, London South Bank University, London, United Kingdom*

The goal of this paper is to explore how an embodied view can redirect our understanding of decision making. To achieve this goal, we contribute a developmental embodied choice perspective. Our perspective integrates embodiment and bounded rationality from a developmental view in which the body provides cues that are used in abstract choices. Hereby, the cues evolve with the body that is not static and changes through development. To demonstrate the body's involvement in abstract choices, we will consider choices in numerical settings in which the body is not necessarily needed for the solution. For this, we consider the magnitude-judgment task in which one has to choose the larger of two magnitudes. In a nutshell, our perspective will pinpoint how the concept of embodied choices can explain the development of numerical choices.

Keywords: embodied choice, fingers, numerical representations, development, bounded rationality, cue, magnitude-judgment task

OPEN ACCESS

Edited by:
Vittorio Gallese,
University of Parma, Italy

Reviewed by:
Michela Ponticorvo,
University of Naples Federico II, Italy
Tom Ziemke,
Linköping University, Sweden

****Correspondence:***
Alexej Michirev
a.michirev@dshs-koeln.de

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *13 April 2021*
Accepted: *09 July 2021*
Published: *20 August 2021*

Citation:
Michirev A, Musculus L and Raab M (2021) A Developmental Embodied Choice Perspective Explains the Development of Numerical Choices. Front. Psychol. 12:694750. doi: 10.3389/fpsyg.2021.694750

INTRODUCTION

Decades of theory in economics assumed *Homo sapiens* to be an agent of rationality. The surprise came when *Homo sapiens* failed to comply with these assumptions. Simon, 1972 identified those failures as the limited human ability to have complete knowledge of the world resulting in states of uncertainty. Together, the dynamic nature of the world and the limits of the human brain restrict human rationality. Simon coined these restrictions "bounds" and introduced *bounded rationality*. Half a century later, rationality is still bounded. To add to bounded rationality theorizing, we distinguish the crucial role of the body in decision making and refer to the concept of *embodied choices* (Raab, 2021).

To demonstrate embodied choices, we use the numerical setting and argue that specific body parts, such as fingers impact numerical choices; therefore, becoming *embodied* choices. Further, we consider how children use their fingers in numerical settings that create choice relevant cues, their development and impact in adulthood; therefore, taking a developmental perspective on embodied numerical choices. To assess these choices, we use the symbolic-magnitude-judgment task stemming from models of numerical cognition (for details see Knops, 2019). In these tasks, the body and its movements are not directly necessary for the choice itself, meaning that you can solve the task without an intact body, such as congenital amputees can choose among magnitudes. Showing that the body influences abstract numerical choices, therefore, would provide a strong case for the crucial role of the body, if it impacts abstract choices. Following this line of reasoning, we propose our theoretical *developmental embodied*

choice (*DEC*) perspective that explains numerical choices relying on *cues* that emerged from *finger-use* and throughout *development*.

THE THREE COMPONENTS CONSTITUTING THE DEC PERSPECTIVE

Fast-and-Frugal Heuristics: The Cues

Fast-and-*frugal heuristics* (Gigerenzer and Todd, 1999) are the first of three components of our *DEC perspective*. Fast-and-frugal heuristics adhere to bounded rationality and are cognitive shortcuts enabling fast choices by relying only on few task-relevant cues. Cue validity indicates how often the cue was successful in producing good or correct choices in similar situations. Thus, within bounded rationality, *we position ourselves within the fast*-and-*frugal heuristics camp* to explain choices and argue for a Homo heuristicus (Gigerenzer and Brighton, 2009) that considers the role of the body and as such constitutes our second theoretical component.

Embodied Cognition: Finger-Use as a Cue

Presupposing bi-directionality and interdependence of body and mind, *embodied cognition* is the second component of our DEC perspective (Wilson, 2002; Barsalou, 2008; Raab, 2017, 2021). In choice settings, the body is mostly neglected because it is not regarded as a source of information that impacts choices (Raab, 2021). Assuming bi-directionality, how would the body and its processes (not) influence cognition? Here, we link fast-and-frugal heuristics to embodied cognition by considering the body as a vital cue: A concept coined embodied choices (Raab, 2017). In numerical settings, children use fingers to help them count (Butterworth, 1999). When children notice that one of their fingers corresponds to one object, they develop an understanding of the one-to-one correspondence principle (Alibali and Dirusso, 1999). In DEC, we propose that fingers are bodily cues that gain cue validity through one-to-one finger-object correspondence. Whenever the child is confronted with a choice in a numerical setting (e.g., "Am I holding one or two cards?") it will frequently rely on its fingers and the representation thereof to choose (Butterworth, 1999). The reliance on mental representations defines the *moderate embodied cognition position* that our DEC perspective adheres to (Goldman, 2012; Raab and Araújo, 2019; overview of embodied cognition positions: Chemero, 2011; Gallagher, 2011). From this moderate position, we argue, children do not necessarily need the fingers to choose but mentally represent and use them as a cue if they made the experience that they are valid.

Development: Finger-Use Impacts Cue Validity

Capturing experiential changes, *development* is the third and final component that we integrate into our DEC perspective. In particular, we argue that the *developing body* fuels embodied choices. Across the life span, the human body undergoes different phases of greater change, especially during childhood. During this rapid development, children fine-tune their motor and cognitive skills (Adolph and Hoch, 2019). From a developmental viewpoint, we suggest that bodily growth and motor-skill development are the foundations of cognitive development (Ridler et al., 2006; Koziol et al., 2012; Gottwald et al., 2016; Musculus et al., 2021) building the basis for learning (Adolph and Hoch, 2019). In the numerical context particularly, developmental studies highlight the positive impact of finger-use in preschool years on children's numerical performance later in school (Fayol et al., 1998; Noël, 2005). Therefore, we argue that a developmental perspective on embodied numerical choices can help to disentangle how finger-use changes with age impacting cue validity of fingers, gestures, and hands and, thereby, numerical choices differentially.

Considering bounded rationality, embodiment, and development jointly, our DEC perspective pinpoints how the developing body and the sensorimotor system in childhood establish fingers as cues. We will make the case by re-interpreting existing studies and show that numerical representations and choices are embodied, developing throughout childhood and persisting in adulthood.

THE SHOWCASE OF FINGER-USE AND NUMERICAL PERFORMANCE

Rationality is as bounded as are children's negative feelings toward mathematics. Indeed, those negative feelings can cause mathematical anxiety in and out of school (Richardson and Suinn, 1972). Approximately, 17% of the population has high math anxiety (Ashcraft and Moore, 2009), which deteriorates with age (Ma and Kishor, 1997; 2–6% in secondary-school children; Chinn, 2009) and is negatively linked to mathematical performance (Foley et al., 2017). Therefore, it is crucial to underpin and promote positive impact factors favoring numerical performance early.

Numerical performance can depend on embodied factors which make mathematics not as abstract as many believe (Lakoff and Núñez, 2000). The body, in particular, the fingers, and the use thereof play a crucial role in numerical development (Barrocas et al., 2020). Here, we focus on different aspects of finger-use in numerical settings, ranging from the use of individual fingers or hands to *finger-gnosis*, and *fine motor skills* (*FMSs*). Finger-gnosis is referred to as the ability to mentally represent your own fingers. Hereby, the experimenter touches the child's two fingers without visual feedback and asks to identify the touched fingers (e.g., Penner-Wilger et al., 2009). FMSs capture how well one can move the fingers and are measured by motor-skill tests (e.g., Gashaj et al., 2019). A recent review summarizes the role of finger-use for preschool children's performance in numerical tasks (Barrocas et al., 2020). The authors conclude that finger-use strongly contributes to counting, knowledge of the number system, number-magnitude processing, and calculation ability in childhood. Crucially, other domain-general cognitive processes, such as reading ability (Noël, 2005) or vocabulary (Asakawa and Sugimura, 2014), do not seem to predict numerical performance better. How is it that specific

bodily based effects, such as finger-use, predict rather abstract numerical performance?

From the DEC perspective, the effects of finger-use on numerical performance provide a good showcase of embodied choice development for two reasons. First, the effects of finger-gnosis and FMSs can be tested using appropriate numerical *choice tasks*. An example of a numerical choice task is the magnitude-judgment task in which participants choose the larger of two magnitudes. Typically, magnitude-judgment tasks show the *distance effect*, that is, it is easier to distinguish two magnitudes that have a larger numerical difference between them resulting in faster and easier judgments (Dehaene et al., 1998). Moreover, performance on the magnitude-judgment task indicates magnitude representations and, therefore, conceptual understanding of magnitudes. Second, children use their fingers to count which has been shown to support their procedural and conceptual understanding of counting principles. Particularly, FMSs are linked to procedural counting skills that, in turn, contribute to conceptual knowledge (U. Fischer et al., 2018). Therefore, using fingers for numerical choices is *developmentally relevant* because it captures the transition from procedural to conceptual knowledge. Given fingers help bridge the transition from procedural to conceptual knowledge, finger-use might also aid abstract mathematical understanding. In the following, we will introduce our theoretical DEC perspective on the role of finger-use (*embodiment*) in the *development* of numerical *choices*.

THE DEC PERSPECTIVE ON FINGER- AND HAND-USE IMPACTING NUMERICAL CHOICES

Childhood

To illustrate our theoretical DEC perspective, first, we reinterpret the results of two exemplary longitudinal studies that depict the intra-individual development of numerical choices in children. We selected these studies because they controlled for the most neglected confounding factors regarding finger-gnosis (visual-spatial skills; Penner-Wilger et al., 2009) and FMSs (executive functions; Gashaj et al., 2019). Hereby, both studies estimated the impact of finger-use on numerical performance with a choice task, the *symbolic-magnitude-judgment task*. Second, we show that the effects of finger- and hand-use are not developmental artifacts and persist through adulthood. Third, we integrate the results of the re-interpretations in our DEC perspective.

The first study (Penner-Wilger et al., 2009) measured finger-gnosis performance by touching the children's fingers and asking them to verbally indicate the touched finger. As children were deprived of any visual-spatial feedback, the task provided a pure assessment of children's mental finger representations. The results showed that children whose mental finger representations were better in grade one (age 6.8 years) performed better in a symbolic-magnitude-judgment task in grade two. In particular, higher finger-gnosis indicated better numerical choices (by distance effect). Most importantly, finger-gnosis uniquely accounted for 10% of the variability in the distance effect.

For these findings, the authors themselves provided two different interpretations. First, they argued that there is a functional link between the mental representation of fingers and numbers established by finger-use to represent numerosities (Butterworth, 1999). From the DEC perspective, we share the interpretation that finger-use establishes a functional link between fingers and numbers. Outside and inside numerical settings, the repeated and practiced use of fingers results in improved finger sensitivity and motility, captured by finger-gnosis (Gracia-Bafalluy and Noël, 2008). Inside numerical settings, number representations become linked to fingers and become finger based. The quality of these finger-based representations constitutes cue validity: the higher the cue validity, the better numerical choices when such cues are used (e.g., in the magnitude-judgment task). Through the course of development, children learn that fingers are valid cues for numerical representations that help them make the correct numerical choices. Thus, we predict that the more frequent use of fingers for numerical choices will lead to higher cue validities attributed to fingers through the course of development. Alternatively, Penner-Wilger et al. (2009, p.524) offered that "the relation between finger and number representations may be one of identity, wherein the relation reflects a shared underlying representational form (Penner-Wilger and Anderson, 2008)." From the DEC perspective, we would not share this interpretation because our moderate-embodiment viewpoint suggests that we represent the body (i.e., fingers) and cognitive processes (i.e., numerical choices) separately but both can activate the other.

The second study (Gashaj et al., 2019) focused on FMSs in three tasks: threading beads, posting coins, and drawing trails (M-ABC-2; Petermann, 2009). The study showed that children with better FMSs performance in preschool (age 6.5 years) concurrently made better numerical magnitude judgments. Additionally, these children performed better in the number-line estimation task reflecting children's understanding of magnitudes. The authors found that the two choice tasks construct a basic numerical skill, which predicted mathematical performance in grade two (age 8 years). Interestingly, there was a significant but weak relationship between FMSs and basic numerical skills (β = 0.31). Here, basic numerical skills strongly predicted mathematical achievement in grade two (β = 0.7). The authors themselves suggest that FMSs can be considered a domain-general skill that contributes to the domain-specific numerical skills (Luo et al., 2007; Cameron et al., 2016). Further, they argue that numbers have finger-based representations (Andres et al., 2007; Penner-Wilger et al., 2007) and that fingers and numbers share cortical connections (Ardila et al., 2000). The DEC perspective specifies that FMSs grant motility to fingers that enables and promotes finger-use. In numerical settings, better FMSs enhance the cue validity of fingers because finger-use gets easier (e.g., for counting and gestures). Here, DEC links FMSs and finger-gnosis and predicts that both are valid cues as basic numerical skills benefit from the ability to move the fingers individually while assigning magnitudes to fingers (Barrocas et al., 2020).

Adolescence and Adulthood

Finger-based representations exist in children (Domahs et al., 2008) and adults (Domahs et al., 2010; Klein et al., 2011) and are therefore not restricted to a certain developmental period. In early development, children first learn to represent numerosity from 1 to 5 on one hand and then transit to represent numerosity from 6 to 10 using both hands. Such representation requires bimanual activation that is often more complex and slower than unilateral activation (Aglioti et al., 1993). Indeed, it results in a strong *five-break effect* during mental calculations in children at the age of 8.5 years showing that children deviate by exactly ±5 from the correct result (Domahs et al., 2008). Importantly, the five-break effect extends beyond childhood and is observed in westernized adults during a symbolic-magnitude-judgment task. Adults make faster choices when both numbers are represented by only one hand (e.g., a choice between 3 and 5). The choice for a set of numbers represented by two hands (e.g., 5 and 7) takes longer because the 5 is represented by one hand and the 7 by both hands (*generation hypothesis*; Domahs et al., 2010). That kind of hand-based representation occurs naturally as it splits the representations of 1–10 in two sets of fives, one for each hand. The five-break effect is systematic and strongly suggests that hand-based representations impact numerical choices. Importantly, it still persists in an adult population manifesting in mental addition (Klein et al., 2011). Together, the evidence of the five-break effect, therefore, suggests robust numerical embodiment effects of finger/hand-use and their representations that are not developmental artifacts.

One interpretation of the five-break effect is that errors in working memory occur while tracking full hands (sets of fives; Domahs et al., 2008) during calculations. The second interpretation comes from the embodied cognition viewpoint and suggests that finger-based representations moderate arithmetic performance even in numerate adults (Klein et al., 2011). Considering the empirical evidence, from the DEC perspective, we predict that both fingers and hands can serve as cues and suggest the following developmental trajectory (also see **Figure 1** for a conceptual summary). When children use fingers to represent sets they start with the understanding that one finger corresponds to one object (one-to-one correspondence). They proceed with counting (ordinality; counting objects in their order) and the representation of sets with gestures (cardinality; understanding that the last object in a set concludes the set; and for an overview of counting principles: Gelman and Gallistel, 1986). By the age of three, children spontaneously produce number gestures (Goldin-Meadow et al., 2014). By the age of 4.4 years, children accurately gesture sets of three or fewer (Gunderson et al., 2015). Our DEC perspective suggests that such a developmental trajectory creates particularly strong cues for the starting hand and starting finger(s) because the fingers are frequently used for counting and gesturing sets. When children learn to represent the full starting hand, the starting hand becomes a cue itself representing the entire set of five. Here, DEC proposes that the establishment of the five-break effect marks a developmental turning point. At the age of 8.5 years, when children intensively learn the mathematical base-10 system and start to count verbally, the five-break effect is particularly strong (Domahs et al., 2008). We argue from DEC that this is because the formerly established, and valid cue of the full hand (base-five) competes with the recently learned cue from the base-10 system. By the age of 8.5–9 years, the competition of base-five and base-10 diminishes and is accompanied by the increase of base-10 errors (Domahs et al., 2008). We would argue that this is another developmental turning point because verbal counting strategies (mostly) replace finger-based strategies. In conclusion, we propose that there is no reason for the five-break effect to exist if the abstract representation was not impacted by hand-based representations (Domahs et al., 2010). After all, advanced mathematical systems operate on a base-10 system, not base-five.

Taken together, we have gathered and reinterpreted evidence favoring numerical finger- and hand-based representations (Domahs et al., 2008, 2010; Penner-Wilger et al., 2009; Klein et al., 2011; Gashaj et al., 2019). From our DEC perspective, **Figure 1** summarizes and illustrates the suggested developmental trajectory for finger- and hand-based representations in relation to numerical choice performance. Last, we propose future directions structured by the three components of our perspective.

POINTING AT FUTURE DIRECTIONS FROM THE DEC PERSPECTIVE

The Heuristic Choice Component

In numerical cognition, participants are asked to make a choice. Finger-gnosis seems to correlate with magnitude judgments (e.g., Penner-Wilger et al., 2009). From an embodied choice viewpoint, it is unclear when and how bodily information is used for such choices. From our DEC perspective, we argue that if fingers are valid cues for a task then finger-gnosis or FMSs will be used in their order of validity (Gigerenzer and Todd, 1999). For this, we need to understand how finger-use manifests as a cue during development. Our DEC perspective suggests that individual finger-use (one-to-one correspondence), counting (ordinality), and gesturing (cardinality) all contribute to the cue validity of fingers. These specific time points could provide the basis for structured interventions to improve their validity.

The Embodied Component

From an embodied cognition viewpoint, finger-gnosis and FMSs are two distinct features. The two are distinct because they might tap into different embodied choice mechanisms (Fischer and Brugger, 2011). Specifying those mechanisms that might play along the sensorimotor-cognitive continuum and to which degree finger-gnosis and FMSs share the same processes would add to future theorizing. In general, new research may want to quantify and specify the embodied effects on numerical cognition. Currently, there is a hen-egg debate whether finger-gnosis enables finger-counting or vice versa (Soylu et al., 2018). That ambiguity, and how FMSs relate to finger-gnosis and finger-counting needs to be empirically tested in cohort-longitudinal designs. Special populations can help to quantify the amount of explained variance

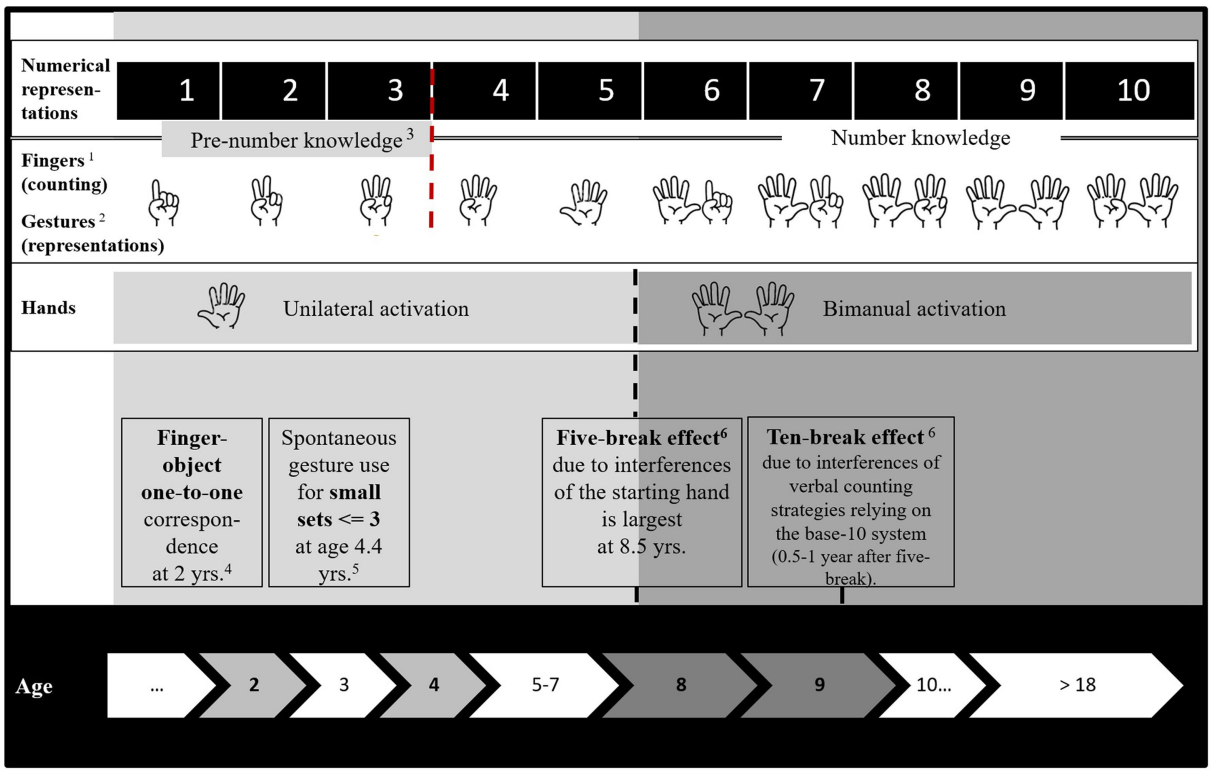


FIGURE 1 | The development of finger/hand-based numerical representations that are relevant for numerical choices. The empirical evidence summarized here stems from the following references: [1]Starting finger/hand for counting: Fischer et al., 2008; Lindemann et al., 2011; [2]Starting finger for gesturing: Wasner et al., 2015; Spontaneous gestures: Noël, 2005; Di Luca and Pesenti, 2008; Goldin-Meadow et al., 2014; [3]Number sense in infancy predicts mathematical performance at 3.5 years: Starr et al., 2013; [4]Pointing gestures: Gelman and Gallistel, 1986; [5]Accurate gesturing for sets of three and fewer: Gunderson et al., 2015; and [6]The five-break and 10-break effects at specific ages: Domahs et al., 2008.

of finger-use, finger-gnosis, and FMSs. For example, children who are born without arms and blind children who cannot rely on vision (Crollen et al., 2011) can do math. Training protocols for special-need groups that acknowledge the importance of the body may enable compensatory mechanisms for children or others at risk (Jung et al., 2015).

The Developmental Component

Fingers, hands, and bodies, as well as their use, undergo lifelong development. While nature and nurture play their role in numerical cognition, the current mathematical education lacks clear directions. It needs to establish how the interaction of finger-gnosis/FMSs and numerical cognition is mediated by age and other individual differences (Moeller et al., 2011). Other factors, such as math anxiety (Richardson and Suinn, 1972), need to be considered because they negatively impact math performance (Foley et al., 2017). As math anxiety deteriorates with age (Ma and Kishor, 1997), preschool interventions are important. Our DEC perspective predicts that repeated use of cues should provide better cues. Therefore, interventions should start early. Interventions, such as playing with cards displaying numerosity (dots and pictures) and Arabic-symbols (mobile card game: Ponticorvo et al., 2019), could improve numerical understanding and benefit future numerical performance. Engaging in physical card games should unfold the full potential of learning because it fully engages the sensorimotor system of fingers and hands. Additionally, our DEC perspective argues that both finger-gnosis and FMSs need to be trained such that the learner is able to use this bodily information as valid cues for a choice (e.g., finger-gnosis training; Gracia-Bafalluy and Noël, 2008). It is crucial to pinpoint the time windows in which finger-gnosis and FMSs training produce the best results. Current recommendations such as longitudinal studies (Moeller et al., 2012) and investigating the timing of developmental changes (Asakawa and Sugimura, 2014) should emphasize choice mechanisms beyond executive functions (Asakawa et al., 2019).

The Take-Home Message

The DEC perspective advocates that rationality is bounded, embodied, and affected by the developing body as well as the sensorimotor system. To pinpoint our perspective, we have considered the role of fingers and hands for numerical choices as a showcase. In sum, we propose a developmental trajectory for developmental turning points at which fingers and hands become cues (**Figure 1**). Cues validity increases by frequent and successful use over the course of development. We argue that at specific time points such as when the base-10 system is introduced, it builds upon our sensorimotor

system (Gallese and Lakoff, 2005) and its cues. Future research should scrutinize when and how exactly the body and bodily information should be considered to improve performance in numerical and other learning environments.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

All authors developed the developmental embodied choice perspective and the outline of the article. AM drafted the article. LM and MR edited the article.

FUNDING

The research was funded by the German Research Foundation (DFG) by RA 940/16-2 and RA 940/21-1 awarded to Markus Raab.

REFERENCES

Adolph, K. E., and Hoch, J. E. (2019). Motor development: embodied, embedded, enculturated, and enabling. *Annu. Rev. Psychol.* 70, 141–164. doi: 10.1146/annurev-psych-010418-102836

Aglioti, S., Berlucchi, G., Pallini, R., Rossi, G. F., and Tassinari, G. (1993). Hemispheric control of unilateral and bilateral responses to lateralized light stimuli after callosotomy and in callosal agenesis. *Exp. Brain Res.* 95, 151–165. doi: 10.1007/BF00229664

Alibali, M. W., and Dirusso, A. A. (1999). The function of gesture in learning to count: more than keeping track. *Cogn. Dev.* 14, 37–56. doi: 10.1016/S0885-2014(99)80017-3

Andres, M., Seron, X., and Olivier, E. (2007). Contribution of hand motor circuits to counting. *J. Cogn. Neurosci.* 19, 563–576. doi: 10.1162/jocn.2007.19.4.563

Ardila, A., Concha, M., and Rosselli, M. (2000). Angular gyrus syndrome revisited: acalculia, finger agnosia, right-left disorientation and semantic aphasia. *Aphasiology* 14, 743–754. doi: 10.1080/026870300410964

Asakawa, A., Murakami, T., and Sugimura, S. (2019). Effect of fine motor skills training on arithmetical ability in children. *Eur. J. Dev. Psychol.* 16, 290–301. doi: 10.1080/17405629.2017.1385454

Asakawa, A., and Sugimura, S. (2014). Developmental trajectory in the relationship between calculation skill and finger dexterity: a longitudinal study. *Jpn. Psychol. Res.* 56, 189–200. doi: 10.1111/jpr.12041

Ashcraft, M. H., and Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *J. Psychoeduc. Assess.* 27, 197–205. doi: 10.1177/0734282908330580

Barrocas, R., Roesch, S., Gawrilow, C., and Moeller, K. (2020). Putting a finger on numerical development – reviewing the contributions of kindergarten finger gnosis and fine motor skills to numerical abilities. *Front. Psychol.* 11:1012. doi: 10.3389/fpsyg.2020.01012

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Butterworth, B. (1999). *The Mathematical Brain*. London: Macmillan.

Cameron, C. E., Cottone, E. A., Murrah, W. M., and Grissmer, D. W. (2016). How are motor skills linked to children's school performance and academic achievement?. *Child Dev. Perspect.* 10, 93–98. doi: 10.1111/cdep.12168

Chemero, A. (2011). *Radical Embodied Cognitive Science (Reprint Edition)*: Radical Embodied Cognitive Science.

Chinn, S. (2009). Mathematics anxiety in secondary students in England. *Dyslexia* 15, 61–68. doi: 10.1002/dys.381

Crollen, V., Mahe, R., Collignon, O., and Seron, X. (2011). The role of vision in the development of finger-number interactions: finger-counting and finger-montring in blind children. *J. Exp. Child Psychol.* 109, 525–539. doi: 10.1016/j.jecp.2011.03.011

Dehaene, S., Dehaene-Lambertz, G., and Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* 21, 355–361. doi: 10.1016/S0166-2236(98)01263-6

Di Luca, S., and Pesenti, M. (2008). Masked priming effect with canonical finger numeral configurations. *Exp. Brain Res.* 185, 27–39. doi: 10.1007/s00221-007-1132-8

Domahs, F., Krinzinger, H., and Willmes, K. (2008). Mind the gap between both hands: evidence for internal finger-based number representations in children's mental calculation. *Cortex* 44, 359–367. doi: 10.1016/j.cortex.2007.08.001

Domahs, F., Moeller, K., Huber, S., Willmes, K., and Nuerk, H. C. (2010). Embodied numerosity: implicit hand-based representations influence symbolic number processing across cultures. *Cognition* 116, 251–266. doi: 10.1016/j.cognition.2010.05.007

Fayol, M., Barrouillet, P., and Marinthe, C. (1998). Predicting arithmetical achievement from neuropsychological performance: a longitudinal study. *Cognition* 68, 63–70. doi: 10.1016/S0010-0277(98)00046-8

Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex* 44, 386–392. doi: 10.1016/j.cortex.2007.08.004

Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2:260. doi: 10.3389/fpsyg.2011.00260

Fischer, U., Suggate, S. P., Schmirl, J., and Stoeger, H. (2018). Counting on fine motor skills: links between preschool finger dexterity and numerical skills. *Dev. Sci.* 21:e12623. doi: 10.1111/desc.12623

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., and Beilock, S. L. (2017). The math anxiety-performance link. *Curr. Dir. Psychol. Sci.* 26, 52–58. doi: 10.1177/0963721416672463

Gallagher, S. (2011). Interpretations of embodied cognition. *Faculty of Law, Humanities and the Arts – Papers (Archive)*. Retrieved from https://ro.uow.edu.au/lhapapers/1373 (Accessed June 30, 2021).

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognit. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Gashaj, V., Oberer, N., Mast, F. W., and Roebers, C. M. (2019). The relation between executive functions, fine motor skills, and basic numerical skills and their relevance for later mathematics achievement. *Early Educ. Dev.* 30, 913–926. doi: 10.1080/10409289.2018.1539556

Gelman, R., and Gallistel, C. (1986). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.

Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.* 1, 107–143. doi: 10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., and Todd, P. M. (1999). "Fast and frugal heuristics: the adaptive toolbox," in *Simple Heuristics That Make Us Smart*. eds. G. Gigerenzer and P. M. Todd, and The ABC Research Group (New York: Oxford University Press), 3–34.

Goldin-Meadow, S., Levine, S. C., and Jacobs, S. (2014). "Gesture's role in learning arithmetic," in *Emerging Perspectives on Gesture and Embodiment in Mathematics*. eds. L. Edwards, F. Ferrara and D. Moore-Russo (Charlotte, NC: Information Age Publishing).

Goldman, A. I. (2012). A moderate approach to embodied cognitive science. *Rev. Philos. Psychol.* 3, 71–88. doi: 10.1007/s13164-012-0089-0

Gottwald, J. M., Achermann, S., Marciszko, C., Lindskog, M., and Gredebäck, G. (2016). An embodied account of early executive-function development. *Psychol. Sci.* 27, 1600–1610. doi: 10.1177/0956797616667447

Gracia-Bafalluy, M., and Noël, M. P. (2008). Does finger training increase young children's numerical performance? *Cortex* 44, 368–375. doi: 10.1016/j.cortex.2007.08.020

Gunderson, E. A., Spaepen, E., Gibson, D., Goldin-Meadow, S., and Levine, S. C. (2015). Gesture as a window onto children's number knowledge. *Cognition* 144, 14–28. doi: 10.1016/j.cognition.2015.07.008

Jung, S., Huber, S., Roesch, S., Heller, J., Grust, T., Neurk, H.-C., et al. (2015). "An interactive web-based learning platform for arithmetic and orthography

an interactive web-based learning platform for arithmetic and orthography," in *Advances in Computers and Technology for Education–Proceedings of the 11th International Conference on Educational Technologies*, 13–22.

Klein, E., Moeller, K., Willmes, K., Nuerk, H. C., and Domahs, F. (2011). The influence of implicit hand-based representations on mental arithmetic. *Front. Psychol.* 2:197. doi: 10.3389/fpsyg.2011.00197

Knops, A. (2019). *Numerical Cognition: The Basics*. London: Routledge.

Koziol, L. F., Budding, D. E., and Chidekel, D. (2012). From movement to thought: executive function, embodied cognition, and the cerebellum. *Cerebellum* 11, 505–525. doi: 10.1007/s12311-011-0321-y

Lakoff, G., and Núñez, R. E. (2000). *Where Mathematics Comes From How: The Embodied Mind Brings Mathematics Into Being*. New York: Basic Books.

Lindemann, O., Alipour, A., and Fischer, M. H. (2011). Finger counting habits in Middle Eastern and Western individuals: an online survey. *J. Cross-Cult. Psychol.* 42, 566–578. doi: 10.1177/0022022111406254

Luo, Z., Jose, P. E., Huntsinger, C. S., and Pigott, T. D. (2007). Fine motor skills and mathematics achievement in East Asian American and European American kindergartners and first graders. *Br. J. Dev. Psychol.* 25, 595–614. doi: 10.1348/026151007X185329

Ma, X., and Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: a meta-analysis. *J. Res. Math. Educ.* 28, 26–47. doi: 10.2307/749662

Moeller, K., Fischer, U., Link, T., Wasner, M., Huber, S., Cress, U., et al. (2012). Learning and development of embodied numerosity. *Cogn. Process.* 13, 271–274. doi: 10.1007/s10339-012-0457-9

Moeller, K., Martignon, L., Wessolowski, S., Engel, J., and Nuerk, H. C. (2011). Effects of finger counting on numerical development the opposing views of neurocognition and mathematics education. *Front. Psychol.* 2, 1–5. doi: 10.3389/fpsyg.2011.00328

Musculus, L., Ruggeri, A., and Raab, M. (2021). Movement matters! understanding the developmental trajectory of embodied planning. *Front. Psychol.* 12, 1–12. doi: 10.3389/fpsyg.2021.633100

Noël, M. P. (2005). Finger gnosia: a predictor of numerical abilities in children? *Child Neuropsychol.* 11, 413–430. doi: 10.1080/09297040590951550

Penner-Wilger, M., and Anderson, M. L. (2008). "An alternative view of the relation between finger gnosis and math ability: redeployment of finger representations for the representation of number," in *Annual Cognitive Science Society Conference Proceedings*, 30, 1647–1652.

Penner-Wilger, M., Fast, L., LaFevre, J., Smith-Chant, B. L., Skwarchuck, S., Kamawar, D., et al. (2007). "The Foundations of numeracy: subitizing, finger gnosia, and fine motor ability," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29.

Penner-Wilger, M., Fast, L., Lefevre, J.-A., Smith-Chant, B. L., Skwarchuk, S.-L., Kamawar, D., et al. (2009). "Subitizing, finger gnosis, and the representation of number," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31.

Petermann, F. (2009). *Movement Assessment Battery for Children-2 (M-ABC-2) (German adaption). 2nd Edn*. Frankfurt/M, Germany: Pearson Assessment.

Ponticorvo, M., Schembri, M., and Miglino, O. (2019). "How to improve spatial and numerical cognition with a game-based and technology-enhanced learning approach," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11486 LNCS*, 32–41.

Raab, M. (2017). Motor heuristics and embodied choices: how to choose and act. *Curr. Opin. Psychol.* 16, 34–37. doi: 10.1016/j.copsyc.2017.02.029

Raab, M. (2021). *Judgment, Decision-Making and Embodied Choices*. London: Academic Publisher.

Raab, M., and Araújo, D. (2019). Embodied cognition with and without mental representations: the case of embodied choices in sports. *Front. Psychol.* 10:1825. doi: 10.3389/fpsyg.2019.01825

Richardson, F. C., and Suinn, R. M. (1972). The mathematics anxiety rating scale: psychometric data. *J. Couns. Psychol.* 19:551. doi: 10.1037/h0033456

Ridler, K., Veijola, J. M., Ivikki Tanskanen, P., Miettunen, J., Chitnis, X., Suckling, J., et al. (2006). Fronto-cerebellar systems are associated with infant motor and adult executive functions in healthy adults but not in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15651–15656. doi: 10.1073/pnas.0602639103

Simon, H. A. (1972). Theories of bounded rationality. *Decis. Org.* 1, 161–176.

Soylu, F., Lester, F. K., and Newman, S. D. (2018). You can count on your fingers: The role of fingers in early mathematical development. *J. Numer. Cogn.* 4, 107–135. doi: 10.5964/jnc.v4i1.85

Starr, A., Libertus, M. E., and Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proc. Natl. Acad. Sci.* 110, 18116–18120. doi: 10.1073/pnas.1302751110

Wasner, M., Moeller, K., Fischer, M. H., and Nuerk, H. C. (2015). Related but not the same: ordinality, cardinality and 1-to-1 correspondence in finger-based numerical representations. *J. Cogn. Psychol.* 27, 426–441. doi: 10.1080/20445911.2014.964719

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322

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.

Copyright © 2021 Michirev, Musculus and Raab. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 09 September 2021
doi: 10.3389/fpsyg.2021.710607

Embodying Bounded Rationality: From Embodied Bounded Rationality to Embodied Rationality

*Enrico Petracca**

School of Economics, Management and Statistics, University of Bologna, Bologna, Italy

Views of embodied cognition vary in degree of radicalism. The goal of this article is to explore how the range of moderate and radical views of embodied cognition can inform new approaches to rationality. In this exploration, Herbert Simon's bounded rationality is taken for its complete disembodiedness as a reference base against which to measure the increasing embodied content of new approaches to rationality. We use the label "embodied bounded rationality" to explore how moderate embodiment can reform Simon's bounded rationality while, on the opposite side of the embodied spectrum, the label "embodied rationality" is employed to explore how radical embodiment can more deeply transform the idea of what is rational. In between the two poles, the labels "body rationality" and "extended rationality" are introduced to explore how also intermediate embodiment can fruitfully inform the research on rationality.

Keywords: bounded rationality, moderate and radical embodied cognition, embodied rationality, embodied heuristics, Herbert Simon

OPEN ACCESS

Edited by:
Shaun Gallagher,
University of Memphis, United States

Reviewed by:
Vicente Raja,
Western University, Canada
Fernando Marmolejo-Ramos,
University of South Australia, Australia

***Correspondence:**
Enrico Petracca
en.petracca@gmail.com

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *16 May 2021*
Accepted: *21 June 2021*
Published: *09 September 2021*

Citation:
Petracca E (2021) Embodying Bounded Rationality: From Embodied Bounded Rationality to Embodied Rationality. Front. Psychol. 12:710607. doi: 10.3389/fpsyg.2021.710607

DISTANCE FROM SIMON'S BOUNDED RATIONALITY AS A METRIC FOR THE EMBODIMENT OF RATIONALITY

In recent years, an increasing number of works have suggested that the study of rationality can be fruitfully informed by the idea of embodied cognition in cognitive science (e.g., Spellman and Schnall, 2009; Mastrogiorgio and Petracca, 2016; Gallagher, 2018; Viale, 2019; Gallese et al., 2020). Despite the mounting interest, no agreement seems to exist, however, about the intellectual foundations of such attempts at integration—the giants upon whose shoulders an embodied notion of rationality is supposed to stand. Which extant notion of rationality, if any, is taken as a reference "to embody?" How does this relate to the strand of embodied cognition selected for the task? These questions still await systematic investigation.

The research program called "embodied bounded rationality" (Gallese et al., 2020) has chosen to stand on the strong shoulders of Herbert Simon's bounded rationality from the very name. This choice has compelling reasons worthy of being mentioned. Introduced more than 70 years ago as the first cognitive science-based approach to rationality (Simon, 1947), bounded rationality has ever since represented a vehicle for introducing cognitive science advances into the study of rationality on a rolling basis. Indeed, scholars have continued to use Simon's label over the decades for bridging the gap with cognitive science, even if ending up proposing versions of bounded rationality quite different from Simon's original one (e.g., Kahneman, 2003; see Fiori, 2011).

On the basis of the above, seemingly no better giant than Simon could have been chosen for the task of embodying rationality. Nevertheless—it needs to be recognized—supporters of embodied cognition might have something to object to. As a founding father of what is called "cognitivism" (Haugeland, 1978), Simon conceived of cognition as a fundamentally abstract and disembodied phenomenon and was as such rather skeptical of embodied cognition since its inception. His skepticism rose to the point of publicly engaging in a controversy with early proponents of embodied cognition (Vera and Simon, 1993) in which his last, peremptory words were: "there is no need [. . .] for cognitive psychology to adopt a whole new language and research agenda" (p. 46). Simon's role and significance in the history of cognitive science are crucial for our discourse, as the entire project of embodied cognition set in motion as a reaction to his cognitivism (Agre, 1993; Petracca, 2017), and the aim to go beyond cognitivist assumptions possibly remains today the only common trait of the many and diverse approaches within embodied cognition.

How, then, to reasonably recruit Simon and his bounded rationality for a project pursuing the embodiment of rationality? Doing so would require, we argue, rethinking Simon's role from that of the godfather—an unfit role for the reasons above—to that, less symbolic but more operational, of a "conceptual yardstick." What does it mean? Because of its fundamental disembodiedness, we suggest taking Simon's bounded rationality as the level zero of a virtual embodiment scale for rationality, which can then be used to assess whether and to what extent new and extant notions of rationality exhibit embodied content. In a nutshell, we suggest using the conceptual distance from Simon's bounded rationality as a metric of embodiment in the field of rationality.

Taking the distance from Simon as a measure of the embodiment of rationality is not conceptually different from what scholars of embodied cognition already do to sort different positions within their own field. Indeed, it is today customary to categorize strands of embodied cognition according to their degree of radicalism (see Goldman and de Vignemont, 2009; Gallagher, 2011), which is just another way to sort them according to their distance from cognitivism. This kind of reconstruction traditionally individuates two poles in a spectrum of positions: a "moderate" embodied cognition that aims to reform cognitivism through selective embodied add-ons and a "radical" embodied cognition that rejects cognitivism as providing no benchmark whatsoever for cognitive activity[1]. In between, a variety of positions target one or more aspects of cognitivism with the aim of either reforming or rejecting them.

The goal of this article is to explore how the range of varying-in-attitude embodied positions may inform new views of rationality. To do so, we first suggest rationalizing the use of two labels currently employed interchangeably in the literature, "embodied bounded rationality" (Viale, 2019; Gallese et al., 2020) and "embodied rationality" (Spellman and Schnall, 2009; Mastrogiorgio and Petracca, 2015, 2016; Gallagher, 2018), by tying each to a different degree of embodied radicalism. By its very name, embodied bounded rationality seems close to Simon's original notion and for this reason especially suited for pursuing a reformistic (embodied) approach. On the opposite side of the spectrum, embodied rationality may be a vehicle for radical (embodied) positions that altogether reject the central tenets of cognitivism—notably, mental representationalism and computationalism—and do not intend to use them for the study of rationality. In between these poles, we also identify two possible intermediate approaches. The one, called "body rationality," is intended for studying the body foundations of cognitive and reasoning shortcuts such as heuristics; the other, "extended rationality," is instead aimed to integrate into rationality insights from the research on extended cognition. To be clear on the increasing order of radicalism, the range goes from embodied bounded rationality through body rationality and extended rationality, and finally gets to embodied rationality. As we will show, the more radical the view of embodied cognition we adopt, the more deeply we will need to rethink the current definition of rationality.

As for what we mean by current definition of rationality, a clarification is required. Although different ideas of what is rational are lumped together under the bounded rationality banner, there is a common core to most of them: the idea that agents' rationality fundamentally lies in their successful adaptation to task environments. Adaptation is the same normative principle underlying Simon's bounded rationality, Gerd Gigerenzer's ecological rationality, and Markus Raab's motor heuristics, although they may differ in the details of adaptation. In this article, adaptation is taken as the higher-order definition of rationality that we will attempt to embody and, at radical embodied latitudes, possibly overcome. In the process, we will also discuss non-adaptive views that understand rationality more traditionally as logical or probabilistic inference, such as Daniel Kahneman's, but broadly construed adaptationism and its possible embodiment(s) will be our primary concern.

The article proceeds by introducing the four notions of rationality in increasing order of embodied radicalism or, equivalently stated, in increasing distance from Simon's bounded rationality. Section Embodied Bounded Rationality: The Reformist Embodied Approach to Bounded Rationality introduces embodied bounded rationality. Sections Body Rationality: The Bodily Roots of Adaptive Heuristics and Extended Rationality: Extended Cognition and Un-Bounded Rationality are devoted respectively to body rationality and extended rationality. Then, section Embodied Rationality: The Radical Embodied Approach to Rationality discusses embodied rationality. Section Discussion and Conclusion concludes by providing some comparative remarks.

Before moving on to the discussion, doing some justice to Simon is in order. Although Simon's thought is presented here as the quintessence of disembodiedness, we will also see that over his long career he foreshadowed, although only sketchily, some of the topics that would later be addressed by students of embodied cognition. A further reason why Simon represents, we

[1]The adjectives "weak" and "strong" are often used to refer to the two poles of embodied cognition in place of, respectively, moderate and radical (see, e.g., Tirado et al., 2018; Khatin-Zadeh et al., 2021).

argue, an inescapable reference for any contemporary discussion of rationality.

EMBODIED BOUNDED RATIONALITY: THE REFORMIST EMBODIED APPROACH TO BOUNDED RATIONALITY

Alvin Goldman has introduced the term "moderate embodied cognition" (Goldman, 2012) as an umbrella for those views of embodied cognition variously compatible with cognitivism. As Goldman claims, his position is moderate in so far as "while highlighting the pervasiveness in cognition of bodily factors, it does not invoke this as a ground for revolutionizing the methodology of cognitive science" (Goldman, 2012, p. 71). Such non-revolutionary intent dovetails quite perfectly with Simon's above-mentioned plea not to change the language and research agenda of cognitivism under pressure from embodied cognition (Vera and Simon, 1993). If moderate embodied cognition has provided a convenient banner for moderate, reformist embodied steps beyond cognitivism, the banner "embodied bounded rationality" (Gallese et al., 2020) may prove to be convenient as well, we argue, for moderate, reformist embodied steps beyond Simon's bounded rationality. This section is devoted to sketching what the reformism of embodied bounded rationality consists of.

From Abstract to Embodied Representations

Much of the debate about cognitivism revolves around the subject of mental representations and variously concerns their existence, nature, role, extent, manipulation, sufficiency, and/or necessity (see Pitt, 2020). While retaining mental representations as a requirement for cognition, the moderate embodied approach is deemed to be a "genuine rival" (Goldman and de Vignemont, 2009, p. 154) of cognitivism in so far as it challenges the disembodied nature of representations. It rejects, in particular, the existence of an abstract, amodal (machine) code of the mind which Fodor (1975) famously called "language of thought," and posits instead that mental representations are rooted in sensorial, motoric, interoceptive (e.g., visceral), and affective neural resources (e.g., Gallese, 2005; Barsalou, 2008; Meteyard et al., 2012) called, in short, "B-formats" (Goldman and de Vignemont, 2009). Currently, a debate exists between those—who may be called the "moderate moderates"—who think of B-formats as just one type of representations alongside amodal ones (e.g., Goldman and de Vignemont, 2009) and those—the "not-so-moderates"—who suggest that all representations are embodied one way or another (e.g., Gallese and Lakoff, 2005). Given this background, moderate embodied cognition can inform embodied bounded rationality suggesting the latter to set its main goal in reforming the amodal representationalism of bounded rationality without putting representations themselves into question.

Before taking this path, some preliminary grasp on the nature of representations in Simon's bounded rationality is needed. What we may expect of any representational approach to rationality is a framework in which agents manipulate their mental representations in some way considered "rational." As Fodor defined it, capturing the gist of what just said, rationality is the "organism's intelligent management of its representational resources" (Fodor, 1975, p. 169). On this view, the core of rationality lies in *how* organisms manage their representations, that is, how well they do so when assessed against a given normative principle. There is no doubt that what just said fits well Simon's idea of rationality (see Simon, 1955, in which the normative principle is represented by an agent's aspiration level), but probably this is not all there is in his thought. Simon not only understood representations as the contents of cognitive activity but also as means for meta-representing cognitive activity. In other words, representations play in Simon also a meta-representational role in so far as they allow the *simulation* of agents' cognitive activity. What Simon calls "symbols"—abstract patterns that obey the rules of formal systems (also called "physical symbol systems")—are pluripotent vehicles able to produce second-order representations, that is, representations of agents' representations (Newell and Simon, 1972). Integrating this meta-representational approach into the study of rationality, Simon and his colleague Allen Newell enunciated the so-called "physical symbol system hypothesis," according to which "a physical symbol system has the necessary and sufficient means for general intelligent action" (Newell and Simon, 1976, p. 116). In other words, Newell and Simon consider symbols necessary and sufficient conditions for any form of rational manipulation of representations and simulation thereof. Disentangling the representational from the meta-representational side of Simon's approach is crucial, we argue, to settle a persistent interpretative controversy over his thought. In the controversy, Felin et al. (2017) consider Simon assuming agents' perceptual omniscience, that is, their capacity to build potentially perfect representations of their environments. Instead, Gerd Gigerenzer and colleagues contend that Simon held a species-specific—far from omniscient—idea of perception (Chater et al., 2018, p. 803–806). To reconcile these views, one likely needs to acknowledge that in Simon's framework what is omniscient and perfect are meta-representations, not representations themselves. Omniscient meta-representations can simulate an endless variety of phenomena at the lower level of agents' representations, from species-specific cognition to any form of perceptual and reasoning bias.

This closer look at Simon is useful if we want to discuss embodied representations. On the one hand, meta-representations seem even harder to reconcile with embodiment as they are of a higher order of abstractness than mental representations. On the other hand, however, Simon's "simulationism" evokes suggestive linguistic proximity to moderate embodiment since the neural mechanisms thought to be at the root of B-formats are called "embodied simulations" (Barsalou, 2008; Gallese and Sinigaglia, 2011; Goldman, 2012). But before expecting too much of this linguistic glimmer, it is important to remark that the two ideas of simulation are quite different. While in Simon a simulation is a method to model cognitive activity, an embodied simulation is instead defined as the "[neural] reenactment of perceptual, motor, and introspective states acquired during experience with the world,

body, and mind" (Barsalou, 2008, p. 618). In other words, what in Simon is a methodological approach is instead a specific neural mechanism in moderate embodied cognition.

If embodied bounded rationality aims to follow the footsteps of moderate embodiment, it needs to leave meta-representations behind and go down the neural path. In this regard, Barsalou (2008) distinguishes between two main neural types of embodied simulation: a "cognitive simulation" and a "social simulation." In cognitive simulation, B-formats are used, among other things, to ground and structure concepts (see Harnad, 1990). For instance, the originally purely sensorial notion of "coldness" is neurally reenacted (in sensory-motor areas) to acquire the affective meaning of "emotional coldness" (e.g., Lakoff and Johnson, 1999). In social simulation instead, B-formats ground and enable social faculties such as mind-reading (this is thought to happen mostly *via* mirror mechanisms). Importantly for our discussion, theories of cognitive simulation seem to be more focused on the representational role of B-formats than theories of social simulation, which are instead more interested in B-formats' function (Gallese and Sinigaglia, 2011, p. 517). For this reason, cognitive theories appear to be more immediately relevant if the aim is to go beyond amodal representationalism. In this regard, Barsalou (1999) has introduced the framework of "perceptual symbol systems" as a way to comprehensively ground Simon's physical symbol systems into embodied simulation. Rather than being amodal, representations are on Barsalou's account entirely rooted in sensorial, motoric, interoceptive, and affective neural systems. In what follows, we will see how embodied moderatism concerns not only the nature of representations but also the very definition of rationality.

Embodied Moderatism and the Definition of Bounded Rationality

So far, the moderatism of embodied bounded rationality has consisted in retaining mental representations by reforming their nature. This section suggests that to unleash the full potential of embodied representationalism, the discussion has to take on directly the definition of bounded rationality. Otherwise, we would find ourselves in the curious situation in which embodied bounded rationality is moderately embodied but is not really bounded rationality.

Exegetical quarrels aside, the gist of Simon's bounded rationality lies in the idea that rationality is the outcome of a process of adaptation of agents' bounded representations and computations to the demands of task environments (Simon, 1955, 1956). Simon conveyed this adaptive message through the famous metaphor of a pair of scissors:

> Just as a scissors cannot cut paper without two blades, a theory of thinking and problem solving [i.e., a theory of rationality] cannot predict behavior unless it encompasses both an analysis of the structure of task environments and an analysis of the limits of rational adaptation to task requirements (Newell and Simon, 1972, p. 55).

Another way to put the metaphor is to say that the rationality of individuals does not depend on absolute cognitive resources but on the adequacy of those resources to task demands (Callebaut, 2007). Ants possess very limited cognitive resources if considered in absolute terms, but assessing them this way would prevent us from realizing that they are enough for ants to succeed—i.e., to survive—in their environment (Simon, 1996a). In Simon's view, organisms are hardwired with, but can also acquire developmentally, undemanding criteria and procedures for decision-making and problem-solving—called heuristics—that permit them to succeed in their environments. The simple but path-breaking idea that rationality lies in the use of adaptive heuristics rather than optimal procedures (Simon, 1955) continues to inspire the current studies on bounded rationality. The goal of a major contemporary strand of research, called ecological rationality, is to make a catalog of the "fast-and-frugal" heuristics used by individuals to make decisions and study in which environments they work (Gigerenzer et al., 1999).

It is crucial for our discourse to recognize that embodied simulations are resource-saving neural mechanisms the same way heuristics are resource-saving cognitive mechanisms. Evolutionarily, heuristics and embodied simulations are two sides of the same coin of adaptation. The parallel between heuristics and embodied simulations was explicitly drawn in Gallese and Goldman (1998)—the first article to hypothesize that the mirror mechanism is a more deeply-rooted neural mechanism for mind-reading than theory of mind. In fact, Gallese and Goldman call the embodied simulation occurring in the mirror system a "simulation heuristic." As they put it,

> MN [mirror neuron] activity seems to be nature's way of getting the observer into the same "mental shoes" as the target—exactly what the conjectured *simulation heuristic* aims to do. [. . .] Our conjecture is only that MNs represent a primitive version, or possibly a precursor in phylogeny, of a *simulation heuristic* that might underlie mind-reading (p. 497–498, italics added).

The meaning of heuristic in this passage is virtually the same as Simon's: embodied simulation is considered to be a species-specific (although not restricted to humans), hardwired mechanism that allows individuals to perform the complex mental faculty of mind-reading fast and frugally. Fastness would be guaranteed by the automaticity of embodied simulation, and frugality by the reuse of sensory-motor resources. On the basis of this, an analogy between Gallese and Goldman's simulation heuristic and theory of mind on the one hand, and heuristics in general and optimal criteria for decision-making and problem-solving on the other hand, does not seem too far-fetched. To complete the analogy, as using heuristics does not preclude resorting to more demanding decision-making and problem-solving mechanisms when need be, the use of simulation heuristics does not likewise preclude resort to more demanding mind-reading mechanisms whenever useful. In both cases, it is situational factors that ultimately decide on the rationality (i.e., adaptivity) of the mechanism.

Other embodied simulation mechanisms can be compared to heuristics. Consider metaphorical simulation discussed above, in which originally sensory-motor, interoceptive, and affective resources are reenacted in wider target domains (Lakoff and

Johnson, 1999). Metaphors do not merely structure concepts but, more exactly, do so in a way that saves neural and conceptual resources. In the metaphorical judgment "this person is cold" we can see in action a resource-saving mechanism that reuses a sensorial resource for affective purposes. As such, a metaphorical judgment can be considered a form of simulation heuristic the same way it is understood by Gallese and Goldman. This line of argument can also extend to decision-making, where metaphors like "heavy decision" or "balanced decision," hinging on the sensory-motor notions of physical weight and physical balance, influence decision-making in the way of making it, again, fast and frugal (see Lee and Schwarz, 2014).

Bounded rationality, however, is not only adaptation. In the post-Simonian version of Daniel Kahneman and Amos Tversky, the normative benchmark of bounded rationality is not environmental fitness but rather the axioms of logic and probability. In this approach that focuses on only one blade of Simon's scissors—limited cognition—, individuals' ability to be rational, i.e., to satisfy the axioms, is seen constantly threatened by perceptual imperfections, cognitive biases, and the use of misleading heuristics (Kahneman, 2003; Fiori, 2011). As this is the currently prevailing interpretation of bounded rationality, the reformism of embodied bounded rationality should say something about it as well. Having no interest in meta-representations and hinging on perception, Kahneman and Tversky's view more naturally than Simon's can join forces with moderate approaches to perception. Barsalou (1999), for instance, discusses how encoding information in different perceptual modalities may give eventually rise to cognitive biases. In his account, embodied simulations are far from perfectly representational as, he says, "simulations are typically partial recreations of experience that can contain bias and error" (Barsalou, 2008, p. 620). Moreover, Kahneman and Tversky introduced their own "simulation heuristic" (Kahneman and Tversky, 1982), presented as a modified version of the more famous availability heuristic. Instead of merely using the ease of retrieving past events to infer their probability (as availability heuristic does), it is the ease of constructing mental representations and counterfactuals that simulation heuristic uses to infer probability. Gallese and Goldman (1998, p. 496) explicitly acknowledge that Kahneman and Tversky's simulation heuristic, particularly when used to construct representations of others' motives and actions, may be founded on their same inter-subjective notion of embodied simulation. Recently, Kahneman has established his view of bounded rationality upon dual-process theories of cognition (Kahneman, 2011), although this new foundation has hardly rendered the approach less disembodied. Petracca (2020) discusses how the slowness and fastness of judgments and decisions can be better understood in the context of embodied mechanisms that also involve embodied simulation.

BODY RATIONALITY: THE BODILY ROOTS OF ADAPTIVE HEURISTICS

Inherent to moderate embodiment is the "neurocentric idea that cognitive states are exclusively realized in neural hardware" (Alsmith and De Vignemont, 2012, p. 5). Such neuro-centrism—often understood as plain brain-centrism—may give rise to a concern about the triviality of embodiment. If cognition were considered to be embodied merely because the brain is part of the body, this would clearly render the embodiment claim trivial. Goldman and de Vignemont (2009) say that to avoid this risk many theorists have come to understand embodiment more specifically in terms of "the whole physical body minus the brain" (p. 154), or, as Damasio (1994) called it, in terms of the "body proper." On their part, moderate theorists find likewise trivial the idea that cognition depends on features of the body, and although they admit that certain body states (such as postures) causally affect cognition, this is not deemed sufficient for considering the body proper a constitutive part of cognition (Goldman and de Vignemont, 2009)[2]. The approach that focuses on the role of the body proper for cognition, called biological embodiment (Gallagher, 2011; see Shapiro, 2004; Gibbs, 2005), represents a sort of intermediate position in the research on embodiment, halfway between neurocentric and more radical views that we will discuss in detail in section Embodied Rationality: The Radical Embodied Approach to Rationality[3]. This section is devoted to exploring how biological embodiment can inform a new approach to bounded rationality that we call "body rationality."

In a sense, Gigerenzer and colleagues' ecological rationality (Gigerenzer et al., 1999) can be considered an as much intermediate position in the field of rationality. On the one hand, supporters of ecological rationality see themselves as heirs of Simon's tradition in its "purest form" (Gigerenzer et al., 1999, p. 14), as they subscribe to Simon's adaptive, scissors-like view of rationality (see also Gigerenzer and Goldstein, 1996). Moreover, they subscribe to Simon's computational program and follow "Simon and Newell's emphasis on creating precise computational models [of heuristics]" (Gigerenzer et al., 1999, p. 26). On the other hand, however, there is a point—an important one—on which ecological rationality does not seem to follow exactly in Simon's footsteps: mental representationalism. As it has been noticed, in ecological rationality

> Mental representations [. . .] are not abandoned, but the fact that simple processing solutions exploit structure in the environment does suggest the possibility of a weaker reliance on internal models of the world (Brighton and Todd, 2009, p. 341).

While the role of mental representations in ecological rationality is the object of debate (Petracca, 2017), as its proponents continue to use them for describing cognitive activity (e.g., Gigerenzer et al., 1991), it is otherwise uncontroversial that ecological rationality is on the whole less dependent on mental representationalism than Simon's bounded rationality. This point

[2]For the idea that the body proper constitutes cognition and the difference between constitution and causality, see Shapiro (2019). The misattribution of the constitutive status to causal determinants of cognition is called "causal-constitution fallacy" (see Adams and Aizawa, 2010).

[3]Witness to the half-wayness of biological embodiment is it having common features with "physical," "organismoid," and "organismic" embodiment as defined by Ziemke (2003) but not being reducible to them.

suggests that it may be a particularly good candidate for building a bridge with biological embodiment.

One simple remark can show why this is the case. There seems to be much more truth than meets the eye in heuristics being also called "rules of thumb." In this expression, the thumb is understood as a resource of the body—a somatic device—that is used for measuring, making judgments, drawing inferences, making decisions, and solving problems (Mastrogiorgio and Petracca, 2015, 2016; Gallese et al., 2020). This remark suggests that it is possible to envisage an entire research program that studies the bodily roots of adaptive heuristics, that is, devoted to identifying those evolutionary and developmental processes that have led structural features of the human body—such as the thumb—to be used for adaptive purposes. The mildly representational view of heuristics in ecological rationality may provide a good starting point for such a new program. A program that can aspire even to reform Simon's scissors metaphor itself: if we put the body in the spotlight, the scissors of bounded rationality result no longer merely double-bladed (composed of cognition and environment) but also comprise a pivot, the body proper, which holds the blades together as an evolutionary and developmental interface (Mastrogiorgio and Petracca, 2015, 2016; Gallese et al., 2020).

In the embodied cognition literature, the body proper has mostly been understood in two ways: as a constraint and as a computational resource (Shapiro, 2019). However, these views are not mutually exclusive as inherent in the idea of a constraint is the complementary idea that it can become an opportunity in the right circumstances. Seen through the lens of the constraint/opportunity duality, it is easy to see how the thumb, along with other somatic devices, can be at the root of normative processes of rule-building. Consider, for instance, the role of somatic devices in the construction of measurement systems (Gibbs, 2005). While somatic devices are usually understood as body resources able to off-load the burden of individuals' cognition (Risko and Gilbert, 2016), they can also substitute for external resources. Over evolutionary and developmental timespans, thumbs and feet have in fact served as "on hand" embodied rulers for measuring or estimating features of the surrounding environment, eventually becoming standard units of measurement (i.e., an inch or a foot).

Proffitt and Linkenauger (2013) provide a productive framework for understanding the role of the body proper in cognition. What they call "phenotype" is deemed to include the three dimensions along which the body proper shapes cognition: the morphological, physiological, and behavioral dimensions. The way the body shapes measurement systems in the example above concerns prominently body morphology that, although being traditionally the least explored embodied dimension, is the one specifically investigated by Proffitt and Linkenauger. In particular, they study the role of body morphology in perception, and do so in two ways: in terms of morphological invariance (e.g., considering five-fingered hands as morphological invariants of the human species) or in terms of individual differences (e.g., considering hands' morphological variations between individuals). Interestingly for our argument, there is a distinct pragmatist undertone in Proffitt and Linkenauger's investigation as they emphasize how body morphology, along with the other phenotypical dimensions, modulates perception in ways that subserve agents' situational goals. For instance, in a task that involves grasping, they say that "apparent distances are scaled with morphology, and in particular, to the extent of an actor's reach or the size of his or her hand" (p. 172).

Gigerenzer's ecological rationality is naturally suited to be understood through the lens of biological embodiment as in some (rare) cases it is already biologically embodied. In the vast repertoire of fast-and-frugal heuristics, Gigerenzer (2007) discusses the "gaze heuristic," which applies whenever individuals try to intercept an object, such as a ball, flying in the air. To trace mathematically the trajectory of the ball one should virtually compute differential equations, which is almost impossible to do (just literally) on the fly. To make the catching job done, the gaze heuristic provides alternative fast-and-frugal rules: "[f]ix your gaze on the ball, start running, and adjust your running speed so that the angle of gaze remains constant" (Gigerenzer, 2007, p. 7). No need to say that these rules are but rational reconstructions of what individuals unknowingly do every time they try to catch a flying ball. The gaze heuristic belongs to that class of fast-and-frugal heuristics that Markus Raab has recently called "motor heuristics" (Raab, 2017), which, concerning specifically the use of the body proper, are biologically embodied by definition.

There are, however, more subtle (but no less pervasive) forms of biological embodiment of fast-and-frugal heuristics. Consider, for instance, the "theory of prominence" (Albers, 2002) and the "QuickEst" heuristic (Hertwig et al., 1999), two judgment processes that exploit so-called "prominent numbers" (1, 2, 5, 10, 20, 50, 100, etc.) in the 10-based number system for fast-and-frugal numerical estimations. Here we are not interested in whether numerical prominence leads to reliable estimates or estimation biases, but in the origins of prominent numbers[4]. Again, fingers and hands feature prominently in this discussion. It is well known, but sometimes not sufficiently appreciated, that the 10-based number system originates in counting processes based on the 10 fingers of the hands (Gibbs, 2005). This leads to plausibly hypothesize that numerical accessibility and prominence have precise roots in body morphology (see also Lakoff and Núñez, 2000). As another instance, consider the 1/N heuristic (Gigerenzer and Gaissmaier, 2011), an evaluation and choice criterion that "weights" different options equally. The very ideas of weighting, pondering, and balancing when used in judgment and decision-making are, as seen, instances of embodied metaphors (Lee and Schwarz, 2014). Rarely, however, it is asked where the accessibility and cognitive relevance of the idea, say, of equal-weight comes from. A biologically embodied answer is that it originates in the morphological symmetry of the body, in the vestibular system, and in the sense of balance it controls (Gibbs, 2005). Similar considerations can be extended to entire classes of heuristics with the aim of uncovering their bodily, and particularly morphological, roots.

[4]When numbers are understood as signifiers of numerousness, they are called numerals. It is plausible that numerals rather than numerousness itself trigger the behavioral responses associated with prominent numbers (see Mastrogiorgio and Petracca, 2014).

EXTENDED RATIONALITY: EXTENDED COGNITION AND UN-BOUNDED RATIONALITY

Biological embodiment is not alone in populating the conceptual space between embodied moderatism and radicalism. The approach of extended cognition pioneered by philosopher Andy Clark also contends for that space. This raises the issue of relative positioning: which one is more leaning toward radicalism? As extended cognition posits that cognitive processes are neither bounded to the brain nor even to the body but also realize through resources of the environment—so going beyond biology as a requirement for cognition—this might suffice, we argue, to consider extended cognition more radical than biological embodiment. However, although it is sometimes presented as a radical position *per se* (e.g., Wilson and Clark, 2009), there are reasons to doubt that this is the case. Clark is well known not to reject mental representationalism and computationalism as he attempts to retain them—however limiting their extent (Clark and Toribio, 1994)—in an integrated framework known as "extended functionalism" (Clark, 2008; Kiverstein and Clark, 2009). According to this framework, what renders a resource cognitive is not its spatial location, inside or outside the body, but its function in the cognitive system (Clark and Chalmers, 1998; Clark, 2008). On this view, notebooks and hippocampal neurons can be seen as functionally equivalent to the extent they both support memory. This section explores how extended functionalism can inform bounded rationality and uses the banner "extended rationality" for the task.

Luckily, Clark has completed much of the preparatory work for us. Particularly in the early stages of extended functionalism, he discussed at length how his view might relate to Simon's. Unusually for a post-cognitivist scholar, he did not criticize Simon for his cognitivism but was even open to recognizing him as a forerunner of extended cognition. "Simon saw, very clearly," Clark says, "that portions of the external world often functioned as a non-biological kind of memory. He thus saw a deep parity (parity, not identity) that can obtain between external and internal resources" (Clark, 2001, p. 139). Clark adds, however, that instead of extending the notion of self to include external resources, "Simon chose to go the other way" (Clark, 2001, p. 139), that is, he shrank the self so much that functions realized through external resources, like memory, ended up being out of its domain. When Clark goes on discussing Simon's bounded rationality, it is only coherent that he considers this concept "probably the first step" (Clark, 1998, p. 184) in the direction of recognizing the importance of external resources for rationality, yet an "insufficiently radical" (Clark, 1998, p. 243) step[5].

According to Clark, there are two main routes for embodying rationality. One is what he calls "biological cognitive incrementalism," a view "according to which full-scale human rationality is reached, rather directly, by some series of tweaks to basic biological modes of adaptive responses" (Clark, 2001, p. 122). As an instance of a basic biological adaptive response, one may think of the already mentioned intuitive use of the thumb for making spatial inferences, an intuitive method which can be eventually "tweaked" into becoming a formal heuristic (e.g., Wong, 2006). As such, biological cognitive incrementalism seems to be in full continuity with biological embodiment and body rationality, and it is not by chance that Clark discusses Gigerenzer's ecological rationality right in this context (Clark, 2001, p. 130). An alternative route—clearly Clark's favorite—for the embodiment of rationality goes instead down the path of extended functionalism. As human cognition is increasingly constituted—not just enabled—by external technological artifacts, the boundaries between biological and non-biological cognitive requirements become blurred. This acknowledgment, Clark suggests, should accordingly turn the discussion of rationality from biological to non-biological cognitive incrementalism.

In recent years, the research on extended cognition has shifted its focus from the study of functional "parity" (e.g., between notebooks and hippocampal neurons) to that of functional "complementary." Functional complementarity means that external resources are not only employed as substitutes for internal resources but can also integrate with the latter in order to enhance individuals' overall cognitive capacity. The subtitle of Menary's (2007) book *Cognitive Integration: Mind and Cognition Unbounded* explicitly suggests that by using external resources cognition can become "unbounded." Rehearsed in the domain of rationality, Menary's unboundedness seems to be rather in contrast—the opposite actually—to Simon's cognitive boundedness, and therefore induces one to wonder whether it is the case that cognitive complementarity leads in the end to an unbounded notion of rationality. As the idea of cognitive unboundedness seems suspiciously reminiscent of the omniscience of rational choice theory that Simon fought (with merit) throughout his career, it is of utmost importance to clarify this point in what follows.

The risk of mistaking the unboundedness of extended cognition as a restoration of rational choice theory occurs only if we adopt a non-adaptive framework. Consider Kahneman's non-adaptive bounded rationality, according to which humans would be fully rational if only they did not use misleading heuristics and were not ridden with cognitive biases. In Kahneman's framework, it is quite natural to think of external resources—understood as "cognitive artifacts" (see Hutchins, 1999)—as means to fix cognitive imperfections and get a step closer to the desired omniscience. But in Clark's framework omniscience does not play any role, not even as a benchmark (Clark, 1998). If it is true that coupled with external resources memory and other cognitive faculties can become virtually limitless (instead of a notebook, think of the far greater potential of a smartphone), the point Clark and other theorists of extended cognition would still raise is: is omniscience desirable from an ecological point of view? Or, is omniscence even meaningful once we come to understand what cognitive faculties are really for (see Glenberg, 1997)? This ecological tone, Arnau et al. (2014) have recently claimed, brings extended cognition quite close to ecological

[5]Instances of Simon's "extended" approach can be found in his study of organizations, where he said that "organizations can expand human rationality" (Simon, 1996b, p. 72).

rationality: in both perspectives, it is rightly noticed, it is the environment and the task at hand that ultimately decide whether more cognitive capacity is beneficial or not. However, although ecological rationality and extended cognition undeniably share the ecological viewpoint, the way they deal with environments makes the two perspectives hard to integrate. The view of adaptation supported by theorists of extended cognition is far from the static and passive process envisioned by Simon and Gigerenzer. Clark's idea of adaptiveness is fundamentally active, so much that another name for his extended approach is "active externalism" (Clark and Chalmers, 1998). In Clark's view, individuals use *actively* the resources of environments to get a step closer to the kind of adaptive, circumstantial unboundedness envisioned by Menary (2007).

Importantly in the active kind of adaptation, individuals do not merely use environmental resources but altogether transform environments. In a rather explicit passage, Clark emphasizes such a constructivist side of his approach when he says that "[o]ur brains make the world smart so that we can be dumb in peace" (Clark, 1997, p. 180). One example he discusses in this regard is that of markets, which Clark sees as constructed environments that "scaffold" agents' cognition and foster their economic rationality. Following this line of argument, Clark's environmental interventionism has been explicitly related to niche constructionism by Sterelny (2004) and to autopoiesis by Di Paolo (2009). In this latter view, agents are considered engaging in a constructive, dynamic relationship with the environment in a way that makes life itself self-sustaining. Rather than with better known bounded and ecological rationality, the constructionism of extended cognition may more easily dovetail with what Shira Elqayam has called "grounded rationality" (Elqayam, 2011)[6]. In grounded rationality, environments acquire their normative status—i.e., ultimately decide whether a cognitive process or behavior is rational or not—only after being constructed as epistemic niches.

EMBODIED RATIONALITY: THE RADICAL EMBODIED APPROACH TO RATIONALITY

Three Challenges for Embodied Rationality

To define radical embodiment, we are faced with the same conceptual difficulties encountered in defining other embodied views and, possibly, even more. A commentator has remarked that "what is common to all versions of radical embodiment is that an agent's possession of her bodily anatomy is taken to be a constitutive part of her mind, in violation of neurocentric assumptions" (Jacob, 2016, p. 44), a definition that as such would also fit what has been called biological embodiment. Although biological embodiment is certainly required for embodied radicalism, it is, however, not sufficient for it. More specifically, Chemero (2011) defines radical embodiment as "the thesis that cognition is to be described in terms of agent-environment dynamics, and not in terms of computation and representation" (p. x). Thus, if we are looking for the core of radical embodiment, anti-representationalism and anti-computationalism are the places to look. Defined this way, we can appreciate how diametrically opposed radical embodiment is to Simon's representational and computational view of cognition. In this section, we use the banner "embodied rationality" (Mastrogiorgio and Petracca, 2016; Gallagher, 2018) for exploring the idea of rationality informed by radical embodiment that, as such, results the most conceptually distant from Simon's.

In pursuing embodied rationality, we face at least three challenges not encountered before. The first, and arguably the main one, is that embodied rationality has no benchmark of rationality to refer to, or, stated otherwise, no extant idea of rationality to build upon, reform, complement, or refound. While Simon's and Gigerenzer's views have been taken so far as conceptual platforms to be provided with new embodied foundations, embodied rationality has nothing preexisting to embody. To find an extant notion virtually compatible with radical embodiment, we should look for a kind of non-computational and non-representational approach to rationality, one that, as Rolla (2019) says, does not equate rationality with reasoning. But, as he adds, we have none of this sort:

> Even the more unorthodox view known as Ecological Rationality, proposed for instance by Todd and Gigerenzer [. . .], holds that a theory of rationality should describe the heuristic reasonings used by real agents, where heuristics involve following certain environmental cues and ignoring excessive information—which is a matter of reasoning nonetheless (p. 2).

For this reason, embodied rationality bears the privilege and the burden of writing its own history. The carte blanche it is given includes, importantly, also the liberty not to follow the usual framework of naturalistic approaches to rationality, adaptationism (see Neemeh, 2021), and therefore to conceive an altogether new definition of what is rational.

The second challenge concerns the intrinsic plurality of the radical field. If it is true that any embodied approach is internally plural, radical embodied cognition is even more plural. Gallagher (2009) has traced the precursors of radical embodiment to American pragmatism, classical phenomenology, and Wittgenstein's philosophy of language, to which can be added, more recently, ecological psychology, situated robotics, dynamical systems theory, and phenomenology-inspired neuroscience. And the list could be easily enlarged. In brief, anti-representationalism and anti-computationalism are only the common denominators of an array of radical positions that can variously inform embodied rationality.

Gallagher's rich list points to the third challenge for embodied rationality. Simon was, among other things, also an economist (of Nobel fame), and much of the debate over bounded rationality has been held in economics. Much of Gigerenzer's fame is also due to economics, for the controversy with Kahneman over the psychological foundations of behavioral economics. And even Clark's extended functionalism crossed paths, although briefly, with economics (Clark, 1997). In striking contrast, drawing upon such varied disciplinary backgrounds and having no extant

[6]Hatchuel (2001) sees constructionism as a possible route, among others, to expand what he calls "the unfinished program of Herbert Simon."

notion of rationality to refer to, embodied rationality seems disciplinary disconnected from economics. Again, the privilege and burden of freedom.

In what follows, we will explore the idea of embodied rationality focusing in particular on two aspects. First, we will see how anti-representationalism and anti-computationalism—taken singularly or together—may radically transform the understanding of rationality. Second, we will see how embodied rationality can propel itself into unchartered routes discussing the view of rationality from the first-person perspective.

Rationality Without Representations and/or Computations

As Rolla (2019) has put it, the challenge raised by radical embodiment to students of rationality is to figure out what "rationality without reasoning" means, which is tantamount to figuring out what a notion of rationality without mental representations and computations looks like. Stating the challenge this way may make one wonder whether representations and computations necessarily come as a bundle or might be thought of separately. In the latter case, radical versions of rationality based on computations but no representations, or, vice versa, on representations but no computations, could be envisaged. Indeed, computations and representations are usually understood as two sides of the same (cognitivist) coin, as it seems hard to think of representations that are not manipulated somehow or computations that have no content[7]. But following the incremental spirit of this article, we will attempt to disentangle their differential contribution to rationality.

Unlike Simon's approach in which representations and computations are equally central, in Kahneman's bounded rationality representations seem to feature more prominently than computations. In associative forms of judgment called System 1 (Kahneman, 2011), it is the content of the representation, and relatedly the semantic proximity of one representation to another, that guide the judgment. In addition, it is curious but telling that Kahneman's representationalism appeals to Freud's associationist concept of a symbol rather than to Simon's idea of symbols as objects of computation (Kahneman, 2011, p. 56; see Petracca, 2017). While associationist forms of reasoning are not bound to irrationality as Kahneman thinks, in so far as fast associations can be adaptive in the right context (see Gigerenzer, 2007), the point we wish to make is that they seem in any case to privilege the semantics of representations over the mechanics of computations.

Rodney Brooks has been one of the first to follow the alternative route, investigating how representations are not necessary for simple forms of intelligent behavior (e.g., Brooks, 1991). As a roboticist, he designed a class of goal-driven robots that, as has become customary to say, used "the world as their own best model." This means that situated interactions with their proximal environments permitted Brooks' machines to accomplish their tasks without relying on representations—such as maps—of the environment. Brooks' robots (one of which was provocatively christened Herbert) were meant to be living falsifications of Newell and Simon's physical symbol system hypothesis, that is, the hypothesis that representations are necessary and sufficient conditions for intelligence (Newell and Simon, 1976). Of course, Brooks' robots were not free of computations, as Simon was eager to rebut (Vera and Simon, 1993), but they were not the kind of serial, centralized, vertically integrated, and content-based forms of computation that cognitivists advocated.

In Brooks' framework, the step from a non-representational form of intelligence to a non-representational form of rationality is not very long. Discussing Brooks' cognitive design, Susan Hurley explicitly speaks of rationality:

> *Rationality* might emerge from a complex system of decentralized, higher-order relations of inhibition, facilitation, and coordination among different horizontal layers, each of which is dynamic and environmentally situated [. . .] *Rationality* reconceived in horizontally modular terms is substantively related to the environment. It does not depend only on internal procedures that mediate between input and output, either for the organism as a whole or for a vertically bounded central cognitive module. Rather, it depends on complex relationships between dedicated, world-involving layers that monitor and respond to specific aspects of the natural and social environment and of the neural network, and register feedback from responses (emphasis added, quoted in Rolla, 2019, p. 4–5).

Yet, these remarks notwithstanding, it is not easy to single out non-representational forms of rationality in the biological domain that do not also qualify as forms of reasoning (remember, absence of reasoning is Rolla's requirement for radical embodied rationality). To have a sense of this difficulty, consider a heuristic discussed by Gigerenzer as a case of fast-and-frugal heuristic:

> To measure the area of a candidate nest cavity, a narrow crack in a rock, an ant has no yardstick but a rule of thumb: Run around on an irregular path for a fixed period while laying down a pheromone trail, and then leave. Return, move around on a different irregular path, and estimate the size of the cavity by the frequency of encountering the old trail (Gigerenzer and Brighton, 2009, p. 107).

(If real ants did not already use such an embodied heuristic, it might very well have been devised by Brooks for his robot-ants). Now, the question is: does this heuristic involve any reasoning? While it likely does not involve representations, this is not enough for disqualifying it as a form of reasoning. In fact, commenting on this very example, Arnau et al. (2014) maintain that "these problem-solving activities qualify as instances of genuine reasoning" (p. 57). And Rolla (2019), as seen, seems to maintain that any use of heuristics qualifies ipso facto as reasoning. Now, if we agree that non-representational heuristics such as ants' qualify as (minimal) forms of reasoning, we should

[7] Miłkowski (2013) has recently substituted Fodor's famous "no computation without representation" with his own "no representation without computation." Although the two mottos reach the same conclusion, they express different nuances.

ask what we need more of (or, perhaps, less of) to achieve rationality without reasoning.

The answer to this question usually given by theorists of radical embodiment is one: extended dynamics. To qualify as a genuine instance of non-reasoning in the radical embodied sense, cognitive processes leading to rational outcomes need to be understood not only as non-representational but also as dynamically extended. This means, in a nutshell, that processes leading to rationality originate in the continuous *interaction* between agents and their environments. On this view, neither internal nor external resources alone would be enough to explain the emergence of a rational outcome, and interaction becomes the new explanatory cornerstone (e.g., Gallagher, 2017). It may seem paradoxical that Simon, the advocate of representations as requirements for intelligence, provided a good example of interactionist explanation that, intriguingly, concerns once again ants. Simon (1996a) discusses the case of the path made by an ant on the sand and wonders why the path is not regular:

> [The ant] has a general sense of where home lies, but he cannot foresee all the obstacles between. He must adapt his course repeatedly to the difficulties he encounters and often detour uncrossable barriers. His horizons are very close, so that he deals with each obstacle as he comes to it; he probes for ways around or over it, without much thought for future obstacles. It is easy to trap him into deep detours. Viewed as a geometric figure, the ant's path is irregular, complex, hard to describe. But its complexity is really a complexity in the surface of the beach, not a complexity in the ant (p. 51).

Although Simon made this example to emphasize the sometime prominence of the environment (the beach, in this case) over agents' cognition in the explanation of complex behavior, his argument can be plausibly understood as if he meant that neither features of the ant nor those of the beach (for different sorts of insects could produce different trajectories) can explain alone the irregular path. Using the words of radical theorists, it can be said that the ant-beach pair forms a "coupled system." Importantly, according to the radical embodied position, the ant-beach system does not merely explain the ant's path, but altogether forms an autonomous cognitive system that constitutes navigational abilities in that circumstance. Compared to the extended notion of constitution encountered in the discussion of the extended mind (section Extended Rationality: Extended Cognition and Un-Bounded Rationality), the idea of constitution held in radical embodiment is more specifically of the interactive, dynamic kind (see Gallagher, 2017).

Rationality in the First-Person Perspective

If the hypothesis of rationality without reasoning seems outlandish enough, this section discusses the possibly more challenging hypothesis that rationality concerns the first-person rather than the third-person perspective. This issue was at the center of an epoch-making controversy in the 1970s between Simon, a staunch advocate of third-person-ism, and supporters of first-person-ism led by phenomenologist Hubert Dreyfus. Dreyfus' book *What Computers Cant' Do* (Dreyfus, 1972) has represented one of the most radical criticisms ever raised against Simon's thought, one that Simon's biographer says "left him angry, sad, and uncharacteristically silent" (Crowther-Heyck, 2005, p. 271). It took 20 years before Simon felt compelled to reply to Dreyfus' sort of criticism (Vera and Simon, 1993), when in the 1990s phenomenology was becoming one of the pillars of embodied cognition (see Petracca, 2017) One of the main points raised by phenomenologists concerned the impossibility to assess rationality objectively, "from the outside," or, equivalently, from a third-person point of view. Famously stating that "intelligence must be situated" (Dreyfus, 1972, p. 62), Dreyfus introduced the idea of a "situation" as a construct critical and alternative to that of "context." While contexts are objectively identifiable states of the world, situations are the outcome of a process of sense-making that can only be carried out by individuals. As Hans-Georg Gadamer put it,

> [t]o acquire an awareness of a situation is, however, always a task of particular difficulty. The very idea of a situation means that we are not standing outside it and hence are unable to have any objective knowledge of it. We are always within the situation and to throw light on it is a task that is never entirely completed (quoted in Winograd and Flores, 1986, p. 29).

What does this mean for rationality? Using the words of another phenomenologist, Maurice Merleau-Ponty, situatedness means that "the world and reason are not problematic [. . .] they are mysterious" (Merleau-Ponty, 2002, p. xxiii). Saying so, Merleau-Ponty appeals to the concept of "mysteriousness" as a way to counter the usual understanding of rationality popularized by Simon in terms of "problematicness,"[8] according to which rationality is equivalent to the capacity of identifying unambiguous procedures to solve as much unambiguously identified problems (see Newell and Simon, 1972). Mysteriousness would instead emphasize the interactive, tentative, and above all non-pre-specifiable process of dealing with the world. On this view, conferring a behavior or an outcome the rationality status cannot be done on a third-person basis but becomes an eminently inter-subjective process, a "we" process. As Merleau-Ponty claims,

> rationality is precisely proportioned to the experiences in which it is disclosed. To say that there exists rationality is to say that perspectives blend, perceptions confirm each other, a meaning emerges. But it should not be set in a realm apart, transposed into absolute Spirit, or into a world in a realist sense (Merleau-Ponty, 2002, p. xxii).

Gallagher (2018) has recently reintroduced the distinction between mystery and problem in the study of rationality, emphasizing that rationality "[is] not an observational or spectatorial stepping back that detaches from the situation to frame the world in abstract concepts" (Gallagher, 2018, p. 91). An important point Gallagher makes in this regard is that concepts such as problem-solving, reasoning, etc. should not be banished from the vocabulary of rationality, but rather reconceived:

[8]The distinction between "mystery" and "problem" was first made by philosopher Gabriel Marcel (see Gallagher, 2018).

TABLE 1 | Main features of the embodied approaches to rationality.

	Degree of embodiment	Normative adaptationism	Representations	Heuristics
Embodied bounded rationality	Moderate	Yes[a]	Yes	Embodied cognitive heuristics
Body rationality	Intermediate	Yes	Weak	Body-based heuristics
Extended rationality	Intermediate	Yes, but also normative constructivism	Yes, but not always necessary	Not the main source of rationality
Embodied rationality	Radical	No	No	Not the main source of rationality

[a]*Except in the case Kahneman's approach is embodied.*

> the alternative [to the classical view of cognition and reasoning] is to think of mental skills such as reflection, problem solving, decision making, and so on, as enactive, non-representational forms of embodied coping that emerge from a pre-predicative perceptual ordering of differentiations and similarities (Gallagher, 2018, p. 86–87).

Here we meet again dynamics as a fundamental ingredient of radicalism (Chemero, 2011; Gallagher, 2017). If problem-solving is, as radical theorists insist, non-representational and non-pre-predicative—that is, if the range of solutions to problems cannot be predicated (let alone predicted) before engaging with the situation—interaction represents the only way for agents to cope with the complexity of the world and give proof of their skills.

As an example of embodied rationality, Gallagher discusses the rationality intrinsic in the use of the hand. Having encountered hands before in our discussion, we can see how now the tone is quite different. The hand seems to show a rationality of its own:

> Consider, that there is a rationality that is implicit in the hand. […] As an agent reaches to grasp something, the hand automatically (and without the agent's conscious awareness) shapes itself into just the right posture to form the most appropriate grip for that object and for the agent's purpose. […] It is sometimes the case that very smart hand-brain dynamics take the lead over a more conceptual, ideational intelligence. For example, a patient with visual agnosia who is unable to recognize objects, when shown a picture of a clarinet, calls it a "pencil." At the same time, however, his fingers began to play an imaginary clarinet (Gallagher, 2018, p. 88).

As another, quite different instance of embodied rationality, Gallagher et al. (2019) show that the dynamic perspective can be employed to explain the emergence of institutional forms of coordination such as markets, thus giving a radical twist to Clark's example of markets as extended forms of rationality (Clark, 1997). The variety of these examples just hints at the wide empirical applicability of the dynamic viewpoint.

It is of utmost importance to remark that insisting on dynamics and the first-person perspective does not take the inquiry of rationality out of the naturalistic heaven in which Simon placed it. In so far as phenomenology and other radical embodiment approaches are not only compatible but also an active part in the construction of a newly naturalized cognitive science (Gallagher and Varela, 2003), the same new naturalistic outlook can be transferred, we argue, into the naturalistic study of rationality.

DISCUSSION AND CONCLUSION

This article has explored how increasingly radical views of embodied cognition may reform or even transform the idea of bounded rationality. To this purpose, four new embodied notions of rationality have been proposed (in increasing order of radicalism): embodied bounded rationality, body rationality, extended rationality, and embodied rationality. Although at this exploratory stage it would be too ambitious to provide self-contained definitions of these notions, their main features are displayed and juxtaposed in **Table 1**. In particular, they are arranged according to four criteria: degree of embodied radicalism, adherence to adaptationism as normative framework, reliance on mental representations, and view of heuristics. This juxtaposition allows us to propose some comparative remarks. The first remark concerns the extent to which the different notions of rationality question adaptation as a normative principle. While embodied bounded rationality (except in the case of embodiment of Kahneman's approach) and body rationality fundamentally retain Simon's adaptationist framework[9], extended rationality seems more compatible with normative constructivism, which considers agents playing an active role in establishing normative standards through environmental manipulations. As for embodied rationality, it more resolutely goes down the post-adaptationist path (see Neemeh, 2021), although adaptation still seems to play a central role in Rolla (2019). For what concerns representations (and computations), embodied bounded rationality proposes to retain them by reforming their abstract nature, while body rationality and extended rationality rely on attenuated or intermittent (i.e., depending on the cognitive task) forms of representationalism. Downright rejection of representationalism is, instead, the trademark of embodied rationality. In this context, it is worth mentioning that a computational view based on the so-called "free energy principle" has recently tried to reconcile representationalism and anti-representationalism (see Constant et al., 2021), although it is doubtful whether it can be the last word on such a controversial matter. Finally, another comparative criterion concerns heuristics. Coherently with all the threads of bounded rationality (Kahneman's included), embodied bounded rationality and body rationality have heuristics as their main objects of inquiry, the only difference between them being that

[9]It should be remarked that adaptation is not always supported by adaptationist mechanisms. In embodied simulation, for instance, the way extant neural resources are reused for different functions is a case of "exaptation." Nonetheless, the extent to which the various mechanisms lead to environmental fitness remains the ultimate normative criterion.

the former focuses on cognitive and neural heuristics, while the latter on body-based heuristics. Extended and embodied forms of rationality, in contrast, do not put special emphasis on heuristics as sources of rationality.

Another point deserves discussion: what kind of evidence can be brought in support of one or the other form of rationality? It seems reasonable, at this early stage, to consider one form of rationality more well-supported than the other depending on the empirical success of the underlying embodied approach. If this criterion is adopted, embodied bounded rationality and body rationality seem to be currently in a better position. But the study of rationality opens, we argue, an entirely new terrain on which the embodied approaches can compete. More radical embodied approaches can prove their merit in this new field in particular the more we switch from the explanation of phenomena at the individual level to those at the collective and social level. Just to make one instance, the study of rationality in institutional settings like markets seems to be addressable in new ways through radical embodied approaches (see Clark, 1997; Gallagher et al., 2019; Petracca and Gallagher, 2020).

Some words should finally be spent on one giant in this article: Herbert Simon. Katsikopoulos and Lan (2011) have argued with reason that one way or another all scholars interested in the naturalistic study of rationality "labor under Herbert Simon's spell" (p. 728). Our take in this article is that, however, when it comes to embodiment, Simon's spell may be a bit less enchanting. This is why rather than taking Simon as a source of inspiration for all the embodied approaches to rationality, we have emphasized his suitability as a reference base for measuring the embodied content of different rationality proposals. The most fruitful way for embodied approaches to stand upon Simon's shoulders is, we argue, dialectical and interactive, taking such a giant of thought as a reference with whom to be in constant dialogue in the spirit of advancing the study of rationality.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I wish to thank Shaun Gallagher and the two reviewers for constructive and helpful comments. The usual caveats apply.

REFERENCES

Adams, F., and Aizawa, K. (2010). *The Bounds of Cognition, 2nd Edn*. Boston, MA: Blackwell Publishers. doi: 10.1002/9781444391718

Agre, P. E. (1993). The symbolic worldview: reply to Vera and Simon. *Cogn. Sci.* 17, 61–69. doi: 10.1207/s15516709cog1701_4

Albers, W. (2002). "Prominence theory as a tool to model boundedly rational decisions," in *Bounded Rationality: The Adaptive Toolbox*, eds G. Gigerenzer and R. Selten (Cambridge, MA: MIT Press), 297–317.

Alsmith, A. J. T., and De Vignemont, F. (2012). Embodying the mind and representing the body. *Rev. Philos. Psychol.* 3, 1–13. doi: 10.1007/s13164-012-0085-4

Arnau, E., Ayala, S., and Sturm, T. (2014). Cognitive externalism meets bounded rationality. *Philos. Psychol.* 27, 50–64. doi: 10.1080/09515089.2013.828588

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi: 10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. *Ann. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Brighton, H., and Todd, P. M. (2009). "Situating rationality: ecologically rational decision making with simple heuristics," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (New York, NY: Cambridge University Press), 322–346. doi: 10.1017/CBO9780511816826.017

Brooks, R. A. (1991). Intelligence without representation. *Artif. Intellig.* 47, 139–159. doi: 10.1016/0004-3702(91)90053-M

Callebaut, W. (2007). Herbert Simon's silent revolution. *Biol. Theor.* 2, 76–86. doi: 10.1162/biot.2007.2.1.76

Chater, N., Felin, T., Funder, D. C., Gigerenzer, G., Koenderink, J. J., Krueger, J. I., et al. (2018). Mind, rationality, and cognition: an interdisciplinary debate. *Psychon. Bullet. Rev.* 25, 793–826. doi: 10.3758/s13423-017-1333-5

Chemero, A. (2011). *Radical Embodied Cognitive Science.* Cambridge, MA: MIT Press.

Clark, A. (1997). "Economic reason: the interplay of individual learning and external structure," in *The Frontiers of the New Institutional Economics*, eds J. Drobak and J. Nye (San Diego, CA: Academic Press), 269–290.

Clark, A. (1998). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (2001). Reasons, robots and the extended mind. *Mind Lang.* 16, 121–145. doi: 10.1111/1468-0017.00162

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195333213.001.0001

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19. doi: 10.1093/analys/58.1.7

Clark, A., and Toribio, J. (1994). Doing without representing? *Synthese* 101, 401–431. doi: 10.1007/BF01063896

Constant, A., Clark, A., and Friston, K. J. (2021). Representation wars: enacting an armistice through active inference. *Front. Psychol.* 11:3798. doi: 10.3389/fpsyg.2020.598733

Crowther-Heyck, H. (2005). *Herbert A. Simon: The Bounds of Reason in Modern America*. Baltimore: Johns Hopkins University Press.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Putnam.

Di Paolo, E. (2009). Extended life. *Topoi* 28, 9–21. doi: 10.1007/s11245-008-9042-3

Dreyfus, H. (1972). *What Computers Can't Do. A Critique of Artificial Reason*. New York, NY: Harper and Row.

Elqayam, S. (2011). "Grounded rationality: a relativist framework for normative rationality," in *The Science of Reason: A Festschrift in Honour of Jonathan St.B.T. Evans*, eds K. Manktelow, D. Over, and S. Elqayam (Hove: Psychology Press), 397–420.

Felin, T., Koenderink, J., and Krueger, J. I. (2017). Rationality, perception, and the all-seeing eye. *Psychon. Bullet. Rev.* 24, 1040–1059. doi: 10.3758/s13423-016-1198-z

Fiori, S. (2011). Forms of bounded rationality: the reception and redefinition of Herbert A. Simon's perspective. *Rev. Polit. Econ.* 23, 587–612. doi: 10.1080/09538259.2011.611624

Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Gallagher, S. (2009). "Philosophical antecedents of situated cognition," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins

and M. Aydede (New York, NY: Cambridge University Press), 35–51. doi: 10.1017/CBO9780511816826.003

Gallagher, S. (2011). "Interpretations of embodied cognition," in *The Implications of Embodiment: Cognition and Communication*, eds W. Tschacher and C. Bergomi (Exeter: Imprint Academic), 59–71.

Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind*. New York, NY: Oxford University Press. doi: 10.1093/oso/9780198794325.001.0001

Gallagher, S. (2018). "Embodied rationality," in *The Mystery of Rationality. Mind, Beliefs and Social Science*, eds G. Bronner and F. Di Iorio (Berlin: Springer), 83–94. doi: 10.1007/978-3-319-94028-1_7

Gallagher, S., Mastrogiorgio, A., and Petracca, E. (2019). Economic reasoning and interaction in socially extended market institutions. *Front. Psychol.* 10:1856. doi: 10.3389/fpsyg.2019.01856

Gallagher, S., and Varela, F. J. (2003). Redrawing the map and resetting the time: phenomenology and the cognitive sciences. *Can. J. Philos.* 33, 93–132. doi: 10.1080/00455091.2003.10717596

Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenol. Cogn. Sci.* 4, 23–48. doi: 10.1007/s11097-005-4737-z

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi: 10.1016/S1364-6613(98)01262-5

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Gallese, V., Mastrogiorgio, A., Petracca, E., and Viale, R. (2020). "Embodied bounded rationality," in *Palgrave Handbook of Bounded Rationality*, ed R. Viale (London; New York, NY: Palgrave Macmillan), 377–390. doi: 10.4324/9781315658353-26

Gallese, V., and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends Cogn. Sci.* 15, 512–519. doi: 10.1016/j.tics.2011.09.003

Gibbs, R. W. Jr. (2005). *Embodiment and Cognitive Science*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511805844

Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious*. London: Penguin.

Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.* 1, 107–143 doi: 10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Ann. Rev. Psychol.* 62, 451–482. doi: 10.1146/annurev-psych-120709-145346

Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–669. doi: 10.1037/0033-295X.103.4.650

Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychol. Rev.* 98, 506–528. doi: 10.1037/0033-295X.98.4.506

Gigerenzer, G., Todd, P. M., and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. New York, NY: Oxford University Press.

Glenberg, A. M. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–19. doi: 10.1017/S0140525X97000010

Goldman, A., and de Vignemont, F. (2009). Is social cognition embodied? *Trends Cogn. Sci.* 13, 154–159. doi: 10.1016/j.tics.2009.01.007

Goldman, A. I. (2012). A moderate approach to embodied cognitive science. *Rev. Philos. Psychol.* 3, 71–88. doi: 10.1007/s13164-012-0089-0

Harnad, S. (1990). The symbol grounding problem. *Phys. D Nonlin. Phenomena* 42, 335–346. doi: 10.1016/0167-2789(90)90087-6

Hatchuel, A. (2001). Towards Design Theory and expandable rationality: the unfinished program of Herbert Simon. *J. Manag. Govern.* 5, 260–273. doi: 10.1023/A:1014044305704

Haugeland, J. (1978). The nature and plausibility of cognitivism. *Behav. Brain Sci.* 1, 215–226. doi: 10.1017/S0140525X00074148

Hertwig, R., Hoffrage, U., and Martignon, L. (1999). "Quick estimation: letting the environment do some of the work," in *Simple Heuristics That Make Us Smart*, eds G. Gigerenzer, P. M. Todd, and the ABC Research Group (New York, NY: Oxford University Press), 209–234.

Hutchins, E. (1999). "Cognitive artifacts," in *The MIT Encyclopedia of the Cognitive Sciences*, eds R. A. Wilson and F. C. Keil (Cambridge, MA: MIT Press), 126–128.

Jacob, P. (2016). "Assessing radical embodiment," in *Foundations of Embodied Cognition: Perceptual and Emotional Embodiment*, eds M. H. Fischer and Y. Coello (London: Taylor and Francis), 38–58.

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475. doi: 10.1257/000282803322655392

Kahneman, D. (2011). *Thinking Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., and Tversky, A. (1982). "The simulation heuristic," in *Judgment Under Uncertainty*, eds D. Kahneman, P. Slovic, and A. Tversky (New York, NY: Cambridge University Press), 201–208. doi: 10.1017/CBO9780511809477.015

Katsikopoulos, K. V., and Lan, C. H. D. (2011). Herbert Simon's spell on judgment and decision making. *Judgm. Decis. Mak.* 6, 722–732.

Khatin-Zadeh, O., Eskandari, Z., Cervera-Torres, S., Ruiz Fernández, S., Farzi, R., and Marmolejo-Ramos, F. (2021). The strong versions of embodied cognition: three challenges faced. *Psychol. Neurosci.* 14, 16–33. doi: 10.1037/pne0000252

Kiverstein, J., and Clark, A. (2009). Introduction: mind embodied, embedded, enacted: one church or many?. *Topoi* 28, 1–7. doi: 10.1007/s11245-008-9041-4

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York, NY: Basic books.

Lakoff, G., and Núñez, R. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics Into Being*. New York, NY: Basic Books.

Lee, S. W., and Schwarz, N. (2014). "Metaphors in judgment and decision making," in *The Power of Metaphor: Examining its Influence on Social Life*, eds M. J. Landau, M. D. Robinson, and B. P. Meier (Washington, DC: APA), 85–108. doi: 10.1037/14278-005

Mastrogiorgio, A., and Petracca, E. (2014). Numerals as triggers of System 1 and System 2 in the 'bat and ball' problem. *Mind Soc.* 13, 135–148. doi: 10.1007/s11299-014-0138-8

Mastrogiorgio, A., and Petracca, E. (2015). Razionalità incarnata. *Sistemi Intelligenti*. 27, 481–504. doi: 10.1422/82223

Mastrogiorgio, A., and Petracca, E. (2016). "Embodying rationality," in *Model-Based Reasoning in Science and Technology*, eds L. Magnani and C. Casadio (Cham: Springer), 219–237. doi: 10.1007/978-3-319-38983-7_12

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. London: Palgrave Macmillan. doi: 10.1057/9780230592889

Merleau-Ponty, M. (2002). *Phenomenology of Perception*. London: Routledge. doi: 10.4324/9780203994610

Meteyard, L., Cuadrado, S. R., Bahrami, B., and Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48, 788–804. doi: 10.1016/j.cortex.2010.11.002

Miłkowski, M. (2013). *Explaining the Computational Mind*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9339.001.0001

Neemeh, Z. A. (2021). Smooth coping: an embodied, Heideggerian approach to dual-process theory. *Adapt. Behav.* 2021:10597123211017337. doi: 10.1177/10597123211017337

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newell, A., and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. ACM*. 19, 113–126. doi: 10.1145/360018.360022

Petracca, E. (2017). A cognition paradigm clash: Simon, situated cognition and the interpretation of bounded rationality. *J. Econ. Methodol.* 24, 20–40. doi: 10.1080/1350178X.2017.1279742

Petracca, E. (2020). Two and a half systems: the sensory-motor system in dual-process judgment and decision-making. *J. Neurosci. Psychol. Econ.* 13, 1–18. doi: 10.1037/npe0000113

Petracca, E., and Gallagher, S. (2020). Economic cognitive institutions. *J. Institut. Econ.* 16, 747–765. doi: 10.1017/S1744137420000144

Pitt, D. (2020). "Mental representation," in *The Stanford Encyclopedia of Philosophy*, ed E. N. Zalta. Available online at: https://plato.stanford.edu/archives/spr2020/entries/mental-representation/ (accessed May 1, 2021).

Proffitt, D. R., and Linkenauger, S. A. (2013). "Perception viewed as a phenotypic expression," in *Action Science: Foundations of an Emerging Discipline*, eds W. Prinz, M. Beisert, and A. Herwig (Cambridge, MA: MIT Press), 171–197. doi: 10.7551/mitpress/9780262018555.003.0007

Raab, M. (2017). Motor heuristics and embodied choices: how to choose and act. *Curr. Opin. Psychol.* 16, 34–37. doi: 10.1016/j.copsyc.2017.02.029

Risko, E. F., and Gilbert, S. J. (2016). Cognitive offloading. *Trends Cogn. Sci.* 20, 676–688. doi: 10.1016/j.tics.2016.07.002

Rolla, G. (2019). Reconceiving rationality: situating rationality into radically enactive cognition. *Synthese* 1–20. doi: 10.1007/s11229-019-02362-y

Shapiro, L. A. (2004). *The Mind Incarnate*. Cambridge, MA: MIT Press.

Shapiro, L. A. (2019). *Embodied Cognition, 2nd Edn*. London: Routledge. doi: 10.4324/9781315180380

Simon, H. A. (1947). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization, 1st Edn*. New York, NY: Macmillan.

Simon, H. A. (1955). A behavioral model of rational choice. *Quart. J. Econ*. 69, 99–118. doi: 10.2307/1884852

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychol. Rev*. 63, 129–138. doi: 10.1037/h0042769

Simon, H. A. (1996a). *The Sciences of the Artificial, 3rd Edn*. Cambridge, MA: MIT Press.

Simon, H. A. (1996b). *Models of My Life*. Cambridge, MA: MIT Press.

Spellman, B. A., and Schnall, S. (2009). Embodied rationality. *Queen's Law J*. 35, 117–164. doi: 10.2139/ssrn.1404020

Sterelny, K. (2004). "Externalism, epistemic artefacts and the extended mind," in *The Externalist Challenge*, ed R. Schantz (Berlin: De Gruyter), 239–254. doi: 10.1515/9783110915273.239

Tirado, C., Khatin-Zadeh, O., Gastelum, M., Leigh-Jones, N., and Marmolejo-Ramos, F. (2018). The strength of weak embodiment. *Int. J. Psychol. Res*. 11, 77–85. doi: 10.21500/20112084.3420

Vera, A. H., and Simon, H. A. (1993). Situated action: a symbolic interpretation. *Cogn. Sci*. 17, 7–48. doi: 10.1207/s15516709cog1701_2

Viale, R. (2019). La razionalità limitata 'embodied' alla base del cervello sociale ed economico. *Sistemi Intelligenti* 31, 193–203. doi: 10.1422/92942

Wilson, R. A., and Clark, A. (2009). "How to situate cognition: letting nature take its course," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (New York, NY: Cambridge University Press), 55–77. doi: 10.1017/CBO9780511816826.004

Winograd, T., and Flores, F. (1986). *Understanding Computers and Cognition: A New Foundation for Design*. Norwood, NJ: Ablex.

Wong, M. (2006). The human body's built-in range finder: the thumb method of indirect distance measurement. *Math. Teach*. 99, 622–626. doi: 10.5951/MT.99.9.0622

Ziemke, T. (2003). "What's that thing called embodiment?," in *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, eds R. Alterman and E. Kirsh (Mahwah, NJ: Lawrence Erlbaum), 1134–1139.

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.

The handling editor declared a past co-authorship with the author of the manuscript.

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.

Copyright © 2021 Petracca. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 30 September 2021
doi: 10.3389/fpsyg.2021.711289

Embodied Heuristics

*Gerd Gigerenzer**

Max Planck Institute for Human Development, Berlin, Germany

Intelligence evolved to cope with situations of uncertainty generated by nature, predators, and the behavior of conspecifics. To this end, humans and other animals acquired special abilities, including heuristics that allow for swift action in face of scarce information. In this article, I introduce the concept of *embodied heuristics,* that is, innate or learned rules of thumb that exploit evolved sensory and motor abilities in order to facilitate superior decisions. I provide a case study of the gaze heuristic, which solves coordination problems from intercepting prey to catching a fly ball. Various species have adapted this heuristic to their specific sensorimotor abilities, such as vision, echolocation, running, and flying. Humans have enlisted it for solving tasks beyond its original purpose, a process akin to *exaptation*. The gaze heuristic also made its way into rocket technology. I propose a systematic study of embodied heuristics as a research framework for situated cognition and embodied bounded rationality.

Keywords: embodied heuristics, gaze heuristic, interception problems, sensorimotor abilities, bounded rationality, adaptive toolbox

OPEN ACCESS

Edited by:
Riccardo Viale,
University of Milano-Bicocca, Italy

Reviewed by:
Elisabetta Versace,
Queen Mary University of London, United Kingdom
Colin Allen,
University of Pittsburgh, United States

***Correspondence:**
Gerd Gigerenzer
gigerenzer@mpib-berlin.mpg.de

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *18 May 2021*
Accepted: *31 August 2021*
Published: *30 September 2021*

Citation:
Gigerenzer G (2021) Embodied Heuristics. Front. Psychol. 12:711289. doi: 10.3389/fpsyg.2021.711289

BRIEF SUMMARY

Bounded rationality is the study of how humans and other animals rely on heuristics to achieve their goals in situations of uncertainty. It differs from axiomatic rationality, which asks whether humans conform to logical principles such as transitivity. This paper contributes to the emerging field of embodied bounded rationality, which studies how the body supports rational behavior. Specifically, I propose the concept of embodied heuristics, along with a program on how to study these. An embodied heuristic requires specific sensory and motor abilities to be executed. I provide a case study of the gaze heuristic, which solves visuomotor coordination problems when capturing or avoiding a moving target, from intercepting prey to catching a Frisbee. I show how various species adapted the heuristic to their specific sensory and motor abilities, allowing it to solve interception problems in both two dimensions (on the ground) and three dimensions (in the air or water), and for vision and echolocation. Humans rely on the heuristic for catching fly balls and other tasks beyond its original domain, a process akin to exaptation. The heuristic has been built into rocket technology. This article is of programmatic nature, outlining a novel research program for situated cognition and embodied bounded rationality.

INTRODUCTION

Jean Piaget once said that he cannot think without a pen in hand. For him, writing *was* thinking, not the translation of thought onto paper (Gruber and Vonèche, 1977). Accordingly,

his theory of cognitive development begins with the child's sensory and motor processes, which are eventually transformed into mental life, where they become cognitive operations and structures. The general idea that cognition is closely intertwined with action was later called *embodied cognition*. This term, however, has been used for a highly diverse set of ideas, including the role of gestures, narratives, and physical proximity in behavior. An early version was *ecological psychology*, most prominently Gibson's (1979) view that perception requires movement to detect the invariants in ambient light: "So we must perceive in order to move, but we must also move in order to perceive" (p. 223). In the field of robotics, Brooks (1991) embraced a Gibsonian-inspired architecture, where robots need no symbolic representation of their world; their sensors are connected directly to their behaviors, enabling them to "use the world as its own model" (p. 139). What unites these various approaches, which have been called the four "E's" – *embodied*, *embedded*, *extended*, and *enactive cognition* – is their critique of theories that explain behavior on the basis of internal processes only (e.g., theory of mind or computational theories of cognition) without considering the role of the body and the environment (Wilson, 2002; Shapiro and Spaulding, 2021).

In the present article, I begin from a different perspective, the evolution of rational behavior.

One might think that rational choice theory – choice axioms and subjective expected utility maximization – has long investigated how humans and other animals make decisions. Yet most theories of rational behavior assume that humans have mental capacities for foreseeing the future that real humans can only dream of: perfect foresight of all future events, along with their consequences and probabilities (Hammerstein, 2012). These assumptions are made not because they are realistic but because they are needed to apply the convenient mathematical tools of optimization. Economist Milton Friedman (1953) famously defended these "as-if" models by arguing that their purpose is prediction, not psychological realism. Their strength lies in the beauty of abstract models, where humans are pictured as econometricians. Their downside is that everything psychological plays little if any role, except as a source of irrationality. This methodological choice has left us with an unsatisfying situation. It has promoted a flood of theories that neither describe actual behavior nor intend to do so. Furthermore, contrary to Friedman's vision, expected utility models appear barely able to predict behavior. According to a review, "their power to predict out-of-sample is in the poor-to-nonexistent range" (D. Friedman et al., 2014). Logical axioms hence may not have been the best route to understanding rational behavior in the real world.

In this article, I start with an evolutionary view on decision-making. I introduce the concept of *embodied heuristics*, that is, rules of thumb that exploit specific sensory and motor capacities in order to facilitate high-quality decisions in an uncertain world. Instead of taking an axiomatic approach, models of heuristics take an *algorithmic* approach to represent the sequential process of decision-making in time. Following that, I present a case study of the gaze heuristic that illustrates how an embodied heuristic exploits sensory and motor abilities and how the heuristic has been adapted to the specific abilities of different species. Moreover, by a process akin to exaptation, the heuristic ended up solving new tasks created by human culture. I begin with what might have been the first decisions made by living organisms.

THE DAWN OF DECISION-MAKING

The earth is about 4.5 billion years old. Life emerged some 3.8 billion years ago and animals much later, about 1 billion years ago. It began in the form of single-celled organisms equipped with early versions of sensors and a small repertoire of actions. The best-studied single-celled organism is a bacterium called *E. coli* (named after its discoverer, the pediatrician Theodor Escherich). Its popularity is based on the observation that it does not appear to die but instead splits into two daughter bacteria, which again split, and so on (Khamsi, 2005). It can be found in the lower intestine of humans and other warm-blooded organisms. *E. coli* can perform two motions, run or tumble, that is, move in a straight line or randomly change course. It continuously switches between these actions, although tumbling is reduced when its sensors detect increasing concentrations of food (see Godfrey-Smith, 2016, for a philosopher's account of this behavior). Here we observe the earliest form of decision-making: bacteria choosing between two actions, run or tumble, guided by chemical cues in their environment. These actions serve adaptive goals, finding food and avoiding toxins. The bacteria rely on decreasing or increasing rates of various chemicals as cues. In decision theory, a cue is a sign, or clue, of something that is not directly accessible, such as food or toxins.

Bacteria are *prokaryotes*, cells without a nucleus. Much later, *eukaryotes* arose from a merger of bacterial cells and eventually formed plants, mushrooms, and animals. Eukaryotes also formed "eyespots," which mark the beginning of vision and allow for further cues to guide action. One of these, light, has a dual function. For some organisms such as single-celled organisms and plants, it is mainly a source of energy, supplying solar power. Although humans and other animals also sunbathe, for them light is primarily a source of information. Humans *infer* the outside world from patterns of light.

Inference is crucial, as we cannot directly see the world. Our inferences, albeit more elaborate than those of single cells, remain intelligent "bets" based on uncertain cues. The great physiologist Hermann von Helmholtz spoke of "unconscious inferences" because even humans are not aware of how they make these inferences, such as reconstructing a three-dimensional world from a two-dimensional retinal image. Unconscious inferences border on magic, given that an infinite number of states of the world are consistent with this retinal image. Through millions of years of learning, sensory and motor abilities have evolved in tandem with heuristics that help make good inferences in such situations of uncertainty – to find food and mates, to avoid toxins and predators, and to solve the basic goals of organisms.

Along with individual inferences, social behavior evolved. Consider *E. coli* again. It reacts not only to signs of edible food and dangerous toxins, but also to chemicals that signal

the presence of other bacteria. This reaction opened the door to the evolution of *coordination* between organisms, that is, social behavior. An example is *quorum sensing* among bacteria living inside of squids. Bacteria produce light through a chemical reaction, but only if enough other bacteria are around to join in. They appear to follow a simple heuristic: The more of the signaling chemical one senses, the more light one produces (Godfrey-Smith, 2016, p. 19). The light produced serves its host, the squid, as camouflage. Without this light, predators from below would see the shadow of squids, which are nocturnal animals, as cast by the moonlight. In humans, social coordination takes many forms, including communication, cooperation, and competition, and has led to cultural systems such as churches, political parties, and the market.

Let us now consider a concrete example of how inferences are made based on an embodied heuristic.

EMBODIED HEURISTICS: AN ILLUSTRATION

Ants, like humans, make real-estate choices, that is, decisions about where to live, which are essential to their fitness. Consider *Leptothorax albipennis*, a small ant approximately 3 mm long that lives in colonies with up to 500 workers and a single queen. When their old nest is destroyed, the ant colony sends out scouts to locate a new site that is sufficiently large to house the entire colony. The ants prefer nest sites consisting of narrow cracks in rocks with flat areas. How can a scout ant estimate the irregular area of a candidate site? A series of ingenious experiments revealed that scout ants use a smart rule called "Buffon's needle algorithm," named after the French eighteenth-century mathematician Buffon, who discovered it millennia after the ants did (Mallon and Franks, 2000).

To determine the size of the area, the scout ant first moves for a fixed period (less than two minutes) on an irregular path that covers the area fairly evenly. While doing so, it leaves behind a trail of pheromones. After that the ant exits the area, and then returns and repeats the procedure of walking around randomly. In this second round, the ant counts how often it crosses its own pheromone trail and uses the count to estimate the area of the site: the larger the number of crossings, the smaller the area. This heuristic is amazingly accurate: For a site that is half the size of the area needed, the frequency of crossing is 1.96 times greater (Mugford et al., 2001).

In Buffon's needle problem, the question is asked, what is the probability p that a needle dropped on a floor made of parallel and equally wide strips of wood will end up lying across a line between two strips? For a needle of length l, $p = 2l/\pi t$, where t is the width of the strips. Buffon used the solution to calculate the number π. In the ant's heuristic, the lines are the ant's pheromone trail and the needles lying across lines are the ant's crossings of its own trail. The ant is not interested in π, but in the length t between lines, which indicates the area.

The ant's heuristic involves its body in several ways. First, the ant needs to move around. The heuristic would not work if the ant simply sat still and looked around. Second, the ant's body produces a pheromone trail, and its sensory system has the ability to recognize its own trail. These biological functions are necessary for the heuristic to be executed, but not sufficient. In addition, the ant needs cognitive abilities such as counting crossings and retaining a memory of the count. Many insects can in fact measure and memorize the rate at which they encounter stimuli (Stephens and Krebs, 1986). All in all, ants have evolved an embodied heuristic to infer the area of potential nest sites.

AXIOMATIC RATIONALITY, BOUNDED RATIONALITY, AND ECOLOGICAL RATIONALITY

The scout ant solves an adaptive problem, finding a nest site. The bacteria *E. coli* solves its own adaptive problems, finding food and avoiding toxins. Adaptive problems relate to survival and reproduction, such as finding a safe location, food, and a sexual partner, or cooperating and competing in social groups (Tooby and Cosmides, 1992). A common characteristic of adaptive problems is the presence of uncertainty, that is, when full knowledge of all options together with their consequences and probabilities is not attainable. Theories of rational behavior, in contrast, have mostly studied artificial lotteries and well-defined games where all is known for certain, including the probabilities. These are known as situations of risk (Knight, 1921).

Axiomatic Rationality

The best-known theory of decision-making goes by many names: axiomatic rationality, expected utility maximization, or rational choice theory. Given the many definitions of rational choice theory, *axiomatic rationality* is a more precise designation. In the axiomatic approach, the term *rationality* has little to do with solving adaptive problems. Instead, it refers to a set of choice axioms, such as completeness and transitivity, and to expected utility maximization. It is also not meant to describe the process of how ants choose a new nest site or how humans make decisions. All it might offer is a model in which ants are assumed to have complete knowledge about the features of all sites in reach and choose a site that maximizes their utility. Such a model would neither help an ant to know what to do nor aid a behavioral biologist in understanding what ants do nor guide an AI engineer in building a robot ant. The theory is deliberately abstract and "as-if."

In fact, its originators, von Neumann and Morgenstern (1944), never intended axiomatic rationality to describe what humans and other animals do or what they should do. Instead, these authors derived the necessary and sufficient conditions to represent choices on a number line, called *utility function*. These conditions are the choice axioms and are similar to the properties of real numbers (Cantor, 1954). Von Neumann and Morgenstern's great contribution was to prove that if an individual satisfies the set of axioms, then their choices can be represented by a utility function – *nothing more*. Nowhere in the three

editions of their landmark book did the founders speak of axioms as a description of how people behave or should behave.

Ten years later, the normative interpretation of choice axioms was promoted by Savage (1954), known as the father of Bayesian decision theory. Yet Savage explicitly limited the theory to *small worlds (S, C),* that is, situations in which the exhaustive and mutually exclusive set of future states *S* and their consequences *C* are known. That is why choices between lotteries have become a standard task in decision research, from behavioral economics to cognitive neuroscience. Here, all future states (the tickets), their outcomes (the prizes), and their probabilities are known (as mentioned before, these are also called *situations of risk*.) However, Savage maintained that it would be "utterly ridiculous" (p. 16) to apply utility theory beyond small worlds, that is, to well-defined situations that are intractable, such as chess, or to ill-defined situations where one cannot know all possible future states and their consequences. Savage's example of an ill-defined situation was planning a picnic (p. 16), which is prone to unexpected events.

In a surprising turn in history, quite a few psychologists and economists disregarded Savage's restrictions and began to assert that axiomatic rationality applied to *all* situations (Binmore, 2008). At the same time, it was known since the demonstrations of Maurice Allais and Daniel Ellsberg that people systematically violate the theory *even in small worlds*. Note that both Allais and Ellsberg criticized the rationality of rational choice theory, not the rationality of people. Nevertheless, many psychologists (mis)construed the theory to be normative and routinely blamed deviations on people, not the theory (e.g., Thaler and Sunstein, 2008; Kahneman, 2011). These deviations were attributed to *bounded rationality*, implying that humans are innately susceptible to cognitive illusions or even irrationality.

Bounded Rationality and Ecological Rationality

But that was not what Herbert Simon, who coined the term *bounded rationality*, meant. In fact, Simon (1989, p. 377) argued in favor of studying how humans and other animals actually make decisions when the conditions for axiomatic rationality are not met, that is, under uncertainty. His revolutionary proposal required leaving the safe haven of small worlds, or situations of risk, and sailing out to study the actual process of decision-making under uncertainty. That proved too much for most neo-classical economists, who reinterpreted bounded rationality as optimization under constraints, which is also not what Simon meant. The psychologists who studied deviations from axiomatic rationality attached yet another meaning to bounded rationality, as deviations between judgment and rational choice theory that signify irrationality. While these latter two definitions contradict each other, one signifying rationality, the other irrationality, what they share is their embrace of rational choice theory as the unconditional benchmark for all behavior (Gigerenzer, 2020). This double takeover has been so successful that few people have noticed how *bounded rationality* has been decoupled from Simon's revolutionary program.

As in axiomatic decision theory, the study of the mind's evolved psychology, not to speak of the body, appears irrelevant for human decision-making in the present definitions of bounded rationality. To avoid confusion, my colleagues and I instead refer to *ecological rationality* in our work on extending Simon's original program (Gigerenzer et al., 1999). **Figure 1** shows the general framework. The left side represents mind and body; the right side represents the environment in which a decision needs to be made. These two sides specify the two blades of Simon's "scissors," an analogy he used to explain why one needs to investigate the interplay between cognition and environment to understand behavior: Looking at only one blade of a pair of scissors does not explain how it cuts so well.

The study of ecological rationality analyzes the match between the adaptive toolbox of an individual or species, and the environment. A *match* refers to the likelihood that a given heuristic achieves a given goal in a given environment. Heuristics exploit sensory capacities and motor abilities and are in this sense embodied heuristics. Together, they constitute the adaptive toolbox, which specifies the first blade of Simon's scissors.

The second blade is the environment. It contains the goals of the organism, such as a good nest site. Note that the *environment* here refers to the world as experienced by animals or humans, as in von Uexküll's (1957) *Umwelt*, not to an exhaustive description in terms of molecular biology or geophysics.

The study of ecological rationality addresses three questions (Gigerenzer and Gaissmaier, 2011; Todd et al., 2012). The first concerns the repertoire of tools: What are the heuristics in the adaptive toolbox of an individual, institution, or species? The second concerns the organism's environment: What are the relevant environment structures? The third concerns the match between mind and environment: What are the environmental conditions conducive to the success of particular heuristics with respect to a goal? Together, the answers to these three questions enable us to comprehend why heuristics evolved and the conditions under which a given heuristic is likely to succeed.

What the study of ecological rationality does *not* ask is whether a behavior departs from logical systems of rationality. Strictly following logical inference can, in fact, even hinder solving adaptive problems. Consider two parties engaged in a social contract of the type "if you take the benefit, then you have to pay the costs" (Cosmides, 1989). Although the heuristic "check whether your partner took the benefit but did not pay the costs" can lead to choices that contradict an interpretation of the social contract as a logical conditional "if p then q," it enables detecting cheaters (Gigerenzer and Hug, 1992). Similarly, a review of deviations from choice axioms and other logical rules – often interpreted as cognitive illusions – found little to no evidence that these deviations are actually associated with lesser health, wealth, happiness, or any other measurable costs (Arkes et al., 2016).

Unlike the ant's implementation of Buffon's needle algorithm, many models of heuristics do not make reference to specific sensory or motor abilities. An example is the investment heuristic 1/N, which solves the problem of how to invest a sum of money into *N* assets by allocating it equally. In the uncertain world of stocks, this fast-and-frugal heuristic has

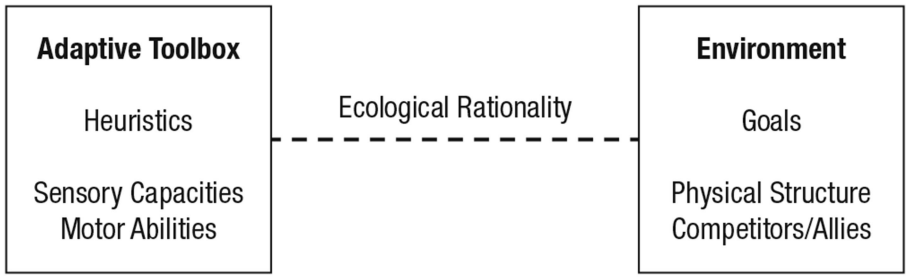


FIGURE 1 | Rationality as the match between heuristics and environment. Left side: The adaptive toolbox of an individual or species, with heuristics that are embodied in sensory capacities and motor abilities. Right side: The environment, including the goals of individual or species and their physical and social structure. The ecological rationality of a heuristic is measured by the degree to which it can attain a goal.

been shown to be able to outperform the Nobel Prize-winning mean-variance portfolio (DeMiguel et al., 2009). However, 1/N does not specify or require specific sensorimotor abilities; dividing a sum by the number of assets can be performed by a pocket calculator as well. Similarly, heuristics such as *minimax* (determine the worst outcome of each option and choose the option with the least undesirable outcome) and *tallying* (count the positive reasons for each option and choose the option with the highest number) do not specify or require any abilities apart from calculation (Gigerenzer and Gaissmaier, 2011).

I will reserve the term *embodied heuristic* for rules that require specific sensory and/or motor abilities to be executed, not for rules that merely simplify calculations. In the next section, I describe in more detail an embodied heuristic that humans share with animal species.

THE GAZE HEURISTIC

When faced with a ball high up in the air, experienced baseball outfielders know where to run in order to catch it. How do they solve the task? There are two visions for finding an answer. The first is to treat the question as an optimal control problem and assume close-to-omniscient players who can make complex calculations unconsciously. That is how Richard Dawkins (1989), p. 95) thinks a player catches a ball:

> He behaves as if he had solved a set of differential equations in predicting the trajectory of the ball. He may neither know nor care what a differential equation is, but this does not affect his skill with the ball. At some subconscious level, something functionally equivalent to the mathematical calculations is going on.

To determine the trajectory of the ball, consciously or unconsciously, the player has to estimate the parameters in this formula:

$$z(x) = x\left(\tan\alpha_0 + \frac{mg}{\beta v_0 \cos\alpha_0}\right) + \frac{m^2 g}{\beta^2}\ln\left(1 - \frac{\beta}{m}\frac{x}{v_0\cos\alpha_0}\right) \quad (1)$$

where $z(x)$ is the height of the ball at flight distance x, measured from the position where the ball was thrown. At $z(x) = 0$, the ball hits the ground. To calculate $z(x)$, the player has to estimate both the initial angle α_0 of the ball's direction relative to the ground and the initial speed v_0 of the ball; know the ball's mass m, the friction β, and that the acceleration of earth g is 9.81 m/s^2 (meter/s squared); and be able to calculate tangent and cosine. Even then, the formula is overly simplified in that it considers only two dimensions and ignores wind and spin. Importantly, the true challenge is not computing the equation, but *estimating* its parameters, such as the initial angle and the initial speed.

Note that Dawkins put the term "as if" into his explanation of how players solve the goal. He was well aware that players do not calculate trajectories; they only behave *as if* they did. What players actually do at the subconscious level remains a mystery in his account. Yet that mystery has been resolved by experimental studies. Experienced players catch a fly ball by using a heuristic that has absolutely nothing to do with calculating a trajectory (**Figure 2**).

Gaze heuristic: *Fixate your eyes on the ball, run, and adjust your speed so that the angle of gaze remains constant.*

The gaze heuristic ignores all the information necessary for computing a trajectory and attends to one variable only, the angle of gaze. In this way, it avoids any measurement errors when estimating the parameters in Equation 1. It consists of three "building blocks" – fixating, running, and adjusting – and works in situations where the ball is already high in the air. If that is not the case, the player needs to adapt the third building block:

Fixate your eyes on the ball, run, and adjust your speed so that image of the ball rises at a constant rate.

One can easily see the logic. If the image of the ball rises at an accelerating rate, the ball will hit the ground behind the player's present position, meaning that the player needs to run backward. If it rises at a decreasing rate, the ball will hit the ground before the player, who then needs to run faster. If the image of the ball rises at a constant rate, the player is running at the correct speed (McBeath et al., 1995; Shaffer and McBeath, 2002).

The gaze heuristic is an embodied heuristic. It requires the ability to hold one's gaze on an object, to run, and to adjust one's running speed. These abilities are learned early in development. For instance, babies begin to exercise visual tracking of moving objects around 2 months of age, such as tracking the objects in mobiles (Jonsson and von Hofsten, 2003).

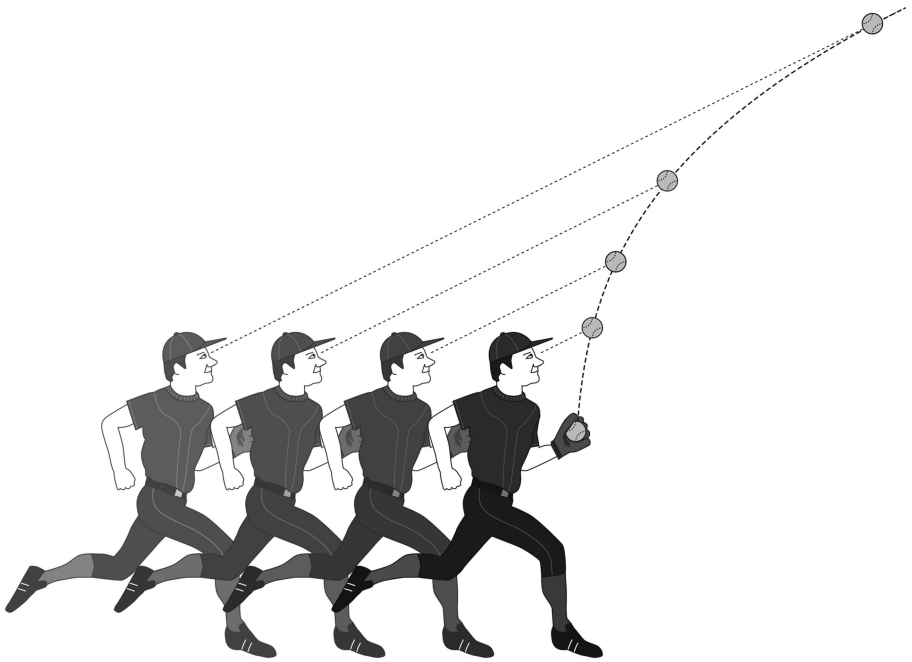


FIGURE 2 | Gaze heuristic. The player adjusts the running speed so that the angle of gaze remains constant. The angle of gaze is the angle between the line from eye to ball and the ground. Shown is the player's position relative to the ball for four points in time.

TABLE 1 | The trajectory calculation model and the gaze heuristic make different predictions about both behavior and cognitive processes. In addition, they imply different specifications of the player's goal. The checkmarks show the predictions supported by experimental studies.

	Trajectory Calculation	**Gaze Heuristic**
Player's goal	Compute landing point	Intercept ball
Prediction 1: Speed	*Runs full speed to landing point.*	*The angle of gaze controls running speed and its change.*✓
Prediction 2: Interception	*At the landing point, player waits to catch ball.*	*Intercepts ball while running.*✓
Prediction 3: Course	*Runs in a straight line.*	*Runs in a slight arc.*✓
Prediction 4: Landing point	*Knows where the ball is landing.*	*Does not know landing point.*✓

The body is part of the solution. In contrast, state-of-the-art bipedal robots cannot implement the gaze heuristic because they lack the ability to run and to securely hold their gaze on a moving object against a noisy background.

Predicting Behavior: As-if Models vs. Embodied Heuristics

Let me now make two more general points. First, reliance on as-if models rather than process models can mislead researchers regarding the actual goal of an organism. The trajectory calculation model suggests that the player's goal is to determine the point where the ball hits the ground (or is at a height in reach of the player) and then run to this point (**Table 1**). The gaze heuristic, in contrast, implies that the goal is to intercept the ball. No knowledge about the landing point is necessary; the heuristic leads the player to the ball. A heuristic is not a just an efficient means toward a given end. It can specify what exactly the player wants to achieve. Means can determine ends, not just the other way round.

Now consider the argument by Milton Friedman that models need not be concerned with psychological realism, only with good predictions. The gaze heuristic and the study of embodied heuristics in general, however, show that psychological realism can lead to better predictions than as-if models. Because as-if models do not care about cognition, only about behavior, let us have a closer look at four predictions about behavior (**Table 1**).

Consider first the running speed. The trajectory model suggests that players would perform better the faster they run to the expected landing point, so that they have time for last-second adjustments. In contrast, the gaze heuristic makes a very specific prediction: that players' speed is controlled by the angle of gaze, which determines speed and its change. If players run too fast, they will miss the ball.

Second, consider interception. According to the trajectory model, players should ideally arrive at the landing point before the ball and wait for it. The gaze heuristic, in contrast, implies that players catch the ball while running. The reason is that they adjust their running speed until they catch the ball. In both cases, the predictions following from the gaze heuristic have been supported by experimental studies (e.g., McBeath et al., 1995; Shaffer and McBeath, 2002).

Third, consider the course of running. According to the trajectory model, the player will run straight toward the landing

point. In contrast, the gaze heuristic can imply in certain situations that players run a slight arc to keep the angle of gaze constant. These arcs have been demonstrated in experiments with skilled outfielders (Shaffer and McBeath, 2002).

Finally, if players consciously or unconsciously computed the landing point, as assumed by the trajectory model, they would know where the ball will land. No such knowledge is implied by the gaze heuristic. Studies show that even experienced players (just like ordinary people) have difficulties estimating the trajectory of the ball, its apex, and the landing point yet are nevertheless able to catch the ball (Shaffer and McBeath, 2005).

The general point is that the as-if trajectory model is ignorant about the process and objectives of decision-making and thus makes incorrect predictions about the resulting behavior. It treats the problem as one of calculating landing points, while the heuristic treats it as one of coordination between body and ball.

Coordination Problems

The gaze heuristic and its relatives can resolve various coordination problems. These include interception, such as when athletes catch balls, but also avoidance of collisions, as in sailing and flying. When beginners learn to sail, they are taught a version of the gaze heuristic to infer whether another boat is on a collision course: Fixate your gaze on the other boat; if the angle of gaze remains constant, change your course quickly. When beginners learn to fly a light aircraft, they may be taught a further version of the same rule: If another plane approaches and you fear collision, look at a scratch in your windshield and observe whether the other plane moves relative to that scratch. If not, dive away immediately – otherwise, the plane might end up colliding with this scratch.

The "miracle on the Hudson River" is a famous case where reliance on the gaze heuristic saved lives. On January 15, 2009, US Airways Flight 1549 collided with a flock of Canada geese shortly after take-off, which shut down both engines. The pilots had to make a life-and-death decision: to try to reach the next airport or attempt a risky landing in the Hudson. Landing at the next airport would have been the safer option, but only if the plane could actually make it that far. As co-pilot Jeffrey Skiles explained, to determine whether the sailing plane could safely make it to the airport, they did not try to calculate the trajectory of the plane but instead relied on a version of the gaze heuristic (Rose, 2009):

> It's not so much a mathematical calculation as visual, in that when you are flying in an airplane, a point that you can't reach will actually rise in your windshield. A point that you are going to overfly will descend in your windshield.

The point in the windshield rose, which meant the plane would have crashed before reaching the airport. The heuristic helped to make the right decision; all passengers and crew survived (Gigerenzer, 2014, pp. 27–29).

Note that the heuristic can be used both consciously and unconsciously, as illustrated by the pilots and the outfielders, respectively. Most outfielders rely on the gaze heuristic without being able to explain how they catch a ball. Their behavior is intuitive, not consciously deliberative (Gigerenzer, 2007). In general, heuristics may be learned consciously, by instruction, or unconsciously, by trial and error learning or imitation. The process is the same, a fact overlooked by dual-process theories that align heuristics with unconsciousness and, moreover, assume different processes (see Kruglanski and Gigerenzer, 2011).

Exaptation

The gaze heuristic was not invented by baseball outfielders. Bats, birds, fish, and other animals rely on it for intercepting prey and mates (e.g., Collett and Land, 1975). The observation that different species rely on the same heuristic invites two possible explanations, *homology* and *analogy*. Homology means that common structures between different species – here, common heuristics, – are due to a common evolutionary ancestor. Analogy means that there is a functional similarity based on something other than common ancestors. Whatever the correct explanation is, we can safely assume that the gaze heuristic evolved for predatory-prey interaction and not for baseball or cricket.

Sperber (1994) distinguished the proper domain of a cognitive module from its actual domain, that is, the domain for which a module actually evolved from a domain to which it was extended or transferred. Similarly, the term *exaptation* means that a trait or feature acquires a new function beyond its original one derived by evolution. It was introduced by Gould and Vrba (1982) as an alternative to the concept of *preadaptation* in order to emphasize that the original function was not connected to the new function. A classical example is the argument that feathers were not evolved for flight in birds, but originally had the function of temperature regulation in their ancestors, reptiles. Eventually, feathers became enlisted for a new function, sailing and, eventually, flying. I have not yet seen a discussion of exaptation with respect to heuristics, embodied or not. Here, I use the term *exaptation* in a more general sense, beyond its original biological meaning, namely, for cultural exaptation where humans find new functions for evolved heuristics. The gaze heuristic is a candidate in point. Its proper domain, or original function, is described in the next section.

PREDATOR-PREY COORDINATION

How does a hawk intercept a duck? **Figure 3** (top) shows two strategies for interception. The first is *direct pursuit*, where the hawk flies straight at the duck, that is, takes the shortest path. When the duck changes its position, the hawk changes its direction accordingly, so that the distance between it and the duck is always the shortest possible. The top left panel shows a case of direct pursuit that ends in a failed interception with a characteristic wavering tail chase (Hamlin, 2017). The second strategy is a version of the gaze heuristic, where the hawk does not fly in a straight line toward the duck. Rather, it initially flies toward an expected point X where it would intercept the duck if the latter did not change course (top right panel). The

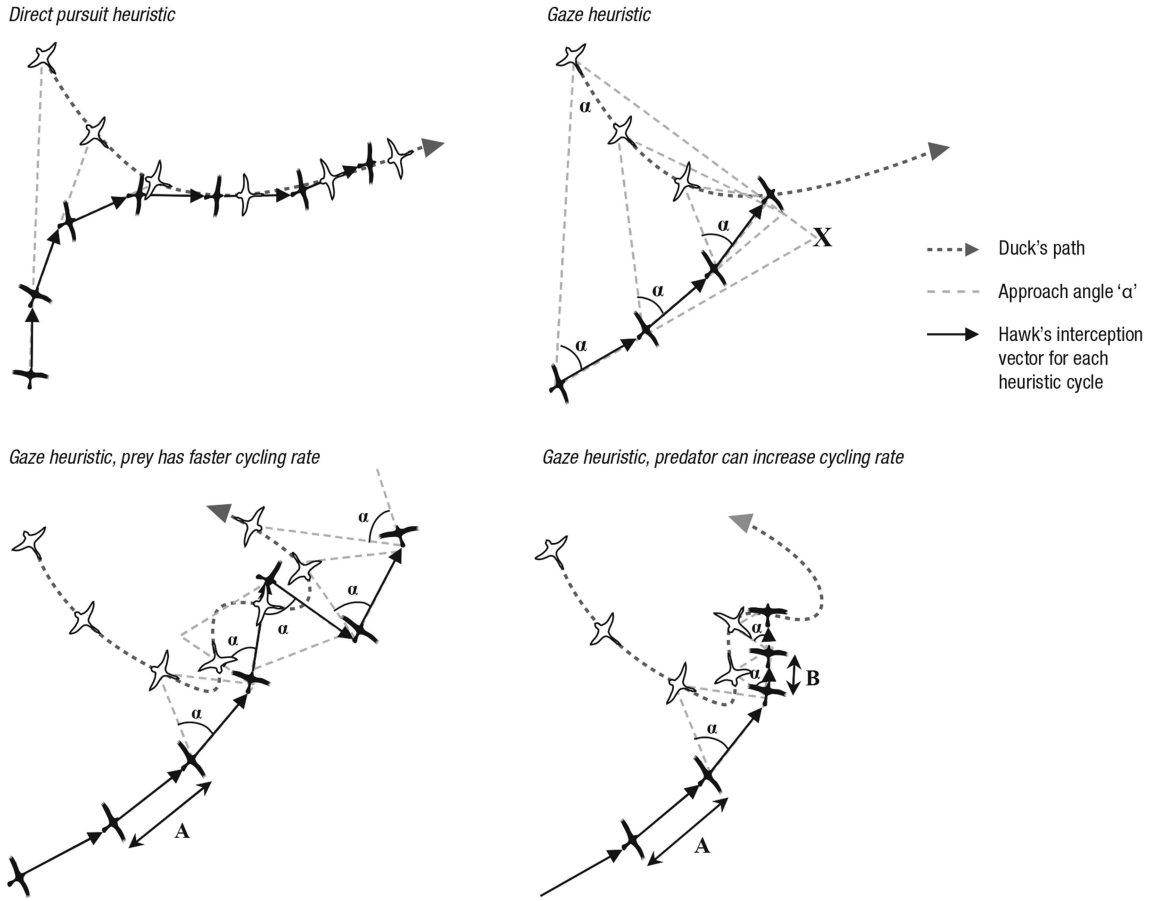


FIGURE 3 | Predators (dark hawks) pursuing prey (white ducks). From top left to bottom right: *Direct pursuit heuristic:* A predator flies in the direction that is the shortest path to the prey, and adjusts the direction when the prey changes its course. If predator and prey fly at the same speed, the result is a characteristic wavering pursuit pattern. *Gaze heuristic:* A predator determines the angle α between the direct line to the prey and the initial estimate *X* of the intersection point, and then adjusts its direction to the subsequent flight path of the duck so that α remains constant. Even when both predator and prey fly at the same speed, the predator can intercept the prey. *Predator relies on gaze heuristic, but prey has faster cycling rate*: The number of adjustments an animal can make per second is its cycling rate, represented by the length *A* of its path before it can change its direction. Here, the prey has a faster cycling rate than the predator, which enables it to evade the predator. *Predator relies on gaze heuristic and can increase cycling rate*: Here, the predator has the ability to increase the cycling rate from *A* to *B*, which is higher than that of the prey, resulting in fast interception. Adapted from Hamlin (2017).

angle α between the duck, the hawk, and the interception point *X* defines the angle of gaze. When the duck changes course, the hawk also changes its course so that the angle of gaze remains constant. In geometric terms, the angle of gaze is the base angle of a triangle with equal sides and apex *X*.

Which of the two heuristics do hawks employ? Studies with headcams mounted on hawks showed that they rely on the gaze heuristic (Kane et al., 2015). The comparison between direct pursuit and the gaze heuristic in **Figure 3** indicates why: Relying on the latter allows for faster interception and avoids the wavering tail chase. Moreover, because the hawk does not fly directly toward the duck, its attack is less obvious. Only when the target is stationary do hawks rely on direct pursuit, that is, fly directly toward the prey.

To be successful in pursuit, an organism needs the ability to adjust speed and direction quickly when the angle changes (due to wind in the case of the fly ball, or due to evasive movements in the case of the duck). The number of possible adjustments per second is the *visual cycle rate*. Raptors have a visual cycling rate of about 200 per second, whereas humans have a much lower rate of about 10 per second (Hamlin, 2017). The cycling rate corresponds to the length of the path *A* before it can be adjusted to maintain a constant angle of gaze. The smaller *A* is, the faster the hawk's cycling rate. **Figure 3** (bottom left panel) shows a prey with a faster cycling rate than the hawk that avoids interception by changing its course before the hawk is able to do so. Thanks to a faster cycling rate, the prey can even get behind the predator. Although the hawk keeps the optical angle constant, it is too slow to adjust. Finally, the bottom right panel shows a successful predator that increases its cycling rate in the final stage of the pursuit from *A* to *B*.

From Gaze to Echolocation and Whiskers

Although the gaze heuristic is named after the visual sense, it has been adapted to other senses, too. Bats rely on the

equivalent of the gaze heuristic when hunting moths in darkness, but their interception is based on sound, not vision. They use an echolocation system that emits sound as a series of short "clicks" or "calls" (Denny, 2004). When a target is located, the clicks occur more frequently as the bat closes in on a prey. The echolocation version of the gaze heuristic works as described in **Figure 3**, except that the angle α is based on echolocation rather than visual location. In response to bats, moths have evolved bat-detecting ears capable of hearing the clicks (Hofstede and Ratcliffe, 2016). Outside the bat's detection range, a moth's first reaction is to fly away from the bat. If the frequency of clicks increases, meaning that the bat has detected its prey, this triggers spasms in the moth's wings, resulting in unpredictable flight. Finally, if the clicks peak in a buzz of about 200 clicks a second, the moth's reflex is to instantly freeze to fall out of the bat's path. All this happens within seconds. The bat's clicks correspond to the visual cycles of humans and hawks.

At the final stage of pursuit, the gaze heuristic is supported by tactile senses. Mammals such as cats, rats, and seals use their whiskers to locate the prey. Whiskers are an array of long, coarse hairs around the head and mouth that provide information about the prey's position in the final milliseconds before impact (Grant et al., 2009). Experiments showed that rats were less successful in completing an interception of a mouse when their whiskers were removed, and if they did succeed, the final clean bite to the neck took longer and was messier (Hamlin, 2017).

THE ROYAL AIR FORCE DISCOVERS THE GAZE HEURISTIC

According to a historical analysis, the Royal Air Force (RAF), after some trial and error, was the first to have discovered the gaze heuristic around the beginning of World War II (Hamlin, 2017). The problem was that the British controllers who used radar to direct fighters to enemy planes had failed to reach the required 90% interception rate. Special calculating devices and increasingly complex mathematics were introduced to crunch the numbers, but to no avail. In this situation, an impatient RAF commander demonstrated that he could do a better job by eye, meeting the 90% rate. His system was fleshed out by the Chairman of the "Committee for the Scientific Survey of Air Defence", Sir Henry Tizard, into a fixed angle approach and taught to the controllers. This system became known as the "Tizzy Angle" and used for the remainder of the war.

After being trained to use the gaze heuristic, the British controllers no longer sent pilots directly *via* the shortest distance toward the opponent (the direct pursuit heuristic) but instead estimated an intersection point *X*, which determined the constant angle. If the bomber changed course after having recognized the fighter, the fighter was directed to change course too, but keep the angle constant. Shortly before interception, the faster fighter could turn around and meet the bomber frontally, where it was most vulnerable (**Figure 4**).

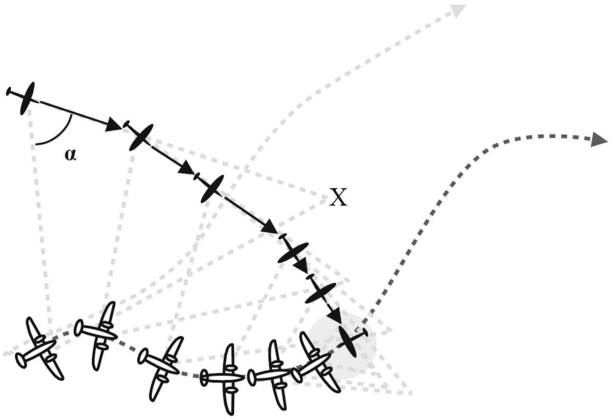


FIGURE 4 | British controllers relied on the gaze heuristic to direct fighter planes to intercept German bombers. From Hamlin (2017).

According to historical records and training materials, the controllers of the German Luftwaffe relied instead on a direct pursuit strategy and appear to have never discovered the gaze heuristic during World War II. In the pursuit control technique, the controller instructs the pilot (who cannot yet see the enemy plane) to fly directly toward the opponent. If the opponent changes course, the pilot is directed to also change course and take the shortest path toward the opponent. The pursuit strategy vectors the fighter behind its opponent, just as the hawk trails behind the duck in **Figure 3** (top left panel), and leads to a smaller rate of interception. Although the Germans' radar system was superior to that of the RAF in several respects, the British use of the gaze heuristic was devastating to the Luftwaffe and decisive to the Battle of Britain. Hamlin (2017) argues that the Germans might have won this battle if they had linked their high-tech radar system with a gaze-based heuristic control system. By the end of the war, the Germans were leading in missile technology, including anti-aircraft missiles based on the direct pursuit strategy, but had missed a smart heuristic.

After World War II, the United States army combined German missile technology with the British gaze heuristic system into a most successful autonomous guided weapon: the Sidewinder A1M9 short-range air-to-air missile (Hamlin, 2017). The missile is a simple, robust interception system whose "gaze" is directed at a point source of heat, which is the target. Once the missile is on its way, it makes continuous inquiries (with a rapid cycle rate) about the changes of the target's position and adjusts its direction so that the angle of "gaze" remains constant. The Sidewinder is still in use in many nations, and new developments appear to be based on the same heuristic maintaining a constant angle of approach.

A RESEARCH PROGRAM ON EMBODIED HEURISTICS

The case study on the gaze heuristic can provide a template for a general research program on embodied heuristics.

Specifically, that program addresses three core questions (see **Figure 1**):

The Repertoire of Heuristics in the Adaptive Toolbox. What are the embodied heuristics used by individuals or groups to solve problems? What sensory and motor abilities do these heuristics exploit to find efficient solutions?

The Structure of the Environment. What is the structure of an environment to which a given heuristics is adapted?

Ecological rationality: Which heuristics are likely to achieve a given goal in a given environment?

This program contrasts with a majority of theories in the cognitive sciences in two respects:

1. The body (e.g., sensory and motor abilities) and the environment "select" the heuristics and are crucial to explaining behavior. This differs from "internalist" theories that explain behavior solely by computational processes inside the mind, such as expected utility maximization, Bayesian probability updating, logical symbol manipulation, System 1/System 2 theories, and theory of mind.
2. As a consequence, behavior can often be explained by simple heuristics rather than by complex computations. An embodied heuristic can exploit innate or learned capabilities and thereby be both simple *and* accurate.

Only the first of these two points is common to all views on embodied cognition. Although the term *embodied heuristics* is used occasionally in the literature on embodied cognition (e.g., Gallagher and Hutto, 2008, p. 27), no programs in existence develop models of embodied heuristics that can be explicated in the form of algorithms and then simulated and tested (see **Table 1**).

WHY STUDY EMBODIED HEURISTICS?

In this article, I introduced the concept of embodied heuristics and provided a case study on a particularly interesting example, the gaze heuristic. This amazing feat of evolution, a dynamic adaptive heuristic, enables animals and humans to make rapid decisions with the help of a highly automatized system superior to conscious reasoning. I end with some general insights this case study provides.

Embodied Heuristics Are Efficient Because They Exploit Sensory and Motor Abilities

To execute an embodied heuristic requires specific sensorimotor abilities. For instance, the gaze heuristic is of little value to a robot that cannot keep its eye on a moving object against a noisy background or cannot run. In the vocabulary of AI, the software needs the proper hardware. This basic insight contrasts with most theories in decision-making that rely exclusively on logic or probability.

Complex Problems Do Not Generally Need Complex Solutions

From machine learning to cognitive sciences, a common assumption is that the more complex a model is, the better it must perform. That is true in situations of risk or well-defined games such as chess and Go, but not in situations of uncertainty, as in interactions with humans and other animals (Katsikopoulos et al., 2020). For instance, between 2007 and 2015, Google Flu Trends tried to predict the proportion of flu-related doctor visits, based on an analysis of 50 million search terms using thousands of big data models. When predictions failed, Google engineers made the algorithm more complex instead of simpler, without any improvement. In contrast, a simple heuristic that relies on a single data point, the most recent number of flu-related doctor visits, predicts better than Google's big data models (Katsikopoulos et al., in press). Similarly, in social encounters, heuristics based on imitation or tit-for-tat can hardly be beaten, even in well-defined games (Duersch et al., 2012). The general methodological lesson is to always test complex models against simple heuristics.

One and the Same Heuristic Can Solve Problems in Stationary and in Nonstationary Worlds

Some scholars have hypothesized that the success of simple heuristics is restricted to stable or nonsocial worlds, and that social interactions need complex strategies (for a discussion, see Hertwig and Hoffrage, 2013). The gaze heuristic is a clear counterexample, as are tit-for-tat and heuristics relying on imitation. Moreover, in the present case, nonstationary problems of predator-prey interaction are the proper domain of the gaze heuristic; those involving inanimate objects such as fly balls are later extensions. In general, complex models with many free parameters are likely to succeed in stable, stationary worlds, while simple heuristics, like human intelligence, evolved for dealing with an uncertain, social world (Katsikopoulos et al., 2020).

Cognition is More Than Symbol Manipulation

Research on embodied heuristics follows and extends Simon (1956) program of bounded rationality. At the same time, it contrasts with Newell and Simon's (1972) *physical symbol hypothesis*, which assumes that symbol manipulation, as in computers, is the essence of all rational systems, implying that sensorimotor abilities are of little relevance (see Gallese et al., 2020). Cognitive and social psychologists have largely taken their inspiration from the symbol manipulation view, assuming that cognition is mainly what statistics and computer do (Gigerenzer, 1991; Gigerenzer and Goldstein, 1996). In these theories, which are often highly complex and "as-if," neither heuristics nor their anchoring in the body play a role.

The gaze heuristic is a simple iterative heuristic that adapts to changes in flight path due to wind in case of a fly ball or due to evasion attempts in the case of prey. It can solve problems in stationary and nonstationary environments and is embodied in the sense that it requires specific sensory and motor capabilities to function efficiently. The astonishing feat is that the heuristic has enlisted different sensory capacities in different species, including vision and echolocation. It also

has enlisted various motor abilities. When dogs catch a Frisbee, they implement the gaze heuristic by running (Shaffer et al., 2004); when teleost fish pursue prey, they implement the heuristic by swimming; and when hawks go after prey, they implement it by flying. Humans implement the heuristic both in two-dimensional space, such as when trying to avoid a collision with another sailboat, or in three-dimensional space, as when trying to avoid a collision in the air.

The heuristic has also inspired rethinking financial regulation. Andrew Haldane, the Bank of England's chief economist, presented his acclaimed Jackson Hole talk entitled "The Dog and the Frisbee" on the gaze heuristic as a model for a safer world of banking. He argued for introducing simple and robust control systems in place of complex regulatory systems, which neither foresaw nor prevented the crisis of 2008 (Haldane and Madouros, 2012). Haldane used the heuristic as an analogy for robustness, not embodiment. For instance, capital requirements are estimated by calculating the value-at-risk of a bank, which may involve estimating thousands of risk factors and millions of covariation coefficients. The limited success of these estimations recalls the calculations made by the RAF before it discovered the gaze heuristic (Gigerenzer and Gray, 2017). The banking system is a fast-changing, nonstationary environment where simple rules can lead to better and more transparent decisions. The standard approach in cognitive science, however, has resembled bank regulation, based on the assumption that more complexity is always better. Journals are filled with highly parameterized models that integrate all possibly relevant information, Bayesian or otherwise. Complexity pays for well-defined situations such as games, but leads to overfitting in ill-defined situations of uncertainty.

Evolution has given us the gaze heuristic, and with it a pointer to study the ingenious solutions it has found for a brain the size of two fists. To do so, we need to embark on a systematic study of embodied heuristics in the real world.

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

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Arkes, H. R., Gigerenzer, G., and Hertwig, R. (2016). How bad is incoherence? *Decision* 3, 20–39. doi: 10.1037/dec0000043

Binmore, K. (2008). *Rational Decisions*. Princeton, NJ: Princeton University Press.

Brooks, R. A. (1991). Intelligence without representation. *Artif. Intell.* 47, 139–159. doi: 10.1016/0004-3702(91)90053-M

Cantor, G. (1954). *The Founding of the Theory of Transfinite Numbers*. transl. P. E. B. Jourdain (New York: Dover). First published 1915.

Collett, T. S., and Land, M. F. (1975). Visual control of flight behavior in the hoverfly, Syritta pipiens *L. J. Comp. Physiol.* 99, 1–66. doi: 10.1007/BF01464710

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31, 187–276. doi: 10.1016/0010-0277(89)90023-1

Dawkins, R. (1989). *The Selfish Gene*, 2nd Edn. Oxford: Oxford University Press.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *Rev. Financ. Stud.* 22, 1915–1953. doi: 10.1093/rfs/hhm075

Denny, M. (2004). The physics of bat echolocation: signal processing techniques. *Am. J. Phys.* 72, 1465–1477. doi: 10.1119/1.1778393

Duersch, T., Oechssler, J., and Schipper, B. C. (2012). Unbeatable imitation. *Game. Econ. Behav.* 76, 88–96. doi: 10.1016/j.geb.2012.05.002

Friedman, D., Isaac, R. M., James, D., and Sunder, S. (2014). *Risky Curves. On the Empirical Failure of Expected Utility*. New York: Routledge.

Friedman, M. (1953). *Essays in Positive Economics*. Chicago: University of Chicago Press.

Gallagher, S., and Hutto, D. D. (2008). "Understanding others through primary interaction and narrative practice," in *The Shared Mind: Perspectives on Intersubjectivity*. eds. J. Zlatev, T. Racine, C. Sinha and E. Itkonen (Amsterdam: John Benjamins), 17–38.

Gallese, V., Mastrogiorgio, A., Petracca, E., and Viale, R. (2020). "Embodied bounded rationality," in *Routledge Handbook of Bounded Rationality*. ed. R. Viale (London: Routledge), 377–390.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gigerenzer, G. (1991). From tools to theories: a heuristic of discovery in cognitive psychology. *Psychol. Rev.* 98, 254–267. doi: 10.1037/0033-295X.98.2.254

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking Press.

Gigerenzer, G. (2014). *Risk Savvy: How to Make Good Decisions*. New York: Viking.

Gigerenzer, G. (2020). "What is bounded rationality?" in *Routledge Handbook of Bounded Rationality*. ed. R. Viale (London: Routledge), 55–69.

Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Ann. Rev. Psychol.* 62, 451–482. doi: 10.1146/annurev-psych-120709-145346

Gigerenzer, G., and Goldstein, D. G. (1996). Mind as computer: birth of a metaphor. *Creativity Res. J.* 9, 131–144. doi: 10.1207/s15326934crj0902&3_3

Gigerenzer, G., and Gray, W. D. (2017). A simple heuristic successfully used by humans, animals, and machines: the story of the RAF and Luftwaffe, hawks and ducks, dogs and Frisbees, baseball outfielders and sidewinder missiles – oh my! *Top. Cogn. Sci.* 9, 260–263. doi: 10.1111/tops.12269

Gigerenzer, G., and Hug, K. (1992). Domain-specific reasoning: social contracts, cheating, and perspective change. *Cognition 42*, 127–171. doi: 10.1016/0010-0277(92)90060-U

Gigerenzer, G., Todd, P. M., and the ABC Research Group (1999). *Simple Heuristics That Make us Smart*. New York: Oxford University Press.

Godfrey-Smith, P. (2016). *Other Minds: The Octopus and the Evolution of Intelligent Life*. London: Harper Collins.

Gould, S. J., and Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology* 8, 4–15. doi: 10.1017/S0094837300004310

Grant, R. A., Mitchinson, B., Fox, C., and Prescott, T. J. (2009). Active touch sensing in the rat: anticipatory and regulatory control of whisker movements during surface exploration. *J. Neurophysiol.* 101(2), 862–874. doi: 10.1152/jn.90783.2008

Gruber, H. E., and Vonèche, J. J. (1977). *The Essential Piaget*. New York: Basic Books.

Haldane, A. G., and Madouros, V. (2012). The dog and the frisbee. *Revista de Economia Institucional*, 109–110.

Hamlin, R. P. (2017). "The gaze heuristic:" biography of an adaptively rational decision process. *Top. Cogn. Sci.* 9, 264–288. doi: 10.1111/tops.12253

Hammerstein, P. (2012). "Towards a Darwinian theory of decision making: games and the biological roots of behavior," in *Evolution and Rationality*. eds. S. Okasha and K. Binmore (Cambridge: Cambridge University Press), 7–22.

Hertwig, R., and Hoffrage, U. (2013). "Simple heuristics: the foundations of adaptive social behavior," in *Simple Heuristics in a Social World*. eds. R. Hertwig, U. Hoffrage and the ABC Research Group (New York: Oxford University Press), 3–36.

Hofstede, H., and Ratcliffe, J. M. (2016). Evolutionary escalation: the bat–moth arms race. *J. Exp. Biol. 219*(11), 1589–1602. doi: 10.1242/jeb.086686

Jonsson, B., and von Hofsten, C. (2003). Infants' ability to track and reach for temporarily occluded objects. *Developmental Sci.* 6, 86–99. doi: 10.1111/1467-7687.00258

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Allen Lane.

Kane, S. A., Fulton, A. H., and Rosenthal, L. J. (2015). When hawks attack: animal-borne video studies of goshawk pursuit and prey-evasion strategies. *J. Exp. Biol.* 218, 212–222. doi: 10.1242/jeb.108597

Katsikopoulos, K., Simsek, O., Buckmann, M., and Gigerenzer, G. (2020). *Classification in the Wild*. Cambridge, MA: MIT Press.

Katsikopoulos, K., Simsek, O., Buckmann, M., and Gigerenzer, G. (in press). Transparent modeling of influenca incidence: big data or a single data point from psychological theory? *Int. J. Forecasting.* doi: 10.1016/j.ijforecast.2020.12.006

Khamsi, R. (2005). Bacteria show signs of ageing. *Nature*. doi: 10.1038/news050131-6

Knight, F. (1921). *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin.

Kruglanski, A., and Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychol. Rev.* 118, 97–109. doi: 10.1037/a0020762

Mallon, E. B., and Franks, N. R. (2000). Ants estimate area using Buffon's needle. *P. Roy. Soc. B–Biol. Sci.* 267, 765–770. doi: 10.1098/rspb.2000.1069

McBeath, M. K., Shaffer, D. M., and Kaiser, M. K. (1995). How baseball outfielders determine where to run to catch fly balls. *Science* 268, 569–573. doi: 10.1126/science.7725104

Mugford, S. T., Mallon, E. B., and Franks, N. R. (2001). The accuracy of Buffon's needle: a rule of thumb used by ants to estimate area. *Behav. Ecol.* 12, 655–658. doi: 10.1093/beheco/12.6.655

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Rose, C. (2009). *The Charlie Rose Show [Television Broadcast]*. New York: PBS.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley and Sons.

Shaffer, D. M., Krauchunas, S. M., Eddy, M., and McBeath, M. K. (2004). How dogs navigate to catch Frisbees. *Psychol. Sci.* 15, 437–441. doi: 10.1111/j.0956-7976.2004.00698.x

Shaffer, D. M., and McBeath, M. K. (2002). Baseball outfielders maintain a linear optical trajectory when tracking uncatchable flyballs. *J. Exp. Psychol: Human.* 28, 335–348. doi: 10.1037/0096-1523.28.2.335

Shaffer, D. M., and McBeath, M. K. (2005). Naïve beliefs in baseball: systematic distortions in perceived time of apex for fly balls. *J. Exp. Psychol: Learn.* 31, 1492–1501. doi: 10.1037/0278-7393.31.6.1492

Shapiro, L., and Spaulding, S., (2021). "Embodied cognition," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Fall 2021 Edition), https://plato.stanford.edu/entries/embodied-cognition/

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychol. Rev.* 63, 129–138. doi: 10.1037/h0042769

Simon, H. A. (1989). "The scientist as problem solver," in *Complex Information Processing. The Impact of Herbert A. Simon*, eds. D. Klahr and K. Kotovsky (Hillsdale, NJ: Erlbaum), 375–398.

Sperber, D. (1994). "The modularity of thought and the epidemiology of representations," in *Mapping the Mind*. eds. L. A. Hirschfeld and S. A. Gelman (Cambridge: Cambridge University Press), 39–67.

Stephens, D. W., and Krebs, J. R. (1986). *Foraging Theory*. Princeton, NJ: Princeton University Press.

Thaler, R. H., and Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Todd, P. M., Gigerenzer, G., and the ABC Research Group (2012). *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press.

Tooby, J., and Cosmides, L. (1992). "The psychological foundations of culture" in *The Adapted Mind*. eds. J. H. Barkow, L. Cosmides and J. Tooby (New York: Oxford University Press), 19–138.

von Neumann, J., and Morgenstern, O., (1944). *Theory of Games and Economic Behavior*. 2nd ed. 1947 3rd 1953. Princeton, NJ: Princeton University Press.

von Uexküll, J. (1957). "A stroll through the worlds of animals and men: a picture book of invisible worlds" in *Instinctive Behavior: the Development of a Modern Concept*. ed. C. H. Schiller (New York: International Universities Press), 5–80.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. B. Rev.* 9, 625–636. doi: 10.3758/BF03196322

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.

Copyright © 2021 Gigerenzer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 26 November 2021
doi: 10.3389/fpsyg.2021.769591

Embodied Irrationality? Knowledge Avoidance, Willful Ignorance, and the Paradox of Autonomy

Selene Arfini and Lorenzo Magnani*

Computational Philosophy Laboratory, Philosophy Section, Department of Humanities, University of Pavia, Pavia, Italy

In the current philosophical and psychological literature, *knowledge avoidance* and *willful ignorance* seem to be almost identical conditions involved in irrational patterns of reasoning. In this paper, we will argue that not only these two phenomena should be distinguished, but that they also fall into different parts of the epistemic rationality-irrationality spectrum. We will adopt an epistemological and embodied perspective to propose a definition for both terms. Then, we will maintain that, while willful ignorance is involved in irrational patterns of reasoning and beliefs, knowledge avoidance should be considered epistemically rational under particular circumstances. We will begin our analysis by considering which of the two phenomena is involved in patterns of reasoning that are still amply recognized as irrational—as wishful thinking, self-deception, and akrasia. We will then discuss the impact of epistemic feelings—which are emotional events that depend on epistemic states—on agents' decision-making. Then, we will consider the impact of willful ignorance and knowledge avoidance on agents' autonomy. By considering these issues, we will argue that when agents are aware that they are avoiding certain information (and aware of what kind of feelings acquiring the information would trigger), knowledge avoidance should be considered a rational, autonomy-increasing, hope-depended selection of information.

Keywords: knowledge avoidance, willful ignorance, embodied cognition, epistemic feelings, self-deception, autonomy, hope, bounded rationality

OPEN ACCESS

Edited by:
Shaun Gallagher,
University of Memphis, United States

Reviewed by:
Iskra Fileva,
University of Colorado Boulder, United States
Marco Viola,
University Institute of Higher Studies in Pavia, Italy

****Correspondence:***
Selene Arfini
selene.arfini@unipv.it

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *02 September 2021*
Accepted: *02 November 2021*
Published: *26 November 2021*

Citation:
Arfini S and Magnani L (2021) Embodied Irrationality? Knowledge Avoidance, Willful Ignorance, and the Paradox of Autonomy. Front. Psychol. 12:769591. doi: 10.3389/fpsyg.2021.769591

INTRODUCTION

Various psychological studies have now confirmed that there are different situations in which the majority of people would not want to know something to avoid pain, regret, or anxiety (Eil and Rao, 2011; Sicherman et al., 2016; Gigerenzer and Garcia-Retamero, 2017). In some cases, people still choose to remain ignorant of something even if they would highly benefit, without apparent material costs, from the act of acquiring that information. For example, many patients who suffer from chronic diseases avoid getting information about their health even if having such knowledge is free and it would permit them to cope better, managing their symptoms and therapy (Oster et al., 2013). Still, a question that current literature strangely avoids is: is this cultivated ignorance epistemically irrational? For example, do these choices imply self-deception or do they affect agents' epistemic autonomy?[1]

Irrationality can be generally defined as a cognitive impediment (Bortolotti, 2010, 2014), and, more specifically, epistemic irrationality defines the creation of those beliefs which "are badly

[1]We will define epistemic autonomy in section.

supported by the evidence available to the agent, or are maintained despite counter-evidence which is available to the agent" (Jefferson et al., 2017, p. 3). Since phenomena of *deliberate not-knowing* (terms that we will use to comprehend both willful ignorance and knowledge avoidance) involve the dismissal or the avoidance of evidence, it is reasonable to believe that there is a strong link between them and epistemic irrationality[2]. Contrary to this idea, in this paper, we argue that while willful ignorance can be rightfully considered as part of epistemically irrational patterns of reasoning, we can judge as epistemically rational the more specific condition of knowledge avoidance.

To advance our arguments, we will adopt an embodied cognition perspective. Thus, in section 1 we will comment on the fact that now discourses of rationality encompass, at various levels, takes from embodied cognition research and from theories of bounded/ecological rationality (Goldstein and Gigerenzer, 2002; Bissoto, 2007; Spellman and Schnall, 2009; Xu et al., 2020). Both these approaches have challenged the idea that irrationality involves only the deviance from rules of logic or probability. Also, emerging theories on ignorance have defied its definition as simply *lack of knowledge* or *true belief*, describing it as a more complex spectrum of states and processes (Arfini, 2019; Werner, 2021). Since now the distinctions between knowledge and ignorance and between rationality and irrationality are more blurred, we will argue that we need to consider *knowledge* avoidance different from willful *ignorance* and that they may fall into different parts of the rationality-irrationality spectrum. We will then propose a definition for both terms, grounded on the literature currently available.

Then, in section 2, we will consider that many irrational phenomena, such as wishful thinking (subsection 2.1), epistemic akrasia (subsection 2.2), and self-deception (subsection 2.3) require deliberate not-knowing. We will discuss which phenomenon between willful ignorance and knowledge avoidance is involved in these irrational processes, and we will argue that they mainly involve wishful ignorance but not knowledge avoidance (subsection 2.4).

In section 3, we will consider possible reasons to judge the phenomenon of knowledge avoidance as epistemically rational. Since the basic tenets of embodied cognition argue that bodily states affect cognitive processes (Chemero, 2011), we will argue that we should consider the emotional impact of certain information (in particular certain epistemic feelings, Arango-Muñoz, 2014a,b) among the costs of acquiring knowledge, contributing to labeling certain situations of knowledge avoidance as forms of rational ignorance. Then, in subsection 3.1, we will discuss the impact of knowledge avoidance on agents' autonomy, which will also bring us to discuss the paradox of autonomy, already introduced in Magnani (2020). By considering these issues and comparing cases of willful ignorance and knowledge avoidance (in subsection 3.2), we will argue that when agents are aware that they are avoiding certain information (and aware of what kind of feelings acquiring the information would trigger), knowledge avoidance should be considered a *rational*, *autonomy-increasing*, *hope-depending* selection of information.

1. EMBODIED RATIONALITY AND THE KNOWLEDGE-IGNORANCE SPECTRUM

"The rational human is neither rational nor human." With these words Spellman and Schnall (2009) begin their essay on how the rationality paradigm has evolved in the last few decades to encompass a more realistic account of the imperfect and limited rational individual.

Bounded rationality theories indeed explained why the majority of people in ordinary situations would not adhere to the rules imposed by logic and probability or would not maximize their utility (Mastrogiorgio and Petracca, 2016). The reason is not that irrationality is a natural human tendency, but that both internal (mental) and external (environmental) constraints limit our possibilities, making us more apt to look out for satisfying (*satisficing*, in Simon's lexicon Simon, 1997) options for decision-making instead of optimal ones. More than a few scholars (Gigerenzer and Goldstein, 1996; Spellman and Schnall, 2009; Xu et al., 2020) have written on the fall of the standard normative paradigm of rationality, and different currents emerged from its ashes (as, for example, theories of "ecological rationality" developed by Todd and Gigerenzer, 2007, 2012). So, yes, the rational human described with the old-fashioned paradigm of rationality could not be classified as *rational* in the same way as the bounded and ecologically rational human—also called *homo heuristicus* (Bardone, 2011)—we are now taking into consideration. But what about the *human* part of it?

Spellman and Schnall (2009) argue, and we agree, that those ideal cognizers who make decisions without considering their context nor their bodily cues have nothing of the human traits that characterize our typical agents. For this reason, bounded rationality today is variously rethought within the broader compass of embodied cognition research (Gallagher, 2018; Xu et al., 2020). Indeed, different principles of embodied cognition have poured into current theories of rationality and orient them into analyzing not only the individual cognizers but the cognitive system that comprehends them (Gallagher, 2018). Nonetheless, some patterns of reasoning, such as epistemic akrasia, self-deception, and wishful thinking, are still clearly epistemically irrational. They usually compromise instead of favoring good decision-making performances, and they do involve forms of deliberate not-knowing. Our question here is: which phenomena of deliberate not-knowing do these irrational conditions involve?

To answer this question, we should first provide reasons to consider willful ignorance and knowledge avoidance two different phenomena. To do that, we will rely on two main arguments: the difference between the current epistemological analysis of "knowledge" and "ignorance" and the specific different

[2]Despite our intention to focus on whether we should consider willful ignorance and knowledge avoidance as part of *epistemically* irrational reasoning, we acknowledge that further considerations may be put out regarding how these conditions can be also part of irrational behaviors (deeming them as *pragmatically* irrational). However, discussing how knowledge avoidance and willful ignorance may be pragmatically rational or irrational is outwith the scope of this paper.

usage of "knowledge avoidance" and "willful ignorance" in philosophical, psychological, and cognitive literature.

The first argument relies on the complexity of the terms "knowledge" and "ignorance" in current epistemology. Knowledge, considered either with the traditional tripartite view that sees it as composed by *true and justified beliefs* (Gettier, 1963) or with more fallibilist accounts (Haack and Kolenda, 1977), is considered a more or less stable but peculiar phenomenon. On the contrary, ignorance has been recently depicted as a more nuanced and diffused condition since its concept encompasses not only the epistemic status of agents but also their attitudes toward it (Haas and Vogt, 2015). For example, we take for granted that ignorance is involved in cases in which agents do not know facts, but also when they do not realize they are not able to do something, or when they do not realize they have committed some errors doing a particular task, or if they have doubts about their competence, or if they do not know that they are competent in certain areas. These cases are, of course, very distinct and differently refer to first-order ignorance (subjects do not know *p*), second-order ignorance (subjects do not know whether they know *p*), or a mix of both, and in specialized literature they come with specific terminology, as *factual ignorance*, *procedural ignorance*, *doubt*, *uncertainty*, *error*, *tacit knowledge*, and so on[3]. Moreover, some cases involve both agents' ignorance and partial knowledge, as know-that or know-how. Still, we resist the attribution of knowledge, even partial knowledge, in these cases, while we have no problem in recognizing how the agent's beliefs system, reasoning, and behavior are affected by ignorance. The reason is that we hold a higher standard for the attribution of knowledge rather than ignorance, and so we tend to distinguish, for example, knowledge from mere belief, while we use a broad meaning for ignorance to generally speak of lack of knowledge, but also lack of awareness, comprehension, or confidence. This lower standard for the attribution of ignorance explains why definitions of ignorance as lack of knowledge (Le Morvan, 2013) or lack of true beliefs (Le Morvan and Peels, 2016) are now broadly challenged. Indeed, they seem to defy the common use of ignorance as a broader term, which refers to a combination of epistemic lack (lack of information, knowledge, competence, etc.) and specific attitudes of self-awareness (doubt, uncertainty, unawareness, etc.).

We also need to point out that in recent times, externalist approaches have also grounded emerging theories on ignorance, defying its definition as something that has to do with only higher cognitive functions of the individual. Different scholars are proposing embodied, extended, and distributed approaches to the idea of ignorance in both epistemological and psychological fields (Arfini, 2021; Arfini and Magnani, 2021; Werner, 2021). Thus, since there is an ampler spectrum of possibilities that defines what we call ignorance rather than what we recognize as knowledge, it is not unreasonable to argue that *knowledge avoidance* should be considered reasonably different from a state of *willful ignorance*.

In the usage of the two concepts, we can even see this difference. Knowledge or information avoidance is generally seen as the choice of not getting *specific information* for *particular reasons* (Sweeny et al., 2010). To make some examples, people may avoid acquiring certain knowledge:

- to postpone anxiety or pain regarding a specific situation (e.g., some patients avoid knowing if they have the genetic markers of a hereditary illness) (Sweeny et al., 2010; Eil and Rao, 2011);
- to preserve positive emotions, as awe and wonder, or even neutral ones, as surprise and suspense (e.g., some people avoid knowing the sex of the unborn child) (Gigerenzer and Garcia-Retamero, 2017).
- to preserve a fair judgment (e.g., the double-blind peer-review process) (Gigerenzer and Garcia-Retamero, 2017).

In all these cases, the agents avoid knowing a particular piece of information that may affect their judgment and reasoning. Instead, scholars often use "willful ignorance" to speak of the more general avoidance of situations that make someone aware of certain information, evidence, or knowledge. So, willful ignorance could prevent agents from knowing about the social impact of their decisions (Grossman and van der Weele, 2017), the law (Zimmerman, 2018), available information (Rubin, 2018), privileged perspectives (May, 2006), and make them disrespect the truth (McIntyre, 2015).

Various articles claim that the idea of "not wanting to know" must be a phenomenon so particular that it does not need any differentiation—which leads them to not distinguishing between willful ignorance and knowledge avoidance (Bertolotti et al., 2016; Gigerenzer and Garcia-Retamero, 2017). The problem with this kind of narrative is that it assumes that ignorance is more or less of one kind. However, intuitively and logically, we consider ignorance to be broader and more differentiated than knowledge. So it is not sufficient to say that ignorance is "what the agent is not aware of," but also, for example, the kind of metacognitive judgments surrounding that ignorance.

Providing a functional definition, we can say that people are willfully ignorant of something when they avoid all circumstances that would allow them to acquire that knowledge, even by accident. Instead, people in a condition of knowledge avoidance do not perform the necessary steps to get a specific piece of information, which could not fall in their laps otherwise. As a last point of characterization, in the case of proper knowledge avoidance, the reasons for not wanting to know have nothing to do with the material costs of acquiring this knowledge, and the agent is also personally interested in acquiring this knowledge[4].

[3] "Tacit knowledge," an epistemically positive term, may seem off place in a list of types of ignorance. On the contrary, the idea of tacit knowledge relies on the partial "unawareness" of the agent who is nonetheless competent. Polanyi's very motto "we can know more than we can tell" (Polanyi, 1966, p. 4) can be rephrase as "we can tell less than we can know," and still it would describe a positive situation for the tacitly knowing agent. Indeed, in this paper, we will not use the word "ignorance" as loaded with a negative connotation, and we will comment on that point on the very beginning of section 2.

[4] Of course, since no scholar presented this distinction before, various authors tried to specify the phenomena they were interested in by coining other formulas—as "deliberate ignorance," used by Gigerenzer and Garcia-Retamero (2017), to speak about what we are calling knowledge avoidance. Here we argue that, since *Ignorance Studies* now propose a more complex view of ignorance, "knowledge avoidance" should be preferred for accuracy in cases where people avoid knowing certain information.

Thus, the distinction between knowledge avoidance and willful ignorance should matter, given the new perspectives on rationality studies. Indeed, since the distinction between rationality and irrationality is now blurred, we need to argue that *knowledge avoidance* and *willful ignorance* may also fall into different parts of the rationality-irrationality spectrum. In the next section, we will then discuss which of the two phenomena is involved in some irrational patterns of reasoning, such as wishful thinking, self-deception, and epistemic akrasia. We will then maintain that most of these forms of irrationality involve wishful ignorance but not knowledge avoidance.

2. DISCUSSION: RATIONAL AND IRRATIONAL IGNORANCE

First, we should point out a rule of thumb that may seem counter-intuitive prima facie but fairly simple to apply after a brief explanation: ignorance is not always epistemically bad for human agents, and knowledge is not always good either. In few words, we should be able to distinguish between a *rational* and *irrational ignorance*. Ignorance is usually presented as the rational choice when the costs of acquiring knowledge outweigh the benefits of possessing it (Mackie, 2012; Somin, 2015; Williams, 2021). In similar ways, also theories of bounded and ecological rationality suggest that we should consider knowledge a limited resource for some good reasons—even pragmatic ones (Jordan, 1996; Reisner, 2009; Star, 2018)[5]. If agents do not have enough time or computational capacity to get the appropriate data to make the most optimal choice, they need to rely on lesser goods.

Moreover, in this part of the analysis, we should consider the difference between the epistemological, logical, and *ideal* definition of knowledge and its phenomenological experience. In few words, what feels like knowledge could be not so: what John Woods (2005) calls "epistemic bubble" defines the easily experienced condition in which we realize that we cannot distinguish, from our first-person perspective, what we know and what we just believe we know. Of course, this condition that feels like knowledge can also involve a form of deliberate not-knowing. However, the mere presence of parts of knowledge or ignorance should imply that we used irrational reasoning to get to that state. As Jefferson et al. (2017, p. 7) point out: "Epistemically irrational beliefs and predictions can be either true or false, but what makes them irrational is that they were not formed on the basis of (sufficiently robust) evidence or are insufficiently responsive to evidence after being adopted."

So, of course, the epistemically problematic trait of irrational reasoning is not that they lead the agents to certain falsity, but that agents delude themselves thinking they have the appropriate epistemic resources to make a decision when it is not the case. Here we will specifically comment on three patterns of reasoning that are considered irrational by most authors in philosophical and psychological literature: self-deception, epistemic akrasia, and wishful thinking. We selected these types of irrational reasoning and not others (as superstition or prejudice) because they all involve types of deliberate not-knowing at their core. So, to discuss the role of deliberate not-knowing in these phenomena, it would not be enough to establish that agents end up being more ignorant than expected in the end, but how and why deliberate not-knowing shapes these kinds of reasoning. To reflect upon these issues, we will briefly present the main definitions of these psychological phenomena, and we will then dedicate a part of the explanation to the comment on the role of ignorance in their maintenance and its motivational character.

2.1. Wishful Thinking

Wishful thinking is commonly described as a positive illusion (Jefferson et al., 2017) which generally moves the agents to believe in statements corresponding to their wishes and to avoid believing ones that are inconsistent with their motivations (Sigall et al., 2000; Mayraz, 2011). This general description does not firmly separate wishful thinking from other kinds of biased reasoning, such as the ones tainted by confirmation bias, and tends to see the motivational character of human reasoning in a theoretical competition with epistemic reasons. Of course, according to the theoretical purposes of different authors, the definition of wishful thinking can become more specific or more general. Some use "wishful thinking" to describe any situation in which "hopes, fears, needs, and other motivational factors combine with, or compete with, prior beliefs as people confront scientific evidence and discourse" (Bastardi et al., 2011) or to refer to how people "avoid information or resist revising their beliefs [. . .] in the competition between cognition and motivation" (Kruglanski et al., 2020).

So it is easy to judge wishful thinking as irrational because, in most cases, the epistemic reasons fall back in the competition with non-epistemic reasons (hope, fear, needs), and the reasoners unjustifiably consider their beliefs epistemically sound. Of course, as Kruglanski et al. (2020) specify, the fact that wishful thinking may reasonably be considered irrational does not mean that it is uncommon in our ordinary decision-making process. On the contrary, we often experience the competition between what we should believe and what we hope/want/need to believe, and we do not always consciously make the epistemically sound "choice" between them.

Thus, in the case of wishful thinking, deliberate not-knowing appears in two ways: a selection of information that discards what goes against the agents' interests and a general unawareness regarding the non-epistemic ground for the doxastic outcomes. So, we claim that willful ignorance, as the "general avoidance of situations that let someone aware of certain information, evidence, or knowledge" could better describe the kind of deliberate not-knowing enacted in these cases. While wishful thinking, people need to avoid certain information and preserve the "wishful" attitude—which consists of a more or less blissful unawareness regarding the effect of non-epistemic reasons on their judgment. Instead, people who avoid particular knowledge

[5]A clarification is needed at this point: in section 3 we will propose one pragmatic reason to consider knowledge avoidance epistemically rational. Even if it is a pragmatic reason to be in a specific epistemic state, we need to say that our argument will only be slightly connected to the debate on the pragmatic reasons for belief. Indeed, knowledge avoidance is not a way to form or maintain a particular belief but a way to avoid forming one. So, the pragmatic reason we will invoke supports the epistemic rationality of "suspending one's belief" instead of forming one.

are well aware of which information they are avoiding and why, so they are, by definition, not wishfully thinking.

2.2. Epistemic Akrasia

The case of epistemic akrasia is complicated since it involves a contrast between first-degree and second-degree orders of beliefs. The general definition says that "epistemic akrasia is possible only if (a) a person's (first-order) beliefs diverge from his higher-order judgments about what it would be reasonable for him to believe and (b) these divergent (first-order) beliefs are freely and deliberately formed" (Owens, 2002, p. 19). In other words, epistemic akrasia describes the situation in which agents hold a belief even though they think it is irrational or unjustified (Greco, 2014; Daoust, 2019; Coates, 2020).

The reasons why they hold this belief is what identifies the akratic pattern of reasoning from other irrational ones: pragmatical akrasia—or weakness of will—is the situation in which people have all qualities, motives, and opportunity to do something that they think would be right for them and fail to do so because they lack conviction, will, and so they give in to the temptation to do easier but less good actions. In a similar way, when people give in and adopt beliefs for epistemic akrasia, they do not want to perform those analytic and epistemically righteous judgments that would allow them to reject some beliefs because of a lack of proof or the presence of counter-proofs to their evidence. They hold on to ignorance as they form false beliefs or insufficiently motivated ones because it is convenient in some respects.

If people choose not to "think hard enough" about what they believe, we can say that they may fall easily into a state of willful ignorance since this condition is broad and general enough to describe deliberate dismissal of adequate reasoning. At the same time, it would be unfair to claim that also knowledge avoidance has a role in this process. People who adopt epistemically akratic reasoning to form beliefs do not exactly know which kind of information, evidence, and knowledge they are dismissing, because they do not put enough effort into knowing that. The akratic reasoning prevents them from precisely selecting which information they are dismissing, so they are not definitely in a condition of knowledge avoidance.

2.3. Self-Deception

Finally, self-deception could represent a challenge to the idea that knowledge avoidance is less involved in irrational beliefs and patterns of reasoning than willful ignorance. The main reason is that we currently do not have only one definition of the phenomenon but multiple descriptions, upon which scholars are still debating. Indeed, Deweese-Boyd (2021) in the Stanford Encyclopedia of Philosophy presents the issue as such: "Virtually every aspect of self-deception, including its definition and paradigmatic cases, is a matter of controversy among philosophers [...] self-deception involves a person who seems to acquire and maintain some false belief in the teeth of evidence to the contrary as a consequence of some motivation, and who may display behavior suggesting some awareness of the truth. Beyond this, philosophers divide [...]" and begins a long list of issues that pertain to this topic.

We need to say, though, that even if it is fascinating to ponder the controversial issues surrounding self-deception, most of its problematic traits do not matter in this particular discussion—as, for example, its morality or practical efficacy. Instead, one controversial but relevant issue at hand is its *intentional* character[6]. Adopting accounts that differ on this particular matter can dramatically change its definition. According to Pedrini (2012) self-deception could have three distinct definitions relative to its intentional character:

1. people hold false beliefs while simultaneously knowing that they are false. They hold dear these false beliefs because it would be too painful to accept that they are false. This is usually called the *intentionalist account* (Davidson, 2004);
2. people delude themselves and believe something false because they have a desire that trumps epistemic reasons to believe otherwise. In this case, they do not know that they hold a false belief, but a thematic desire compromises their rational processes. This is usually presented as the *anti-intentionalist account of self-deception* (Mele, 2000);
3. people shift between believing a certain painful proposition to be true and a condition of self-delusion, in which they believe that proposition is false—*weak intentionalist account* (Pedrini, 2018);

In all three definitions, deliberate not-knowing is involved since agents believe in false statements for different reasons. The first definition is the easiest to dismiss as a case of knowledge avoidance. If self-deceiving people at the same time believe/suspect p (or have enough reasons/evidence to believe/suspect p) and refuse to acknowledge those beliefs and suspicions, they would no longer be in a position to avoid the information/knowledge that they wished they did not acquire. So, this condition would more easily encompass a state of willful ignorance, taken as a comprehensive phenomenon that includes the denial of evidence.

At this point, we need to point out that this definition of self-deception has been heavily criticized by the current philosophical literature, especially by Alfred Mele (2000), who talked about the paradox that surrounds it. Indeed, if we take "believing" as the condition that makes people say that something is true, then it is doubtful to assert that a person can believe at full force that something is both true and false. For this reason, Mele and other scholars have proposed the anti-intentionalist account of self-deception.

Mele and other anti-intentionalists (or non-intentionalists) (Johnston, 1995; Barnes, 2007), indeed, offer this description: subjects fall into self-deceiving patterns of reasoning when their epistemic motivations are compromised by the desire to believe a particular proposition. So, self-deceiving people would believe a particular false proposition p just because their *initial emotional motivation* to believe that p was more

[6]A terminological note here might be useful. In the philosophical debates on self-deception, the word "intentional" is only used to describe in which sense the agent who falls into a state of self-deception does it "deliberately." Here we use the term with this meaning, and we will not refer to the notion of "intentionality" as "aboutness," as it is commonly used in philosophy of mind.

successful than epistemic motivations. Unfortunately, this poses another theoretical problem to the depiction of self-deception: within this theory, self-deception is only the *initial cause* for believing a false proposition, not the explanatory reason for its lasting effect. As Pedrini comments: "if a full-blown belief that *p* is successfully reached, then there is no trace of the psychological tension that seems, instead, to be highly typical of self-deception. For this tension is obviously due to the fact that the motivationally distorted self-deceptive process runs counter to evidence that not-*p* that is at hand, or that is easy available" (Pedrini, 2018, p. 2).

Within this account, we could not attribute the self-deceiving state to knowledge avoidance exactly because self-deceiving people do not recognize certain knowledge as available for emotional reasons (so they do not put any effort into avoiding certain information). In that sense, we are not even discussing a case of deliberate not-knowing since there is no non-epistemic motivation involved in the actual preservation of the state of ignorance.

On the contrary, weak intentionalist accounts of self-deception open the possibility that self-deceiving people would shift from a state of willful ignorance to knowledge avoidance and even self-delusion. Indeed, Pedrini argues that there is a tangible tension between believing and not believing a false proposition; it does not end up being a paradoxical situation, but the agents keep getting back and forth between believing the false proposition and recognizing it is false.

This definition of self-deception incorporates both kinds of deliberate not-knowing because when people are in a self-deluded state, they do not know they are ignorant even if this ignorance comes from their choices (willful ignorance). Instead, when they shift to a more self-knowing state, they still avoid gathering evidence in favor of the true proposition, so they forcefully maintain a condition of knowledge avoidance. Of course, neither willful ignorance nor knowledge avoidance depicts the complex process of self-deception entirely, even in this last and more complex characterization. Self-deception is the shifting between willful ignorance and knowledge avoidance, but neither of these conditions can comprehend the process of self-deception.

2.4. The Rationality of Knowledge Avoidance

As argued so far, commonly defined irrational phenomena mainly involve willful ignorance, not knowledge avoidance. Indeed, returning to the definition we offered of knowledge avoidance, we said that it describes a condition in which agents avoid some knowledge to refrain from anticipated costs (in terms of pain, anxiety, or regret) of possessing it. So knowledge avoidance does not technically involve the willful preservation of false beliefs or the generic dismissal of evidence in favor of certain theoretical positions. It instead refers to situations in which agents have not (nor look for) evidence to fixate a particular belief regarding a specific situation. If these cases do not fit the range of the irrational reasoning we so far described, how should we judge them? In these situations, people avoid acquiring those pieces of information that would impact their emotional state, reasoning abilities, and decisions. Is this another form of epistemic irrationality, or are they adopting patterns of reasoning closer to rational ignorance?

To proceed with our argument, we should point out that, so far, the examination of these cases adopted an old-fashioned cognitivist take on the matter. In many papers regarding this topic, the authors account for material costs of acquiring specific knowledge (money, time, etc.) but not the emotional response of the agents—so, nonmaterial costs. In this paper, we aim at partially closing this gap in the literature, discussing the impact of emotions and, in particular, epistemic feelings—which are feelings that depend on epistemic states (Arango-Muñoz, 2014a,b)—have on the human reasoning.

3. EPISTEMIC FEELINGS, ANTICIPATED REGRET, AND THE APPEAL TO AUTONOMY

In the last two decades, philosophers and cognitive scientists have adopted some distinguishing features to discriminate between types of feelings, separating, for example, between emotional feelings and epistemic ones (Arango-Muñoz, 2014a). In particular, in studies regarding metacognition, feelings have been described as experiences that regard objects or states of affairs that affect the subjects' organism in certain specific ways. While emotional feelings are pretty known and fit without issues in this description, epistemic feelings are less understood and need more explanations to be comprehended in this framework. Epistemic feelings are phenomenal experiences regarding agents' cognitive abilities, conditions, or processes. So, while there is a bodily reaction that accompanies these experiences—to make a practical example, we can think of how it feels to have something on the tip of our tongue (tip-on-the-tongue feeling)—the trigger of these experiences is internal (Arango-Muñoz and Michaelian, 2014). Moreover, since epistemic feelings are reactions to internal contents, epistemic and emotional feelings can create loops between each other and chains of reaction. These reactions and loops, of course, happen without the explicit acknowledged approval of the subjects; instead, they profoundly affect them and their rational evaluations.

We here argue that we should consider the importance of epistemic feelings when reflecting upon knowledge avoidance. Indeed, from a theoretical point of view, it would be reasonable to admit that we can describe the anticipated regret of a decision as an epistemic feeling. The feeling of anticipated regret—which is the leading cause of knowledge avoidance according to Gigerenzer and Garcia-Retamero (2017)—rests upon the idea that we could not cope or we would not be happy with acquiring a particular knowledge (either because it would cause us too much pain because it would spoil our surprise, or it would make us unfair judges). So, at this point, we should discuss whether anticipated regret allows agents to perform types of rational reasoning or not.

In the next section, we will discuss possible conditions that may elicit anticipated regret. We will defend the idea that the

anticipated regret of "no longer being as autonomous as before" may be considered a rational reason to avoid specific knowledge in the new bounded and ecological rationality standards. In particular, we will argue that when knowledge avoidance is conscious, and the agent is aware of the information is giving up (and of the feelings that this knowledge would trigger), we should see knowledge avoidance as evidence of embodied bounded rationality.

3.1. The Hoping Stand: Overcoming the Paradox of Autonomy

As already mentioned, (Gigerenzer and Garcia-Retamero, 2017) propose to take into account "anticipated regret" as one of the negative feelings that may arise when considering not acquiring a particular piece of information. We agree that we should consider anticipated regret as one reason for which people avoid acquiring specific knowledge. However, we think that there is more to add to this consideration: what could more accurately describe how anticipated regret works—considering it as an epistemic feeling—in the mind of people who avoid knowing certain things is the specification of its content. So, what is this anticipated regret about, and why should this potential content matter for the agent?

Here we propose to consider as an answer to this question the agent's *anticipated regret of no longer being as autonomous as before knowing certain information*. To defend this claim, we first need to specify what we can describe as "autonomy," what we will name "epistemic autonomy," and what one of us (Magnani, 2020) has named "the paradox of autonomy".

So, autonomy is not an uncontroversial topic in philosophy, especially in ethical discussions. Buss and Westlung (2018) offer three different accounts of personal autonomy that they claim are dominant and interacting in the current philosophical literature. These accounts are labeled "coherentists" since they variously affirm that 1) agents are autonomous if they are motivated to act, and this motivation is coherent with some of their mental states (Frankfurt, 1971); 2) agents are autonomous when their actions are coherent with a "sufficiently wide range of reasons" for and against that behavior (these reasons could be based on facts about their desires and interests, or even false beliefs) that the agents know and can express (Fischer and Ravizza, 1998); then 3) "the essence of self-government is the capacity to evaluate one's motives on the basis of whatever else one believes and desires, and to adjust these motives in response to one's evaluations" (Buss and Westlung, 2018). We take the last account to describe an "epistemic" type of autonomy, to differentiate it from practical forms of it (those that have to do with "what agents can do" instead of "what agents can believe").

Considering this last definition, the "paradox of autonomy" takes shape (Magnani, 2020). It claims that if, on the one hand, agents need reasoning to be autonomous—so they rely on their decisions, rules, preferences, and desires, on the other hand, the same decisions, rules, preferences, and desires can oppress our thinking and reduce our epistemic autonomy. Moreover, since we know that even our autonomous reasoning may lead to a reduction or an enhancement of our practical and epistemic autonomy, we should judge the rationality of our judgments, decision-making processes, and reasoning on how much the consequences of our decisions will allow us to preserve enough epistemic autonomy to make other rational choices.

With these critical points at hand, we need to reconsider the rationality of knowledge avoidance. Indeed, considering what we have described so far, we can provide some reasons to justify knowledge avoidance rationally. The anticipated cost of acquiring specific knowledge could affect the agent's epistemic autonomy and the agent's autonomy in general.

To explain the first reason adequately, we need to get back to discuss the intersections between cognition and emotions. Indeed, there is quite an emerging literature that describes negative emotions as more impactful on the cognitive capacity of agents than positive ones. Indeed, this realization brings out what Eil and Rao (2011, p. 116) call the "good news, bad news" effect:

> Our primary finding is that subjects incorporated favorable news into their existing beliefs in a fundamentally different manner than unfavorable news. In response to favorable news, subjects tended to respect signal strength and adhered quite closely to the Bayesian benchmark, albeit with an optimistic bias. In contrast, subjects discounted or ignored signal strength in processing unfavorable news leading to noisy posterior beliefs that were nearly uncorrelated with Bayesian inference. [. . .] We call this finding the good news bad news effect. The result suggests that bad news has an inherent "sting" that differential processing mitigates.

So, if agents anticipate that, by seeking out specific knowledge, they could receive news so bad that they would compromise their rational decision-making processes, it would be more reasonable to remain ignorant or postpone the acquisition of that knowledge to preserve solid reasoning-making abilities. Since the reasoning capacity is one of the conditions for maintaining both epistemic and practical autonomy, we can also justify this choice to defend one's autonomy in general.

Moreover, it is essential to consider also the degree of certainty that certain information carries. Let us consider two cases: 1) Amanda does not know if she has a genetic marker that would increase her possibility of suffering from a debilitating disease; 2) Beatrice is suffering from a disease now, but she has not received the diagnosis yet. If Amanda gets tested and receives a positive result, she will not be sure that she will suffer from that disease in the future. In the worst-case scenario, by being tested she would only know of a potential restriction of her future autonomy. In that case, she may choose to believe as if that restriction was a certainty, restricting her epistemic and practical autonomy even if she may never suffer from that particular disease. In this case, avoiding that knowledge would be an empowering choice for Amanda, which would increase her perceived epistemic and practical autonomy.

Beatrice, instead, is already in a condition that restricts her autonomy: receiving a diagnosis would allow her to take control and "ownership of her destiny" (Magnani, 2020). Choosing to not know, in her case, would amount to willful ignorance since she would not only need to avoid finding one information attainable

by a medical test, but she would also need to avoid acknowledging any symptoms of her disease, recurring to wishful thinking, self-deception, and other irrational patterns of reasoning.

At the same time, it seems understandable that people would avoid knowledge regarding future states of affairs, especially if they believe they have control over their developments. For example, in a series of studies, Gigerenzer and Garcia-Retamero (2017) asked if people would like to know, with certainty, if their marriage would last or not. Most people refused this possibility. While the authors claim that the "anticipated regret" was at the heart of this decision, we claim that "the loss of perceived autonomy" could very well be the content of that particular epistemic feeling. Indeed, if there were the possibility of foreseeing how long a marriage would last, then it would mean that people do not have any power to change the situation. They would believe that they are not in charge of their relationship. So, choosing not to know seems the only way to preserve their epistemic autonomy, if not autonomy in general. We can put forward almost the same affirmation, even considering the case in which the test would predict with high accuracy (not certainty) the result. How would people know if knowing the result of the test will not affect the duration of their marriage? Knowing would imply a gamble in epistemic autonomy: if knowing the result of the test would affect the perception of people's autonomy, then it is reasonable to stay in a state of not-knowledge and preserve the perception of full epistemic and practical autonomy on the length of their marriage.

So, by not knowing, people are able to overcome the paradox of autonomy: by avoiding or postponing the acquisition of specific knowledge, they can preserve the perception of their epistemic and practical autonomy, and they would not have reasons to doubt that it is genuine. This refusal of "specific certainty or semi-certainty" is also confirmed by studies from (Kruglanski et al., 2020, p. 416):

> Often, individuals crave specific certainty concerning beliefs they find reassuring, flattering, or otherwise pleasing. A student may prefer to know that they passed an exam, a patient may prefer to receive a clean bill of health, a suitor may prefer to have their affections returned. Similarly, one may avoid specific certainties that are troubling or threatening. Not knowing that one failed an exam is more pleasant than knowing that one did. Agnosticism concerning the alleged misconduct of one's child is preferable to unpleasant certainty in this matter. Avoidance of specific uncertainty can lead people to value ignorance.

Moreover, together with a curated perception of their autonomy, by not knowing certain information, people would also preserve a certain optimism that their future choices will be free and rational. So, in these cases, knowledge avoidance of certain data is less a preservation of a "blissful ignorance" and more a form of curation of a pragmatically valuable emotion: *hope*.

As Bloeser and Stahl (2017, p. 11) affirms, "hope is implicit in most pragmatic philosophies," since it has not only to do with agents' expectations and desires, but also with the possibility that certain things will happen and on the actions that agents need to perform to make sure their hopes are not in vain. We need to add that hope is necessary to preserve the perception of both practical and epistemic autonomy. Indeed, it does not only preserves the idea that the future may reserve positive events but also that people can reason, form beliefs, and justify their actions to make them happen. Thus, if it preserves the agent from emotional costs or loss of autonomy, the choice of not-knowing will also preserve hopeful considerations on the future, which, in turn, will help agents to reason and form beliefs toward the further preservation of their autonomy.

3.2. The Appeal to Autonomy: Comparing Cases

As a last consideration, we need to compare how the appeal to autonomy can help us make a case for the rationality of knowledge avoidance but not willful ignorance. Let us review two cases[7]:

- Clara is a lawyer. She thinks she can better defend her client if she believes her client is innocent. Defending her client to the best of her ability is what she wants to do. She acquires evidence that her client is guilty, and she engages in self-deception—she starts looking for reasons to reject the evidence, however strong. She wants to maintain her belief in her client's innocence as then she would be better able to act as she wishes.
- Denise is also a lawyer, with the same ambition of defending her client and the same belief that she will do a better job if she believes the client is innocent. She has the chance to read a potentially incriminating letter. To maintain her epistemic autonomy, she refuses to read it.

Clara engages in self-deception as the intentionalists describe it and so falls into willful ignorance in order to—allegedly—better serve her client. Instead, Denise seems to choose the less committing option of knowledge avoidance: without acquiring the potential evidence of her client's guilt, she allows herself to free her judgment of the idea that her client might be guilty. Even if the situations seem similar, we still maintain that, while Clara is limiting herself by falling into a self-deceiving state, Denise may still have a chance to increase her epistemic autonomy.

Clara, in fact, both *knows* that her client is guilty and is in a state of denial regarding this fact. At the same time, she is not preventing others from finding evidence for her client's guilt because she is fooling herself regarding the client's innocence. So, instead of considering, for example, extenuating circumstances for her client's actions—which would increase her ability to have a fair trial for her client and ultimately serve better the client's interests—she is just burying her head in the sand. So, she is not increasing in any way her pragmatic or epistemic autonomy: she is just trapping herself in a self-deceiving pattern of reasoning, limiting her options to defend her client better.

[7]We need to thank one of the anonymous reviewers for challenging our theory by providing us these captivating scenarios.

Instead, Denise is in a precarious situation: by not reading the potentially incriminating letter, she has avoided acquiring the belief that the client might be guilty—or not. So, if that letter is the only potential evidence of her client's guilt and only she could present the evidence in court, by avoiding reading the letter, she is precluding herself to either acquire further evidence of her client's innocence or to have the chance to get rid of the only evidence that may prove her client's guilt. If the letter does not contain evidence of guilt, Denise is just preserving a belief that the letter would confirm–that her client is innocent. If the letter contains evidence of her client's guilt, by not reading the letter, she is not putting herself in a position of choosing between defending her client as guilty or defending her client as innocent and destroying the evidence. If Denise is aware that she would not easily make this choice or she would choose to destroy the evidence, potentially ruining her career if caught, she is preserving both her epistemic and pragmatical autonomy by not reading the letter.

Now, suppose that Denise's client is guilty, and that letter is *not* the only potential evidence of her client's guilt: other people could present evidence of her client's guilt in court. In that case, she will need to face another choice: falling into self-deception as Clara did or still defending her client by looking for extenuating circumstances for the client's actions.

So, by considering all the options the two lawyers face, we can still defend the idea that people who avoid specific knowledge, if they are aware of the emotional toll the acquisition of that knowledge would take on them, do increase their epistemic autonomy, while willfully ignorant people do not.

4. CONCLUSIONS

In this paper, we offered some reasons to defend the rationality of knowledge avoidance. To fully explain this epistemic right to not-know, we have first distinguished between "willful ignorance" and "knowledge avoidance": while the former amounts to all cases in which people try to preserve a general state of ignorance (as doubt, uncertainty, indecision, etc.) also avoiding all circumstances that would allow them to stumble on particular knowledge by accident, the latter describes the agents' avoidance of a particular piece of information, which could not fall in their laps otherwise.

To defend the rationality of knowledge avoidance, we used takes from embodied cognition research and theories of bounded/ecological rationality (Goldstein and Gigerenzer, 2002; Bissoto, 2007; Spellman and Schnall, 2009; Xu et al., 2020). Even if the rationality-irrationality spectrum recently became more nuanced with the contribution of these theories, we reflected on the fact that there are still states and processes deemed irrational in the current literature. So, we asked ourselves which kind of deliberate not-knowing had a role in irrational patterns of reasoning, such as wishful thinking, self-deception, and akrasia, and we argued that, while willful ignorance has a significant role to play in these states, knowledge avoidance does not play a crucial part in most of them.

Then, we focused on the reasons for which knowledge avoidance could be considered rational. To proceed with our argumentation, we discussed the impact of certain feelings—epistemic ones—on people's reasoning abilities. Following the basic tenets of embodied cognition, we argued that the emotional impact of certain information should be considered among the costs of acquiring knowledge, contributing to judging certain situations of knowledge avoidance as rational. Moreover, we discussed the impact of knowledge avoidance on the agents' sense of autonomy, which also brought us to discuss the concepts of epistemic autonomy and the paradox of autonomy.

In sum, we maintained that if knowledge avoidance is fully conscious and agents are aware both of the information they are giving up and of the emotional impact that information would have if acquired, then rejecting to seek that knowledge is a form of rational and autonomy-increasing *hope-depended selection of information*. We, of course, do not claim that the appeal to autonomy is the only argument we can advance to defend the epistemic rationality of knowledge avoidance[8]. Nevertheless, we believe it is one reason to consider knowledge avoidance rational in a perspective of embodied bounded rationality.

AUTHOR CONTRIBUTIONS

SA wrote the first draft of the manuscript. Both authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

Open access funding were provided by MIUR, Ministry of University and Research, Rome, Italy (grant no. PRIN 2017 Research 20173YP4N3) and by the University of Pavia (INROAd+ − ENIGMA).

ACKNOWLEDGMENTS

We are profoundly grateful to Wendy Ross, Samantha Copeland, Marco Viola, Debbie Jenkins, and Alger Sans Pinillos for their valuable comments and constructive criticism on some ideas presented in this manuscript. We also need to express our gratitude toward the two anonymous referees, for their crucial remarks and knowledgeable suggestions.

[8]One of the anonymous reviewers pointed out, and we agree, that, for example, the account of the rationality of belief put forward by Fileva (2018) may justify the rationality of knowledge avoidance as well. On her view, the belief of the agent who engages in knowledge avoidance is consistent with the agent's evidence, so it does not violate evidential constraints, and so is at least minimally rational or rationally permissible even if it is not rationally ideal. Not so in the case of the willfully ignorant agent who forms beliefs contrary to the evidence.

REFERENCES

Arango-Muñoz, S. (2014a). Metacognitive feelings, self-ascriptions and mental actions. *Philos. Inquiries* 2, 145–162. doi: 10.4454/philinq.v2i1.81

Arango-Muñoz, S. (2014b). The nature of epistemic feelings. *Philos. Psychol.* 27, 193–211. doi: 10.1080/09515089.2012.732002

Arango-Muñoz, S., and Michaelian, K. (2014). Epistemic feelings, epistemic emotions: review and introduction to the focus section. *Philos. Inquiries* 2, 97–122.

Arfini, S. (2019). *Ignorant Cognition. A Philosophical Investigation of the Cognitive Features of Not-Knowing*. Cham: Springer.

Arfini, S. (2021). Situated ignorance: the distribution and extension of ignorance in cognitive niches. *Synthese* 198, 4079–4095. doi: 10.1007/s11229-019-02328-0

Arfini, S., and Magnani, L. (2021). *Embodied, Extended, Ignorant Minds. New Studies on the Nature of Not-Knowing*. Cham: Synthese Library. Springer International Publishing.

Bardone, E. (2011). *Seeking Chances: From Biased Rationality to Distributed Cognition*. Berlin; Heidelberg: Springer Science & Business Media.

Barnes, A. (2007). *Seeing Through Self-Deception*. Cambridge: Cambridge University Press.

Bastardi, A., Uhlmann, E. L., and Ross, L. (2011). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychol. Sci.* 22, 731–732. doi: 10.1177/0956797611406447

Bertolotti, T., Arfini, S., and Magnani, L. (2016). Abduction: From the ignorance problem to the ignorance virtue. *J. Logics Appl.* 3, 153–173.

Bissoto, M. L. (2007). Self-organization, embodied cognition and the bounded rationality concept. Ciências Cognição 11, 80–90.

Bloeser, C., and Stahl, T. (2017). *Hope*. Stanford: The Stanford Encyclopedia of Philosophy.

Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

Bortolotti, L. (2014). *Irrationality*. Malden, MA: John Wiley & Sons.

Buss, S., and Westlung, A. (2018). *Personal Autonomy*. Stanford: The Stanford Encyclopedia of Philosophy.

Chemero, A. (2011). *Radical Embodied Cognitive Science*. Cambridge: MIT press.

Coates, A. (2020). Rational epistemic akrasia. *Am. Philos. Q.* 49, 113–124.

Daoust, M.-K. (2019). Epistemic akrasia and epistemic reasons. *Episteme* 16, 282–302. doi: 10.1017/epi.2018.6

Davidson, D. (2004). *Problems of Rationality*. Oxford: Oxford University Press.

Deweese-Boyd, I. (2021). *Self-Deception*. Stanford: The Stanford Encyclopedia of Philosophy.

Eil, D., and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *Am. Econ. J. Microecon.* 3, 114–138. doi: 10.1257/mic.3.2.114

Fileva, I. (2018). What does belief have to do with truth? *Philosophy* 93, 557–570. doi: 10.1017/S0031819118000335

Fischer, J. M., and Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *J. Philos.* 68, 5–20. doi: 10.2307/2024717

Gallagher, S. (2018). "Embodied rationality," in *The Mystery of Rationality*, eds G. Bronner and F. Di Iorio (Cham: Springer International Publishing), 83–94.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis* 23, 121–123. doi: 10.1093/analys/23.6.121

Gigerenzer, G., and Garcia-Retamero, R. (2017). Cassandra's regret: the psychology of not wanting to know. *Psychol. Rev.* 124, 179–196. doi: 10.1037/rev0000055

Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–669. doi: 10.1037/0033-295X.103.4.650

Goldstein, D. G., and Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychol. Rev.* 109, 75–90. doi: 10.1037/0033-295X.109.1.75

Greco, D. (2014). A puzzle about epistemic akrasia. *Philos. Stud.* 167, 201–219. doi: 10.1007/s11098-012-0085-3

Grossman, Z., and van der Weele, J. J. (2017). Self-image and willful Ignorance in social decisions. *J. Eur. Econ. Assoc.* 15, 173–217. doi: 10.1093/jeea/jvw001

Haack, S., and Kolenda, K. (1977). Two fallibilists in search of the truth. *Proc. Aristotelian Soc. Suppl.* 51, 63–104. doi: 10.1093/aristoteliansupp/51.1.63

Haas, J., and Vogt, K. M. (2015). "Ignorance and investigation," in *Routledge International Handbook of Ignorance Studies*, eds M. Grossand L. McGoey (Abingdon: Routledge), 17–25.

Jefferson, A., Bortolotti, L., and Kuzmanovic, B. (2017). What is unrealistic optimism? *Conscious Cogn.* 50, 3–11. doi: 10.1016/j.concog.2016.10.005

Johnston, M. (1995). "Self-deception and the nature of mind," in *Philosophy of Psychology: Debates on Psychological Explanation*, ed C. Macdonald (Cambridge: Blackwell), 63–91.

Jordan, J. (1996). Pragmatic arguments and belief. *Am. Philos. Q.* 33, 409–420.

Kruglanski, A. W., Jasko, K., and Friston, K. (2020). All thinking is 'wishful' thinking. *Trends Cogn. Sci.* 24, 413–424. doi: 10.1016/j.tics.2020.03.004

Le Morvan, P. (2013). Why the standard view of ignorance prevails. *Philosophia* 41, 239–256. doi: 10.1007/s11406-013-9417-6

Le Morvan, P., and Peels, R. (2016). "The nature of ignorance: two views," in *The Epistemic Dimensions of Ignorance*, eds R. Peels and M. Blaauw (Cambridge: Cambridge University Press), 12–32.

Mackie, G. (2012). "Rational ignorance and beyond," in *Collective Wisdom: Principles and Mechanisms*, eds H. Landemore and J. Elster (Cambridge: Cambridge University Press), 290–318.

Magnani, L. (2020). Autonomy and the ownership of our own destiny: tracking the external world and human behavior, and the paradox of autonomy. *Philosophies* 5, 12. doi: 10.3390/philosophies5030012

Mastrogiorgio, A., and Petracca, E. (2016). "Embodying rationality," in *Model-Based Reasoning in Science and Technology*, eds L. Magnani and C. Casadio (Cham: Springer International Publishing), 219–237.

May, V. M. (2006). Trauma in paradise: willful and strategic ignorance in *Cereus Blooms at Night. Hypatia* 21, 107–135. doi: 10.1111/j.1527-2001.2006.tb01116.x

Mayraz, G. (2011). Wishful thinking. *SSRN Electronic Journal, Paper n 1955644*.

McIntyre, L. (2015). *Respecting Truth: Willful Ignorance in the Internet Age*. London: Routledge.

Mele, A. (2000). *Self-Deception Unmasked*. Princeton: Princeton University Press.

Oster, E., Shoulson, I., and Dorsey, E. R. (2013). Optimal expectations and limited medical testing: Evidence from huntington disease. *Am. Econ. Rev.* 103, 804–830. doi: 10.1257/aer.103.2.804

Owens, D. (2002). Epistemic akrasia. *Monist* 85, 381–397. doi: 10.5840/monist200285316

Pedrini, P. (2012). What does the self-deceiver want? *Humana Mente* 20, 141–157.

Pedrini, P. (2018). Liberalizing self-deception. les ateliers de l'Ethique. *Ethics Forum* 13, 11–24. doi: 10.7202/1059496ar

Polanyi, M. (1966). *The Tacit Dimension*. London: Routledge & Kegan Paul.

Reisner, A. (2009). The possibility of pragmatic reasons for belief and the wrong kind of reasons problem. *Philos. Stud.* 145, 257–272. doi: 10.1007/s11098-008-9222-4

Rubin, D. I. (2018). Willful ignorance and the death knell of critical thought. *New Educator*. 14, 74–86. doi: 10.1080/1547688X.2017.1401192

Sicherman, N., Loewenstein, G., Seppi, D. J., and Utkus, S. P. (2016). Financial attention. *Rev. Financ Stud.* 4, 863–897. doi: 10.1093/rfs/hhv073

Sigall, H., Kruglanski, A., and Fyock, J. (2000). Wishful thinking and procrastination. *J. Soc. Behav. Pers.* 15, 283–296.

Simon, H. A. (1997). *Models of Bounded Rationality*. Cambridge: MIT Press.

Somin, I. (2015). "Rational ignorance," in *Routledge International Handbook of Ignorance Studies*, eds M. Gross and L. McGoey (Abingdon: Routledge), 274–281.

Spellman, B. A., and Schnall, S. (2009). Embodied rationality. *Virginia Public Law and Legal Theory Research, 35:Paper No. 17*.

Star, D. (ed.). (2018). *The Oxford Handbook of Reasons and Normativity, 1st Edn*. Oxford, United Kingdom; New York, NY: Oxford University Press.

Sweeny, K., Melnyk, D., Miller, W., and Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Rev. General Psychol.* 14, 340–353. doi: 10.1037/a0021288

Todd, P. M., and Gigerenzer, G. (2007). Environments that make us smart: ecological rationality. *Curr. Dir. Psychol. Sci.* 16, 167–171. doi: 10.1111/j.1467-8721.2007.00497.x

Todd, P. M., and Gigerenzer, G. E. (2012). *Ecological Rationality: Intelligence in the World*. Oxford: Oxford University Press.

Werner, K. (2021). Cognitive confinement: theoretical considerations on the construction of a cognitive niche, and on how it can go wrong. *Synthese* 198, 6297–6328. doi: 10.1007/s11229-019-02464-7

Williams, D. (2021). Motivated ignorance, rationality, and democratic politics. *Synthese* 198, 7807–7827. doi: 10.1007/s11229-020-02549-8

Woods, J. (2005). "Epistemic bubbles," in *We Will Show Them! Essays in Honour of Dov Gabbay, Vol. 2*, eds S. Artemov, H. Barringer, A. d'Avila Garcez, L. C. Lamb and J. Woods (London: College Publications), 731–774.

Xu, F., Xiang, P., and Huang, L. (2020). Bridging ecological rationality, embodied emotion, and neuroeconomics: Insights from the somatic marker hypothesis. *Front. Psychol.* 11:1028. doi: 10.3389/fpsyg.2020.01028

Zimmerman, M. J. (2018). Recklessness, willful Ignorance, and exculpation. *Crim. Law Philos.* 12, 327–339. doi: 10.1007/s11572-017-9424-y

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.

Copyright © 2021 Arfini and Magnani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fpsyg.2022.805743

More Thumbs Than Rules: Is Rationality an Exaptation?

Antonio Mastrogiorgio[1], Teppo Felin[2,3], Stuart Kauffman[4] and Mariano Mastrogiorgio[5]*

[1] IMT School for Advanced Studies Lucca, Lucca, Italy, [2] Huntsman School of Business, Utah State University, Logan, UT, United States, [3] Saïd Business School, University of Oxford, Oxford, United Kingdom, [4] Institute for Systems Biology (ISB), Seattle, WA, United States, [5] Department of Strategy, IE University, Segovia, Spain

The literatures on bounded and ecological rationality are built on adaptationism—and its associated modular, cognitivist and computational paradigm—that does not address or explain the evolutionary origins of rationality. We argue that the adaptive mechanisms of evolution are not sufficient for explaining human rationality, and we posit that human rationality presents exaptive origins, where exaptations are traits evolved for other functions or no function at all, and later co-opted for new uses. We propose an embodied reconceptualization of rationality—embodied rationality—based on the reuse of the perception-action system, where many neural processes involved in the control of the sensory-motor system, salient in ancestral environments have been later co-opted to create—by tinkering—high-level reasoning processes, employed in civilized niches.

Keywords: exaptation, embodied rationality, bounded rationality, heuristics, neural reuse, spandrels

OPEN ACCESS

Edited by:
Vittorio Gallese,
University of Parma, Italy

Reviewed by:
Marco Viola,
University Institute of Higher Studies in Pavia, Italy
Guido Baggio,
Roma Tre University, Italy
Antonella Tramacere,
University of Bologna, Italy

****Correspondence:***
Antonio Mastrogiorgio
mastrogiorgio.antonio@gmail.com

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *30 October 2021*
Accepted: *03 January 2022*
Published: *24 February 2022*

Citation:
Mastrogiorgio A, Felin T, Kauffman S and Mastrogiorgio M (2022) More Thumbs Than Rules: Is Rationality an Exaptation? Front. Psychol. 13:805743. doi: 10.3389/fpsyg.2022.805743

INTRODUCTION

> I counted the panda's other digits and received an even greater surprise: there were five, not four. Was the "thumb" a separately evolved sixth finger?
>
> S. J. Gould, *The Panda's Thumb*

In Herbert A. Simon's view, heuristics are *rules of thumb*—instantiations of bounded rationality—that produce solutions adapted to specific task environments, given limited information, time and cognitive capabilities. This adaptive dimension is part of the very definition of *bounded rationality*, also known as the scissors' metaphor: "Just as a scissors cannot cut paper without two blades, a theory of thinking and problem solving cannot predict behavior unless it encompasses both an analysis of the structure of task environments and an analysis of the limits of *rational adaptation* to task requirements" (Newell and Simon, 1972, p. 55, emphasis added). The adaptive view of rationality—'rational adaptation' in Simon's own words—has been transposed into contemporary views of heuristics. While Kahneman's 'heuristics and biases' focus on the nature of dis-adaptation, in the sense that heuristics, automatically triggered, do not fit specific task environments (Tversky and Kahneman, 1974; Kahneman et al., 1982), Gigerenzer's 'ecological rationality' emphasizes the fact that fast and frugal heuristics produce satisficing solutions, through ecological correlations (Gigerenzer and Selten, 2002; Gigerenzer and Gaissmaier, 2011).

In this contribution, we criticize the emphasis on the adaptative logic (e.g., Cosmides and Tooby, 2006; Tooby and Cosmides, 2007), arguing that adaptive mechanisms are not a unique or sole explanation for human rationality. Relying on the old but still relevant critique to adaptationism—initially raised by Stephen J. Gould (i.e., Gould and Lewontin, 1979; see also Andrews et al., 2002)—we discuss a fundamental limit of adaptative explanations: the difficulty to make a distinction between contingent utility and reasons for origins. For instance, consider the popular and frequently discussed gaze heuristic (e.g., Gigerenzer and Gray, 2017; Hamlin, 2017; Höfer et al., 2018), which is used to track the motion of a moving goal by keeping constant the angle between a catcher and the goal. The fact that a soccer player uses the gaze heuristic to catch a ball (contingent utility) tells us nothing about how such a heuristic came into being in the first place (reasons for origins). We of course realize that the gaze heuristic did not come into being for playing soccer, since it was present thousands of years before soccer was invented and other species also use it (in particular it is heavily used by predators to catch prey). In short, the idea that heuristics are effective decision rules for contingent task environments does not strictly explain their origins.

We argue that exaptive mechanisms are fundamental for explaining the origins of rationality. While adaptations are traits gradually evolved *via* natural selection in order to meet pre-existing functions, *exaptations* are traits evolved for other functions, or no function at all, and later co-opted for new uses (Gould and Vrba, 1982). We propose an embodied reconceptualization of rationality—so-called, *embodied rationality* (e.g., Mastrogiorgio and Petracca, 2012, 2015, 2016)—based on (non-adaptive but) exaptive evolutionary mechanisms. In particular, we amend and reconceptualize bounded and ecological rationality, by discussing the reuse of the perception-action system: many neural processes involved in the control of the sensory-motor system, which were salient in ancestral environments, have been later co-opted to shape—*via* tinkering—high-level cognitive faculties employed in civilized niches.

RATIONALITY AND ADAPTATION

Evolutionary explanations are elegant from the point of view of Occam's Razor: by identifying some criteria to explain the factual diversity of nature, evolutionary theories aim to establish theoretically plausible solutions to the problem of origins. In its general definition, also known under the general label of Universal Darwinism, evolution is instantiated by the processes of variation, selection and retention—processes that account for the diversity that composes natural and cultural systems (cf. Campbell, 1960; Lewontin, 1970; Dawkins, 1983; Hodgson, 2005).

The program of research on bounded rationality, started by Herbert A. Simon, owes much to evolutionary frameworks, as it is argued that human behaviors must be studied with respect to specific task environments. Simon emphasizes that minds are adapted to real-world environments and must be evaluated in terms of their adequacy to specific environmental instances: in the scissors' argument (summarized above), cognition and environment are two cutting blades that make sense precisely because they are conjointly defined in a unitary analytical framework (Simon, 1956, 1990; Newell and Simon, 1972).

Modern transpositions of bounded rationality are sympathetic to this adaptive, evolutionary framework. On the one side, Kahneman's *heuristics and biases* are based on the evidence that specific heuristics, which are automatically triggered, violate specific rules of logic and probability so as to produce biased judgments (Tversky and Kahneman, 1974; Kahneman et al., 1982). On the other side, Gigerenzer's *ecological rationality* relies on the idea that fast and frugal heuristics, by exploiting ecological correlations, provide satisficing solutions to specific task environments (Gigerenzer and Goldstein, 1996; Gigerenzer and Gaissmaier, 2011), where 'satisficing' is a well-known neologism coined by Simon, given by the combination of 'satisfy' and 'suffice.'

These two colliding research programs are at the center of so-called 'rationality wars' (see Samuels et al., 2004 for a discussion). In fact, Gigerenzer criticizes Kahneman's heuristics as being "vague, undefined, and unspecified with respect both to the antecedent conditions that elicit (or suppress) them and also to the cognitive processes that underlie them" (Gigerenzer, 1996, p. 592). Generally speaking, Gigerenzer criticizes the incorrectness of deducing a positive framework of rationality by relying on the experimental evidence built upon the normative benchmarks based on general rules of logic and probability calculus. This critique (Gigerenzer and Murray, 1987; Gigerenzer, 1991) created a dialectical interaction and debate (see the replies of Gigerenzer, 1996; Kahneman and Tversky, 1996). In such rationality wars, while Gigerenzer's view remains "panglossian," in the sense that fast and frugal heuristics generate satisficing solutions, Kahneman's view is "meliorist," in the sense that heuristics are sources of cognitive errors and biases and produce misfits with respect to specific normative requirements.

Inside the Adaptive Toolbox

Fast and frugal heuristics generate satisficing outcomes to the extent that they fit the specific structure of the task environment. This idea is a pillar of ecological rationality, which investigates "in which environments a given strategy is better than other strategies" (Gigerenzer and Gaissmaier, 2011). The research program on ecological rationality, developed by Gigerenzer and Gaissmaier (2011), emphasizes that heuristics compose an *adaptive toolbox*, where human behavior is described by a series of "cognitive heuristics, their building blocks (e.g., rules for search, stopping, and decision), and the core capacities (e.g., recognition memory) they exploit." In Gigerenzer (2008, p. 20) own words: "The adaptive toolbox is a Darwinian-inspired theory that conceives the mind as a modular system that is composed of heuristics, their building blocks, and evolved capacities."

Using an adaptive framework to explain the nature of heuristics would imply that heuristics are (casual) variations selected by the environment because of their comparatively better fitness, which are then retained. For instance, let us consider the evidence that some fast and frugal heuristics predict heart attack

in a manner that is comparable with complex, effortful (and slow) medical procedures (Todd and Gigerenzer, 2000; Marewski and Gigerenzer, 2012). If we adopt a strict, adaptationist, evolutionary framework—based on variation, selection and retention—, we should hypothesize that these heuristics are the result of a gradual refinement of older ones, that have been selectively retained by the environment. Despite the emphasis on the adaptive toolbox, theoretical inquiries on the role of selection—which is fundamental for adaptation—seem to be overlooked in ecological and bounded rationality, where the selective mechanisms remain underexplored.

Interestingly, the absence of theorizing on the nature of adaptation is not a bug of ecological rationality but a deliberate theoretical choice. On this point, Hutchinson and Gigerenzer (2005; see also Sanabria and Killeen, 2005) clarify that the adaptive view of heuristics is not an argument about their origins. That is, the fact that a heuristic is ecologically rational does not imply that it has been shaped by the forces of selection for that specific task. Interestingly, ecological rationality avoids both trivial adaptive explanation ("just-so stories") and the necessity of accurate theorizing on the nature of origins: "It thus would be a weak argument [. . .] to find a heuristic that humans use, then search for some environment in which that heuristic works well, and then claim on this basis alone that the heuristic is an adaptation to that environment. The heuristic may work well in that environment, but that need not be the reason why it evolved or even why it has survived" (Hutchinson and Gigerenzer, 2005, p. 109; see also Navarrete and Santamaría, 2011). This clarification, though, looks like an *excusatio non-petita*, where the caveat substitutes the claim for an adaptive toolbox. As the authors add (p. 109): "Ecological rationality might then be useful as a term indicating a more attainable intermediate step on the path to a demonstration of adaptation. There is nevertheless a risk that a demonstration of ecological rationality of a given heuristic in a given environment will mislead someone who uses this evidence alone to infer adaptation."

The risk of making casual and cursory claims about the evolutionary origins of heuristics is real, but the above clarification is hopelessly insufficient. That is, if we cannot infer adaptation, then why even speak of an *adaptive* toolbox? The authors' clarification—about the problem of inferring adaptation from ecological rationality of a given heuristics—therefore, raises important issues, which are crucial in the critique of adaptationism.

Adaptationism at Stake

For decades, S. J. Gould, in his broad program of research, has tried to demonstrate that there is an unjustified, paradigmatic correspondence between the general problem of evolution as originally formulated by Charles Darwin and its transposition into the neo-Darwinian synthesis. According to the dominant paradigm, there are no radical alternatives to 'adaptationism.' When scholars in different fields refer to evolutionary theories, they are, implicitly and unwittingly, appealing to a mechanism that they consider necessary and sufficient: adaptation. Therefore, according to Gould—and consistently with Kuhn's idea of a paradigmatic science—adaptationism signals more a faith in evolutionary theorizing (where the risk of "just-so stories" is always present) than a deep understanding of its related questions. The theoretical mechanisms and implications of adaptationism are a matter that cannot be informally treated in a few words. Generally speaking, as suggested by Gould (and by Darwin himself), the laws of change, more than a nomothetic necessity, should be considered as extrapolations whose instrumental value is the understanding of empirical evidence (Gould and Eldredge, 1977; Gould and Lewontin, 1979; Gould and Vrba, 1982; Gould, 2002, see also Williams, 1966). In particular, according to Gould and Lewontin (1979, p. 581), adaptationism "is based on faith in the power of natural selection as an optimizing agent. It proceeds by breaking an organism into unitary 'traits' and proposing an adaptive story for each considered separately. Trade-offs among competing selective demands exert the only brake upon perfection; non-optimality is thereby rendered as a result of adaptation as well".

Although we acknowledge the disapproval of some scholars (in particular John Maynard Smith and Richard Dawkins) of Gould's ideas (see Gould, 1997 for a reply), we believe that Gould's critique of adaptationism matters a great deal for understanding the contemporary state of the art of human rationality, when related with alternatives to adaptive processes. Gould criticizes adaptationism for its unwillingness to consider alternatives to adaptive processes. According to Gould and Vrba (1982, p. 5), there are two meanings of the word 'adaptation': "the first is consistent with the vernacular usage [. . .]: a feature is an adaptation only if it was built by natural selection for the function it now performs. The second defines adaptation in a static or immediate way as any feature that enhances current fitness regardless of its historical origin." Adaptationism fails because of its impossibility to make a distinction between current utility and reasons of origin. Importantly, the fact that a trait satisfies (more or less effectively) a particular function cannot strictly be an explanation of its origins.

According to this critique, adaptation often presumes an unjustified teleological perspective, where things (like biological traits, cognitive faculties or, in our case, rationality) are explained in terms of their final causes, and implicitly represent the best state of the world precisely in virtue of their existence. This part of Gould's critique—according to which adaptationism is Panglossian—calls to the mind Voltaire's novel of Candido. Things are made for the best purpose, as suggested by Dr. Pangloss (a character of the novel): "Legs were clearly intended for breeches, and we wear them." The consideration of the current traits as adaptations often collapses onto the problematic statement that they represent the best status possible in nature, or, quoting again Dr. Pangloss: "Things cannot be other than they are."

Generally speaking (as we discussed in the previous Section "Inside the Adaptive Toolbox"), both ecological and bounded rationality tend to propose an adaptive view of heuristics, where heuristics are instrumental to the evolution of complex cultures (for instance, Finnish mushroom foragers have learned some rules of thumb to deal with poisonous species, Kaaronen, 2020). Oddly, bounded and ecological rationality overlook the role of selective processes, which might play a fundamental role in the

origins of heuristics. This somehow represents an inconsistency, since selection is a crucial mechanism of adaptation. The reasons for this apparent inconsistency are essential and find an adequate answer in Gould's critique of adaptationism. Bounded and ecological rationality implicitly assume that adaptive mechanisms require *contingent utility*: that is, if a heuristic works in a specific task environment—in the sense that it produces a satisficing outcome—, then it is adapted or 'ecologically rational,' in the words of Gerd Gigerenzer. However, the comparatively better performance of a heuristic with respect to alternatives is not, strictly speaking, an explanation of its origins. As we will explain in the next sections, the distinction between 'contingent utility' and 'reasons for origins' thus represents a fundamental argument that can cast new light on non-adaptive mechanisms at the origins of human rationality.

THE EXAPTIVE ORIGINS OF RATIONALITY

The use of adaptive explanations to understand the nature of human rationality represents a significant innovation with respect to standard economic theories, which assume that economic agents are optimizers, thus possessing unconstrained knowledge, time and computational power. Arguments based on bounded rationality approach reasoning processes not in absolute terms (i.e., in terms of logic and probability rules that abstract from a specific environment), but as a matter of domain-specific adequacy of reasoning processes to specific environmental instances. However, such an adaptive framework suffers from the same limits that characterize adaptationist explanations in evolutionary theory. As we discussed in Section "Rationality and Adaptation," we cannot, strictly speaking, explain the origins of heuristics by considering their contingent utility.

Insights for understanding the exaptive nature of rationality can be drawn from a more pluralistic view on evolution, often called 'extended evolutionary synthesis' (Gould, 1982; Pigliucci and Müller, 2010; Laland et al., 2015), which also includes the theory of punctuated equilibrium (Eldredge and Gould, 1972). Far from being an alternative to the Darwinian framework—a paradigm-shift, *stricto sensu*—the extended evolutionary synthesis calls for an exegesis of the original ideas of Darwin, who argued that selection, despite being central, is non-exclusive. As put by Darwin: "it has been stated that I attribute the modification of species exclusively to natural selection [. . .] I am convinced that natural selection has been the main, but not the exclusive means of modification" (Darwin, 1872). In other words, the extended evolutionary synthesis claims the right to adopt a pluralistic approach to evolutionary explanations, challenging the limitations of the modern synthesis.

An extended taxonomy of fitness suggests that we should include exaptive mechanisms, along with adaptive ones, in evolutionary frameworks (Gould, 2002). Such an extension does not deny that adaptation is a fundamental mechanism of evolution, as it just places emphasis on the fact that mutations can occur only 'given a structure': current structures, *de facto*, constrain the possibility of evolution more than selective mechanisms effectively do. As suggested by Gould and Lewontin (1979, p. 581; see also Andrews et al., 2002), "the constraints themselves become more interesting and more important in delimiting pathways of change than the selective force that may mediate change when it occurs." According to this perspective, evolution is more a matter of possibilities and constraints than a matter of optimal fit to a given environment. In particular, Gould and Lewontin, in their manifesto—The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist program—propose a structuralist view of evolution, that separates function from structure, by focusing on the so-called *spandrels*. Spandrels are architectural features, given by the roughly triangular spaces between the top of an arch and the ceiling (like the ones in the Basilica of San Marco Basilica in Venice). Evolutionary speaking, spandrels are phenotypic traits that arose as a by-product of evolution of some other traits rather than a product of adaptive selection.

Under such a framework, exaptive mechanisms are complementary to adaptive ones. The basic idea of an extended taxonomy is that a trait contributing to fitness is, simply speaking, an 'aptation', until we have sufficient evidence to consider it an ad-aptation or an ex-aptation (see Pievani and Serrelli, 2011). The word 'ex-aptation', which etymologically contrasts with 'ad-aptation' (where '*ex*' is the Latin correspondent of 'from' and '*ad*' is the Latin correspondent of 'to'), refers to the process through which existing traits, originally developed for a certain use, are employed for uses that are entirely different from the original one (see **Table 1**). That is, *exaptations* are "characters, evolved for other usage (or no function at all) and later "co-opted" for their current role" (Gould and Vrba, 1982, p. 6). That is, exaptations can be based on either a (1) functional shift or a (2) cooptation of a non-aptation. A common example of functional shift are the feathers of birds, first evolved for thermal regulation, then co-opted for flight. An example of a cooptation of a non-aptation are the sutures in the skulls of young mammals, a byproduct of the laws of growth and then co-opted for aiding parturition. Hence, characteristics or traits must be evaluated not only in terms of effectiveness to a pre-stated *function* but with respect to the affordable *effects* they can produce because of their specific morphological features.

Now, in order to understand the exaptive origins of heuristics, let us consider, again, the gaze heuristic (defined in Section "Introduction," see Raab and Gigerenzer, 2005; Hamlin, 2017; Höfer et al., 2018), which requires an ongoing adjustment between gaze and movement, linking perception and action, to accomplish tasks such as catching a prey or a ball. We can, speculatively, consider the gaze heuristic as a:

(i) *Cooptation from non-aptation.* The morphological traits of the human body strongly constrain the heuristics that can be generated. We can easily realize that the link between gaze and movement, which requires a well-developed sensory-motor integration between perception and action, is a necessary condition for the come-into-being of the gaze heuristic. This consideration seems trivial, unless we consider that this integration is not obvious, since we can reasonably hypothesize that, for some species, such integration is not well developed. For such species the gaze heuristic would not be an affordable option (e.g., we

TABLE 1 | A typology of fitness.

Process	Character		Usage
Functional adaptation Natural selection shapes the character for a current use	**Ad-aptation**	**Aptation**	Function
Functional shift A character, previously shaped by natural selection for a particular function (an adaptation), is co-opted for a new use	**Ex-aptation**		Effect
Cooptation of non-aptation A character, whose origins cannot be ascribed to the direct action of natural selection (a non-aptation), is co-opted for a current use			

Adapted from Gould and Vrba (1982).

can, speculatively, hypothesize that sloths do not use the gaze heuristic). If we consider this integration as a non-aptation, as we do not assume that this integration is the product of natural selection, we can hypothesize that this integration has been co-opted to create a mechanism—the gaze heuristic—that presents specific advantages (i.e., catching prey) in specific ancestral environments.

(ii) *Functional shift*. If we hypothesize that the gaze heuristic was selected in ancestral environments to be incrementally refined as a successful strategy for catching prey, we can easily realize that such a heuristic has been further exapted, as it readily admits a number of different 'applications' in non-ancestral environments. Far from being domain-specific, the gaze heuristic is successfully applied in pedestrian behavior, sailing, landing an airplane, kicking a ball, and so on. Such functional shifts require an isomorphism in the structure of the environment, where the spatial analogies are meaningful. However, analogical correspondence is not necessary. As in the case of feathers of birds, first evolved for thermal regulation then used for flight, we do not always need an isomorphism.

Cross-Level Mechanisms: Exapting the Perception-Action System

Exaptive mechanisms are not bounded to a specific level but operate, upward and downward, in the hierarchy—genes, organisms and species—of biological systems (Vrba and Gould, 1986). Generally speaking, an adaptation at one level could become an exaptation at another level, as the unit of selection pertains to all the different levels of biological organization (Lewontin, 1970). Interestingly, cross-level mechanisms could represent an interface between nature and culture (Uchiyama et al., 2020), where genetic adaptations could be culturally exapted. For instance, skin pigmentation—a genetic adaptation (a protection against UV radiation)—has been culturally exapted to become a signal of socio-economic status, because over centuries light pigmentation was a signal of high socio-economic status and prestige. But during the 60s, tanning started to be associated with wealth, leisure and prosperity. Gaze heuristics, in this regard, is quite representative, as—we suppose—it first glimmered as a cooptation of non-aption, and then, after the refinements of selective adaptation, it was further exapted in cultural environments through functional shifts.

Here we embrace the hypothesis that such cross-level mechanisms connote the exaptation of the perception-action system, where the sensory-motor devices and their mechanism admit very different applications—high-level cognitive faculties involved in reasoning processes—with respect to the ones for which they evolved through functional adaptation. In particular, an unconstrained view of the unit of selection, along with the presence of exaptations at different levels, allows for the integration of evolutionary and task-relevant timescales. We propose to conceptualize heuristics not only in terms of their original adaptive function (many of them arose in hunter-gatherer environments), but also with respect to their contingent effects, based on exaptive mechanisms, in civilized niches (For instance, the gaze heuristic presents quite different applications such as landing a plane or catching the ball, in civilized environments).

The morphological features of organisms—*bauplan* or *baupläne* (plural)—constrain which heuristics can be developed in the first place, before they are subject to any selective pressures. Hence, human rationality can be considered an 'adjacent possible' (cf. Kauffman, 2000) since the enabling constraints, exerted by the sensory-motor system in delimiting the evolution of higher cognitive faculties, are probably more important than the selective forces in mediating changes, once they've occurred. (Notice that we use the expression 'sensory-motor system' in the singular form as an exemplification, but we are aware that there is a multitude of sensory-motor systems.) Our perspective shifts the emphasis from selective processes to the mechanisms of variation, where the randomness of mutations deserves a clarification. That is, mutations are not equiprobable, as extant biological structures delimit the degrees of freedom and the probabilistic topology of evolutionary possibilities.

The exploitation of exaptive arguments in psychology is not new. However, they are not the orthodoxy, since exaptive arguments have been strongly criticized and ostracized (see Buss et al., 1998 for a critical review of exaptation in psychology). In spite the dialectical role of exaptive arguments in evolutionary psychology (Gould, 1991), we cannot avoid mentioning the heated debate where Darwinian fundamentalism was opposed to a more pluralistic approach (see Dennett, 1997; Gould, 1997; Kalant et al., 1997). We believe and, in this regard, we agree with a number of scholars—though this is not the place for this discussion—that a significant part of the arguments for ostracisms are not persuasive and can be readily

neutralized with a philological exegesis of S. J. Gould's arguments (Lloyd and Gould, 2017).

Neural Reuse

A fundamental evolutionary argument for an embodied reconceptualization of rationality proposed here consists in the general hypothesis that the brain might be seen as structured into layers, each one presenting a distinctive evolutionary dimension. Neural circuits evolved for specific uses (ad-aptations) or no uses at all (non-aptations) can be co-opted for novel uses, while retaining their original function. This view has gained importance in the last decade, under the general hypothesis that evolutionarily-speaking older brain areas are recruited to support different and relatively novel cognitive functions: recent layers (dedicated to high-level cognitive processes) exploit the 'lower' layers, dedicated to the sensory-motor control (Anderson, 2007a).

This perspective can captured under the notion of *neural reuse*, an umbrella term for a heterogenous group of overlapping theories (see Rathkopf, 2021) sharing this view of the brain. This group includes the 'Massive Redeployment Hypothesis' (Anderson, 2007a, 2010), 'neuronal recycling' (Dehaene and Cohen, 2007), 'neural exploitation' (Gallese, 2008), 'neural repurposing' (Parkinson and Wheatley, 2015), 'cognitive recycling' (Barack, 2017) and 'neural exaptation' (Chapman et al., 2017).

Neural reuse, in particular in its version known as the 'Massive Redeployment Hypothesis', represents an alternative to both strict brain localization of cognitive functions—the orthodox position on the functional topography of the brain—and holistic approaches to the brain (Anderson, 2007b, 2014, see also Favela, 2021 for a critical discussion of the evolutionary foundations of neural reuse). This novel view of the brain, with respect to the orthodox perspective—where specific cognitive functions are strictly localized in specific, non-overlapping areas—builds upon a methodological consideration: subtractive methods used in brain imaging are problematic when it comes to interpreting data and the conclusions these interpretations support. As put by Anderson (2007b, p. 148), "while difference images can show areas that participate in one task and not another, they cannot show that the area is limited to that task." Indeed, the evidence is that "there are very few specialists in the brain, supporting only tasks from a single task category such as semantics or visual perception. Most regions of the brain are active during multiple tasks in different task categories" (Anderson, 2016, p. 2; see also Anderson et al., 2013). There is of course ongoing debate about whether specific areas of the brain can be mapped onto specific cognitive processes (e.g., Poldrack, 2006). But, the Massive Redeployment Hypothesis posits that a typical cognitive function involves more than one brain area, and each brain area may be redeployed in support of other cognitive functions, according to a three-tier architecture characterized by many-to-many relationship between each level (brain area, component function, and functional complex). Again, in the words of Anderson (2007b, p. 163), we expect "each functional complex to have more than one component, each of which in turn will involve more than one area; likewise, we should expect areas to be members of more than one component, and components to be members of more than one functional complex, and we should not expect that such cross-participation will respect traditional functional-anatomical boundaries."

Evidence of the Massive Redeployment Hypothesis (where 'massive' indicates that redeployment is the norm in the brain), such as the sensorimotor coding in working memory or motor simulations in language understanding (see Anderson, 2006, 2007b), thus represents an argument in favor of cross-level exaptive mechanisms, where lower brain areas participate in higher cognitive processes. Such a thesis is also corroborated by the evidence that brain lesions can produce deficits across multiple domains, thus representing a solid counterargument to the modularity thesis (Prinz, 2006).

We believe that the Massive Redeployment Hypothesis can serve as a building block for an embodied foundation of human rationality. Indeed, the Massive Redeployment Hypothesis, by its own tenets, sheds light not only on the origins of novel, high-level, cognitive functions (in which a given circuit was redeployed), but also on the older functions and structures from which it originates (Anderson, 2008), or, put differently, it is theoretically salient on both sides.

The Case of Fingers in Numerical Cognition

The anecdotal case of the aforementioned exapted gaze heuristic (discussed in Section "The Exaptive Origins of Rationality")—both as functional shift and cooptation of a non-aptation—is relatively trivial, since the gaze heuristic by definition remains bounded to perception-action mechanisms. But what about the exaptive origins of higher cognitive processes involved in human rationality? In order to shed light on this point, consider the case of fingers in numerical cognition.

A crucial argument that is central in the 'rationality wars' (discussed in Section "Rationality and Adaptation") is related to the nature of numerical cognition, and in particular to frequency formats (Gigerenzer and Hoffrage, 1995). Gigerenzer (1996) and his colleagues have tested models that predict when frequency judgments are valid and when they are not. According to Gigerenzer, some cognitive biases—for example, the well-known conjunction fallacy (e.g., Tversky and Kahneman, 1983)—can be neutralized if probability information, in specific task environments, is given in absolute frequencies, not percentages (the so-called 'natural frequency hypothesis': Hertwig and Gigerenzer, 1999; see Amitani, 2015 for a discussion). What seems to be interesting in this 'frequency battle' is that both perspectives are quite silent on the nature of numerical cognition and, in particular, on the evidence that numerical processing is significantly embodied.

An important tradition of research over the last two decades highlights that numerical processing is constitutively dependent on the sensory-motor system: "Adults can be said to rely on an abstract representation of number if their behavior depends only on the size of the numbers involved, not on the specific [. . .] means of denoting them" (Dehaene et al., 1998, p. 356). Numerical processing depends on the surface format, as magnitudes are denoted by employing numeral systems (e.g.,

decimal system, binary system, graphical systems, etc.) and their respective notations (Arabic notation, Roman notation, etc.) (for an articulate view on this debate, see Cohen Kadosh and Walsh, 2009). For instance, 'four' can be expressed as 'OOOO,' '4' in the decimal numeral system, '100' in the binary numeral system, 'IV' in the Roman numeral system. What we call numbers are actually numerals, namely artifacts that humans manipulate in order to perform computations. As suggested by Lakoff and Núñez (2000, p. 86): "when we learn procedures for adding, subtracting, multiplying, and dividing, we are learning algorithms for manipulating symbols-numerals, not numbers." With a specific reference to heuristics, Mastrogiorgio and Petracca (2014) show that what selectively activates automatic or deliberate systems, in the well-known 'bat and ball' problem (Frederick, 2005), are the specific numerals involved in the task. This type of evidence corroborates the idea that heuristics processing of magnitudes is not neutral to the format of the task. Actually, the format of the task is part of the task and is precisely what enables a specific type of solving process (Mastrogiorgio, 2015).

The embodied, non-abstract view of numerical cognition represents an argument that is just as crucial to the rationality debate as it is overlooked. The perception-action system is not an accessory of mathematical cognition, but it is precisely the embodied substrate exapted for the emergence of higher-level faculties. The embodied dimension of numerical abilities, where numerical processing is grounded on the perception-action systems, represents a fundamental argument to consider such abilities as exaptations of the perception-action system. Indeed, logical and mathematical abilities—which are considered as pillars of rationality—are actually embodied and require neural reuse. Walsh (2003) highlights that the sensory-motor system provides a common metric for different types of magnitudes (space, time, and numbers). The idea is consistent with the hypothesis that the manipulation of sizes is embodied (see Bueti and Walsh, 2009; Ranzini et al., 2011), where reasoning processes are grounded on the perception-action system (Gallese and Lakoff, 2005). In particular, numerical processing involves gestures of hand (e.g., Chiou et al., 2012) and fingers (e.g., Sato et al., 2007).

More specifically, finger gnosis represents an alternative mechanism that is contrary to the general hypothesis that counting on fingers is the main mechanism on which numerical processing is grounded. *Finger gnosis*—which is the ability to distinguish which finger has been lightly touched without relying on the visual feedback—enables, *via* neural reuse, numerical processing. Indeed, finger gnosis represents the embodied register for storing the numbers to be manipulated. Such a finger register is co-opted for numerical processing, and potentially for all those functions that can exploit such type of biological structure (see Penner-Wilger and Anderson, 2008). Finger gnosis is a good predictor of children' mathematical performance, but the same cannot be said for the generalist idea of using fingers to count. As discussed by Anderson (2008, p. 432) "children with developmental coordination disorder (DCD) have poor finger agility, but most have preserved finger gnosis, and do not generally evidence significant mathematical difficulties" (see also Cermak and Larkin, 2001).

MORE THUMBS THAN RULES

Bounded rationality flourished in the *cognitivist paradigm*, according to which reasoning processes are conceived as computational rules to manipulate symbolic representations of the environment (cf. Newell and Simon, 1972). Accordingly, in Gigerenzer's ecological rationality human behavior can be described by a number of "cognitive heuristics" and "their building blocks (e.g., *rules* for search, stopping, decision)..." (Gigerenzer and Gaissmaier, 2011, p. 456 emphasis added). The cognitivist paradigms focus on abstract computation is further evident in the emphasis that is placed on humans as "intuitive statisticians" (Gigerenzer and Murray, 2015). That is, heuristics are said to be based on various computational and statistical techniques including statistical sampling, threshold and signal detection, just-noticeable-differences, and Bayesian inference (e.g., Gigerenzer and Hoffrage, 1995; Dhami et al., 2004; Karelaia and Hogarth, 2008; Hertwig and Pleskac, 2010; Luan et al., 2014; Gigerenzer, 2019).

However, what is missing in much of this cognitivist focus is that since the 90s, cognitive sciences have been subject to a radical renovation that challenges the assumption of the cognitivist paradigm, through the general hypothesis of a constitutive dependence of cognition on the traits of the human body. Far from being a unitary epistemic attempt, such renovation—known under the general label of *embodied cognition*—includes a pluralism of approaches, differing in their epistemic, theoretical and methodological dimensions (for an overview see Wilson, 2002; Calvo and Gomila, 2008; Clark, 2008; Kiverstein and Clark, 2009; Newen et al., 2018). The flourishing field of embodied cognition—which places a novel emphasis on the sensory-motor system as a constitutive component of cognitive processes—represents a fresh theoretical viewpoint for a reconceptualization of bounded and ecological rationality. This reconceptualization—so-called, *embodied rationality*—considers human rationality as an embodied phenomenon (Spellman and Schnall, 2009; Mastrogiorgio, 2011; Mastrogiorgio and Petracca, 2012, 2015, 2016; Gallagher, 2018; Gallese et al., 2021; Petracca, 2021). Embodied rationality, by endorsing an anti-Cognitivist stance, is critical toward the idea that human rationality is based on symbolic manipulations of a represented environment and, *de facto*, rejects the cognitivist pillar of a cognition implementable on artificial architectures (i.e., the so-called 'physical symbol system hypothesis,' Newell and Simon, 1976).

As we discussed in the previous Section "The Exaptive Origins of Rationality," the exaptation of perception-action systems represents a fundamental mechanism for the come-into-being of higher-level cognitive faculties involved in reasoning processes. Exaptive mechanisms are able to cast light on how the morphological traits of the human body are co-opted to give rise to cognitive mechanisms. Embodied rationality—by claiming an embodied view of cognition and by endorsing (as we propose here) exaptive evolutionary mechanisms—radically challenges the two pillars of evolutionary psychology and ecological rationality: computationalism (i.e., cognitivism) and adaptationism.

Putting Embodied Rationality Into the Evolutionary Psychology Debate

The idea that biased minds make better inferences is a central argument of ecological rationality (Gigerenzer and Brighton, 2009), antithetical to Kahneman's (2011) focus on cognitive biases, which are considered sources of systematic irrationality. Evolutionary psychology, in its foundational principles, is sympathetic with this argument as it views such 'biasedness' as a constitutive property of the mind, ascribed to a natural endowment. With reference to the frequency format (discussed in Section "The Case of Fingers in Numerical Cognition"), Tooby and Cosmides (2005, p. 23) endorse natural frequencies (i.e., absolute frequency), admitting that "Giving people probability information in the form of absolute frequencies—an ecologically valid format for hunter-gatherers—reveals the presence of mechanisms that generate sound Bayesian inferences." Cosmides and Tooby's (2013) emphasis on the environment of evolutionary adaptedness argues that the modules of the functional architecture of the mind were the product of selective pressure in hunter-gatherer environments but not in civilized ones. Generally speaking, a distinctive mark of evolutionary psychology lies in the general hypothesis that the psychological architecture consists of reasoning and learning processes that are *not* general-purpose, content-independent and somehow equipotential (Tooby and Cosmides, 1992; Pinker, 2002). Mind is not a *tabula rasa* (blank-slate) and organisms come "factory-equipped" with evolutionary endowments allowing specific reasoning and learning processes, which are salient in the respective environments.

We agree with such principles of evolutionary psychology to the extent that we here propose an embodied theory of human rationality that takes into account specific evolutionary mechanisms. Nevertheless, we think that adaptationism and computationalism—both central in evolutionary psychology and the associated literature on ecological rationality—are quite problematic for a theory of embodied rationality, which calls for an embodied view of cognition and asks us to carefully consider (as we propose here) non-adaptive evolutionary mechanisms. Moreover, we cannot avoid noticing that ecological rationality is still—in our opinion—far too anchored on the adaptationism and computationalism of evolutionary psychology, as it deliberately relies on the adaptive arguments of the cognitivist framework (where heuristics are computationally modeled as search and stopping rules for information processing). In the next sections, we propose five arguments that we deem constitutive of embodied rationality and that are dialectically critical against some of the foundational tenets of adaptationist approaches to evolutionary psychology.

Tinkering and Rationality

According to Tooby and Cosmides (2005, p. 16), "the brain was designed by natural selection to be a computer. Therefore, if you want to describe its operation in a way that captures its evolved function, you need to think of it as composed of programs that process information." (This strong focus on information processing readily carries into current work within ecological rationality as well – for example, see Gigerenzer, 2019.) It's important to point out that Tooby and Cosmides (2005) foundational claim—that the brain is computational and that it is composed of programs of information processing—is extremely provocative and strong, at least for endorsers of embodied cognition. Although they add that "its programs were designed not by an engineer, but by natural selection" (p. 16), we cannot but be puzzled by the juxtaposition of 'design' and 'natural selection,' also considering that the statement seems to denote a teleological perspective, where things are "designed . . . to be" and Pittendrigh's (1958) teleonomy/teleology distinction is not declaratively assumed.

A well-known alternative to unitary-design arguments (i.e., the design of a computer, in Cosmides and Tooby's words) is that of *tinkering* (Jacob, 1977), according to which the outcomes of evolution do not resemble perfect products of engineering but the ones of a tinkerer, "who does not know exactly what he is going to produce but uses whatever he finds around him" (p. 1163). As further put by Solé et al. (2002, p. 21), "evolution is limited by the constraints present at all levels of biological organization as well as by historical circumstances." Evolution does not somehow produce novelty from scratch but works on what already exists, as natural selection is strongly dependent on historical contingencies. With reference to the brain, Jacob (1977) adds: "Although our brain represents the main adaptive feature of our species, what it is adapted to is not clear at all" (p. 1166), arguing that the human brain is the product of evolutionary tinkering: "brain development in mammals was not as integrated process as, for instance, the transformation of a leg into a wing. The human brain was formed by superposition of new of new structures on old ones" (p. 1166). This idea critically departs from Cosmides and Tooby's view (though it is three decades antecedent), in the sense that the brain is not only far from being a perfect device, but also a layered structure where new structures of the neocortex were awkwardly superposed on the old ones through a tinkering process resembling the process of "adding a jet engine to an old horse cart" (p. 1166).

In this contribution, we are sympathetic with tinkering as we hypothesize that human rationality presents exaptive origins, where the ancestral sensory-motor system represents a structure that enables and constrains the emergence of specific reasoning processes, through neural reuse. From this perspective, rationality—far from being the apex of evolutionary processes—is essentially an accidental byproduct whose specificities are evolutionary constrained by contingency. Rationality plausibly resembles a "kluge" (see Marcus, 2008) and can be considered as an 'adjacent possible' (Kauffman, 2000), where the historical contingency defines the specificities for a (re)use of bodily structures.

Rationality, Out of the Vat

According to Tooby and Cosmides (2005, p. 17): "Individual behavior is generated by this evolved computer, in response to information that it extracts from the internal and external environment." The emphasis on information processing and the declared computationalism of Cosmides and Tooby is antithetical to the anti-cognitivist arguments of embodied cognition scholars.

Critical precursors of this tension can be found in a quite-known special issue (led by Herbert A. Simon and his colleague Alonso Vera on *Cognitive Science* in the early 90s) on the nature of situated cognition. Traditionally, situated cognition emphasizes that humans think on the fly—through an extemporaneous interaction with the environmental contingencies—, rather than storing and retrieving conceptual knowledge (e.g., Chiel and Beer, 1997; Clancey, 1997; Greeno, 1998). In this debate, Vera and Simon defended the compatibility of situated cognition with the 'physical symbol system hypothesis,' since "complex human behavior can be and has been described and simulated effectively in physical symbol systems" (Vera and Simon, 1993, p. 46).

A fundamental—still recent—counter-argument against Vera and Simon's defense lies in the consideration that the physical symbol system hypothesis projects first-person cognitive processes onto third-person computational rules able to model them (see Clancey, 1993). By doing this it conflates the possibility of emulating, through a computer, a number of cognitive processes with a nomological necessity. In the words of Greeno and Moore (1993, p. 56): "the question should not be whether a system that uses symbolic processes is sufficient, but whether the symbolic processes that are hypothesized are necessary."

We believe that this conflation of first- and third-person accounts is also the unwitting assumption of evolutionary psychology, where the independence between computational programs (composing a computer-like brain) and flesh-and-blood organisms represents a legitimation principle: separating computation from the body is precisely the theoretical argument that makes evolutionary psychology a domain of investigation independent from the biological realm and matters of morphology. And, interestingly, when "flesh and blood" are washed out, cross-level mechanisms (which are central in our speculation) also disappear.

Provocatively, if we endorse the view that such computational programs were sculpted by evolution, should we also assume that such programs admit artificial, out-of-the body, evolution? Can we implement such processes on a computer and simulate natural evolution through environmentally-calibrated evolutionary algorithms? We think that an embodied reconceptualization of rationality represents an alternative to the current views, which are still rooted in the cognitivist framework assumed in evolutionary psychology, where the body of the organism seems to be nothing more than hardware that is merely instrumental to the allegedly-computational processes occurring in the brain.

From Massive Modularity to Massive Redeployment

A fundamental tenet of evolutionary psychology is modularity, where the functional decomposition of biological systems—into functional sub-systems (e.g., organs) incrementally adapted for specific tasks—is extended to cognitive processes and endowments (Carruthers, 2006; see also Barrett and Kurzban, 2006). Innateness, along with Chomsky's thesis on the poverty of stimulus, represents an argument in favor of the selective adaptation of cognitive faculties, as a set of evolved mechanisms that instantiate human problem solving abilities, substituting the necessity of learning everything from scratch. Such an argument endorses a specialized view of the human brain: "Natural selection ensures that the brain is composed of many different special-purpose programs and not a domain general architecture" as suggested by Tooby and Cosmides (2005, p. 17), adding that "this is a ubiquitous engineering outcome. The existence of recurrent computational problems leads to functionally specialized application software."

A fundamental evolutionary counter-argument—central in Gould and Lewontin 's critique to adaptationism—is the rejection of a fixed modularity of organisms, where specialization is functionally defined. Gould and Lewontin strongly criticize the claim of adaptationism that "proceeds by breaking an organism into unitary 'traits' and proposing an adaptive story for each considered separately" (Gould and Lewontin, 1979, p. 581). Adaptations are generally referred to as structural (e.g., features of the human body), physiological (e.g., homeostatic mechanisms) or behavioral (e.g., inherited systems of behaviors) traits. Actually, the problem of the *unit of analysis* is crucial, since in Gould and Lewontin's perspective we cannot make ontological distinctions but just the ones that are instrumental to evolutionary contingencies. Indeed, we have precise modularity only if we assume a congruence between structure and function, which is the precise argument criticized by Gould and Lewontin, through the notion of spandrels. Furthermore, the absence of strong empirical evidence in favor of claims about modularity suggests a need for a far more pluralistic approach (e.g., Lloyd, 1999).

Modularity (see Prinz, 2006 for a critique) enters into the rationality debate where a problematic blank-state is substituted by innateness, calling into account Darwinian evolutionary mechanisms (Samuels, 1998; Samuels et al., 1999). Moreover, this modular view is precisely the one that connotes Gigerenzer (2008)'s adaptive toolbox (see Section "Inside the Adaptive Toolbox"): "The adaptive toolbox is a Darwinian-inspired theory that conceives the mind as a modular system that is composed of heuristics, their building blocks, and evolved capacities" (p. 20), where the building blocks are precisely the cognitivist rules for symbolic manipulation—rules for search, stopping and decision (Gigerenzer and Gaissmaier, 2011).

We think that the general claim of embodied cognition—and the hypothesis of a neural reuse (discussed in Section "Neural Reuse")—opens a quite different perspective on the evolution and nature of rationality, where neural substrates are horizontally layered instead of being vertically compartmentalized. This alternative to modularity, building on the Massive Redeployment Hypothesis, represents a significant argument for a radical updating of the current view of human rationality. Importantly, the Massive Redeployment Hypothesis does not represent a radical alternative to modularity in general terms but to such forms of modularity that, *stricto sensu*, assume domain-specificity (Anderson, 2007b).

Our view of embodied rationality, then, represents a radical alternative to the modular mind assumed by the adaptive toolbox of ecological rationality, where heuristics are domain-specific. Indeed, embodied rationality claims a horizontally layered mind—whose evolution is connoted by exaptive mechanisms of older structures—instead of an adaptive toolbox with specialized

modules. Interestingly, this framework might also shed new light on the nature of automatic cognitive systems (sources of cognitive errors, in Kahneman's view), which are plausibly more plastic and flexible than commonly assumed (see Bellini-Leite and Frankish, 2021). Its biasedness can be conceptualized as the instantiation of such evolutionary constraints from which specific reasoning processes originate.

Niche Construction

Evolutionary psychology emphasizes that cognitive programs (the computational rules composing the human brain) were adaptive in ancestral environments but they may not be adaptive in civilized environments (Tooby and Cosmides, 2005; Cosmides and Tooby, 2006; also see Stanovich, 2011), where the mismatch produces an *adaptive lag*. The hypothesis of the adaptive lag is plausible if we endorse a purely adaptationist view that encompasses the bottleneck of time, for the occurrence of incremental refinements. Under this hypothesis, the environment remains somehow fixed and untouched so as to identify a univocal causal direction of selective processes. As bluntly argued by Williams (1992, p. 484): "Adaptation is always asymmetrical; organisms adapt to their environment, never *vice versa*."

Transposing this argument to the rationality debate, and using the Simon's scissors argument, the cognitive blade adapts to the environmental blade. We think that this conception of a fixed and untouched environment in the rationality debate, might originate from Simon's original emphasis on human problem-solving (e.g., Newell and Simon, 1972), according to which the environment is considered a mere *task environment*, representing the experimental setting used to comparatively assess human reasoning abilities (cf. Gray et al., 2006). This experimentally-operationalized environment is problematic because over decades it has forced rationality scholars to unwittingly conflate a methodological expedient into a theoretical assumption (consistently with an adaptationist perspective): organisms are problem solvers, continuously facing survival problems, out there, administered by the environment. This view is endorsed by Tooby and Cosmides (2007, p. 43, emphasis added), who stated that natural selection "favors building special assumptions, innate content, and domain-specific *problem-solving strategies* into the proprietary logic of neural devices whenever this increases their power to solve *adaptive problems*." This conflation of a methodological expedient into a theoretical assumption is, we believe, the main cause of the marginalization of alternative evolutionary logics—in particular, niche construction—in the rationality debates.

Niche construction emphasizes the active role of the organism in manipulating the environment. In particular, the organism significantly modifies the environment thus affecting the selective processes, where such modification invites an evolutionary response of the organism (and/or other species) (Lewontin, 1983; Odling Smee et al., 2003). Niche construction represents a fundamental mechanism in the so-called 'extended evolutionary synthesis,' which, though retaining the fundamentals of evolutionary theory, also emphasizes the role of constructive processes in evolution and development (Laland et al., 2015)—meaning that the organism is not just the object but *the* subject of evolution. We are of course aware (and this is not the place for an extended discussion) that niche construction has been the subject of heated evolutionary debates (as in the case of exaptive processes) among proponents and critics (see Gupta et al., 2017a, and replies: Feldman et al., 2017; Gupta et al., 2017b).

The problem of the adaptive lag remains one of the points of contrast between niche construction and evolutionary psychology (Laland and Brown, 2006), a point that is also critical for the rationality debate. Indeed, evolutionary psychology remains vague on the nature of the misfit between the ancestral cognitive architecture and the civilized environment. As put by Tooby and Cosmides (2005, p. 17): "The industrial revolution—even the agricultural revolution—is too brief a period to have selected for complex new cognitive programs." This has led many to claim that existing cognitive faculties are more and more unsuited for modern decision environments, that "the modern world tends to create situations in which the default values of evolutionarily adapted cognitive systems are not optimal" (Chater et al., 2018, p. 812). This is also an implicit assumption of the extant biases literature and it's *de facto* claim of an "epidemic of human perceptual blindness, irrationality, and delusion" (Felin et al., 2019, p. 109). Ironically, we do not understand why we do not have a cognitive architecture adapted to civilization, but (somehow) we have cognitive structures that deliberately created such civilization. If we consider civilization a non-entropic process, a product of human deliberation (and we exclude that civilization is an accidental byproduct), why should we lack the cognitive endowments to deal with this ordered process?

Now, Cosmides and Tooby might be right that many of our cognitive endowments were adapted to the past and not to our civilized environments. However, we believe that a central matter for rationality is, precisely, what ancestral structures (and how) have been exapted to be reused to create civilized niches. From the perspective of economics (investigating advanced economic systems populated by more or less rational agents), the absence of an adaptive lag is somehow taken for granted and factual: it is of little value assuming that, say, the executive board of the European Central Bank is composed by individuals endowed with ancestral, hunter-gatherer-type cognitive architectures. Interestingly, the clash between task-relevant timescale and evolutionary timescale is central matter for neural reuse (Rathkopf, 2021). Thus we agree with Laland et al. (2007) in recommending a rejection of the adaptive-lag hypothesis "in favor of a niche-construction perspective, which focuses on how human beings respond, and are themselves responses, to self-induced environmental changes" (p. 63; also see Laland and Seed, 2021).

Cognitive arguments in favor of niche construction are abundant and central in externalist perspectives on embodied cognition (Sterelny, 2010), as the environment is *de facto* manipulated through its own artifacts so as to create the conditions which facilitate—extend and scaffold—human behavior and constrain its evolutionary paths. Cumulative technological culture continuously improves, evolves and innovates through the use of tools (e.g., Osiurak and Reynaud, 2020). The notion of 'intoolligence' introduces a unified framework for the cognitive study of

tool use and technology, based on the general idea that making and using a tool are two independent cognitive steps (Osiurak and Heinke, 2018). The existence of cross-level exaptive mechanisms, (discussed in the Section "Cross-Level Mechanisms: Exapting the Perception-Action System"), that also imply jumps from the natural to the cultural dimension, are relevant. An unconstrained view on the unit of selection reveals the possibility of considering the interplay between nature and culture in a more flexible manner. Notice that the notion of exaptation has also been increasingly used to explain the nature of technological innovation in the economic domain, where evolutionary processes apply directly to endosomatic endowments and tools (Dew et al., 2004; Cattani, 2006; Andriani and Cattani, 2016; Felin et al., 2016; Cattani and Malerba, 2021; Cattani and Mastrogiorgio, 2021).

In short, we think that niche construction-related argument represents a fundamental opportunity for innovation for the bounded and ecological rationality literatures (see Callebaut, 2007), by stressing the relativistic and culturally-embedded criteria of normativity (e.g., Elqayam, 2011).

The Environment, in Place of Rationality

A central principle of evolutionary psychology is that describing the evolved computational architecture of our brains also allows us to understand cultural and social phenomena (Tooby and Cosmides, 2005). The idea is that domain-specific programs are not passive but active devices in defining our experience so as to shape cultural practices.

We think that this principle, consistent with a computational mind whose modules were adaptively selected, is problematic. Indeed, we think that cognitive architectures alone are not sufficient for understanding cultural and social phenomena. In many niches—precisely the ones of the civilized world in which rationality is paramount—the opposite consideration is also valid: social and cultural phenomena help to understand the mind. And this is not just because the environment matters as it shapes domain-specific modules by adaptive selection, but because the environment can, in a sense, operate in place of cognition. That is, a relevant perspective in the pluralism of embodied cognition approaches (see Wilson, 2002) is the externalist one, according to which the environment extends and integrates cognition: organisms do not need to gather and process information internally to the extent that they offload cognition onto the environment (Sterelny, 2010). Therefore, the external environment can be functionally equivalent to internal cognitive processes, for instance when we use calendars as external memory tools (e.g., Clark and Chalmers, 1998). And, it can also be complementary as it functionally integrates cognition, for instance when we use a pen and paper to facilitate mathematical reasoning (e.g., Menary, 2010). Moreover, the mind can be socially extended to encompass the social, institutional, and cultural dimensions (Gallagher, 2013).

Furthermore, acknowledging "others" as part of the environment helps us also to consider the role of social cognition and social learning in the development of embodied rationality. Mirror neurons are a crucial substrate in many developmental processes based on imitation (Rizzolatti et al., 2002). While many scholars see mirror neurons as a genetic adaptation for understanding action (Rizzolatti et al., 1996; Cook et al., 2014), others consider their ontogeny, specifically hypothesizing that mirror neurons are related to learning process (Giudice et al., 2009, see also Tramacere et al., 2017). Importantly, the reuse of the perception-action system is not limited to proprioception, but also involves interoception in the domain of affectivity. Higher forms of empathy, in particular mentalizing, are hypothesized to be linked to perception and motor system (specifically associated with mouth/face actions and expressions), subject to a process of exaptation during primate phylogeny (Tramacere and Ferrari, 2016). Generally speaking, social intelligence can be considered an adaptive response to the complexity of the social environment. Specifically, the different views of such complexity represent different versions of the social intelligence hypothesis (Jolly, 1966; Humphrey, 1976). In this debate, we endorse a pluralistic view on evolution where exaptive processes, also due to the unconstrained nature of the unit of selection, allow "jumps" from the genetic to the cultural domain. Hominin evolution can be seen as a response to selective environments that other hominins previously created, consistent with a niche constructionist perspective (cf. Sterelny, 2007).

With specific reference to rationality, understanding the role of social systems as external and distributed cognitive devices is crucial. The development of rationality in humans is enacted in a series of socio-cultural experiences, as the environment is significantly instantiated in interactions with others (cf. Gallagher, 2018). The enactivist approach, in particular, emphasizes the extended, intersubjective and socially situated nature of cognitive systems. From this perspective, the brain does not create an internal model of the world, but is conceptualized as a part of the larger system of brain-body-environment (for an overview see Newen et al., 2018). The existence of so-called minimal/zero intelligence agents operating in complex economic environments represents a persuasive argument against the internalist perspective (assumed in evolutionary psychology). Simple agents making elementary choices (e.g., random choices) can generate outcomes that are substantively rational (Gode and Sunder, 1993). Zero/minimal intelligence occurs precisely because some external structures take the place of an agent's cognition. The case of embodied swarm intelligence is, in this regard, representative: elementary agents (e.g., ants) interacting at the micro-level through simple rules generate self-organized and complex societies. And such societies, in which tasks and roles are differentiated, cannot be reduced to the underlying interacting rules, being rather emergent phenomena (e.g., describing the evolved-though-rudimentary computational architecture of ants' brains tells us little about how ants' societies are organized).

Modernity is populated by devices, and whole cultural systems, that work in place of individuals' cognition. Indeed, zero/minimal intelligence could also be conceptualized in terms of an institutional perspective (Hodgson, 2004; Felin, 2015; Petracca and Gallagher, 2020), where economic institutions work in place of internal cognitive processes, doing most of the cognitive job by scaffolding agents' decisions and actions (Clark, 1997). Externalizing the cognitive burden is not just a

matter of sharing mental models among the economic agents as suggested by Denzau and North (1994), but calls into account the whole existence of so-called cognitive institutions (Gallagher and Crisafi, 2009; Petracca and Gallagher, 2020). Interestingly, markets—composed by a collectivity of (more or less) rational agents—are precisely such cultural devices that offer an ostensive counter-argument to the evolutionary psychologists' claims about a pure internalist perspective. Markets can be considered as socially extended institutions able to solve collective allocation problems unsolvable by agents with an internalized and disembodied rationality. Markets are social institutions that emerge from intersubjective "embodied" interactions, producing a "cognitive economy" as they reduce individual cognitive effort in making decisions and enable specific types of economic reasoning processes (for a discussion see Gallagher et al., 2019).

CONCLUSION

According to the philosopher Wittgenstein (1953) the rules of language are analogous to the rules of games, where words are ambiguous and have a meaning only within the specific linguistic game being played: words do not point to ontologically fixed entities but they are merely instrumental to contingent interaction. In this contribution, 'thumbs' are such an ambiguous word, contingent upon the narrative. Thumbs are exemplifications of what we mean by heuristics (as rules of "thumbs"), they are effects of exaptive processes (as in the case of a small bone of the wrist, exapted to became the panda's opposable thumb), and they are enabling constraints (as in the case of finger gnosis that enables numerical processing).

Our use of 'thumb' therefore is not just a rhetorically expedient device for scientific communication but uncovers an evolutionary matter central to our arguments. Exaptive mechanisms require us to disentangle structure and function to the extent that their interplay cannot be pre-stated, but rather represents precisely a point which evolutionary forces apply to. As in the case of the feathers of birds (first evolved for thermal regulation and then co-opted for flight), functional ambiguity is probably a property of evolution that is able to (re)attribute contingent meanings to biological structures. And this process relativizes the possibility of framing evolutionary units in once-for-all, ontologically-defined categories (Kauffman, 2000, 2008; Longo et al., 2012; Roli and Kauffman, 2020). And at a high level of abstraction, the whole process of evolution toward rationality might be seen as adjacent possibility, emergent—contingently, through reuse—from the embodied endowments of the organism co-evolved with a continually changing environment.

In this contribution, we propose an embodied reconceptualization of rationality—*embodied rationality*—based on the reuse of the perception-action system. We argue that neural processes involved in the control of the sensory-motor system, salient in ancestral environments, can be co-opted to create (by tinkering) high-level cognitive faculties, employed in civilized niches. The idea that 'rationality is an exaptation' is actually an exemplification, as rationality, in our view, is not a unitary system but is made of a stratified mix of many exaptations of the sensory-motor system on which higher cognitive processes are grounded (the exaptation of finger gnosis registers, discussed in Section "The Case of Fingers in Numerical Cognition" is only one of them. Notice that we use the expression 'sensory-motor system' in the singular form as an exemplification, but we actually assume that there is a multitude of exapted systems). To revisit Gould and Lewontin's architectural, "spandrels" metaphor, the cognitive architecture of embodied rationality does not resemble a futuristic and optimized building, as much of evolutionary psychology suggests. Rather, cognition resembles a harmonious, pleasant and effective jumble of stratified architectural structures (similar to Tuscan farmhouses), the products of tinkering. In this stratified architecture, the many spandrels and byproducts matter a great deal in defining the overall result and the adjacent evolutionary possibilities.

But, considering exaptive mechanism a *tout court* alternative to adaptive logics is wrong. Exaptive processes do not rule out adaptive mechanisms—they do not imply a paradigm-shift, *stricto sensu*, on evolution—rather they extend the richness of evolutionary possibilities. Thus, embodied rationality does not represent an alternative to bounded and ecological rationality but precisely an integration, aiming to challenge and update the underlying adaptationist and cognitivist assumptions. However, the innovativeness of our proposal does not rely on a *tout-court* redefinition of rationality, and *a fortiori*, does not imply to identify a novel and different type of heuristics. Nothing changes in the performative dimension of rationality.

The relevant innovation that *embodied rationality* brings to the debate is related to the novel role of the human body, and the attendant notions of bodily, cognitive and material reuse. While bounded and ecological rationality (flourished in the cognitivist paradigm) prescind from the biological endowments, embodied rationality emphasizes the constitutive dependence of heuristics on the human body and in particular on the sensory-motor system. Actually, bounded and ecological rationality do not rule out embodiment in absolute terms: on the one side a number of heuristics of the adaptive toolbox (such as the gaze heuristic, e.g., Raab and Gigerenzer, 2005) require the use of the sensory-motor system, on the other side, gut feelings are determinant for heuristic choice (e.g., Gigerenzer, 2007). Despite such embodied arguments, heuristics in bounded and ecological rationality remain disembodied in the sense that human body plays the role of a neutral hardware on which the cognitive software runs.

That is to say, heuristics are seen as algorithmic rules for information processing, which could hypothetically run on various types of bodily hardware (cf. Simon, 1990). Embodied rationality, on the contrary, claims the non-neutrality of biological endowment for the specification of the cognitive processes, and this argument represents a distinctive mark of embodied rationality, which cannot be found in ecological and bounded rationality.

Hence, embodied rationality invites us to abandon a third-person rationality (where cognitive processes can be expressed as objectified, algorithmic rules for information processing) and calls into account the biological realm. That is, high-level

cognitive processes can be understood precisely as they are grounded on the sensory-motor system, and not prescinding from it, where such grounding can be considered the pivot of Simon's scissor. Such grounding allows us to account for the origins of heuristics. While bounded and ecological rationality have offered us different types of heuristics, they are not able to explain how heuristics came into being. Embodied rationality can be useful to ascribe the origins of heuristics to specific evolutionary constraints that specify the adjacent possible for cognition—which cognitive processes are "affordable" by neural reuse. Investigating the ontogenetic and phylogenetic dimensions, along with task-relevant or evolutionary timescales of neural reuse, represent a future domain of investigation for embodied rationality.

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

AM: conceptualization, supervision, writing – original draft, and writing – reviewing and editing. TF: conceptualization and writing – reviewing and editing. SK and MM: conceptualization and reviewing. All authors contributed to the article and approved the submitted version.

REFERENCES

Amitani, Y. (2015). The natural frequency hypothesis and evolutionary arguments. *Mind Soc.* 14, 1–19. doi: 10.1007/s11299-014-0155-7

Anderson, M. L. (2006). Evidence for massive redeployment of brain areas in cognitive function. *Proc. Cogn. Sci. Soc.* 28, 24–29.

Anderson, M. L. (2007a). Massive redeployment, exaptation, and the functional integration of cognitive operations. *Synthese* 159, 329–345. doi: 10.1007/s11229-007-9233-2

Anderson, M. L. (2007b). The massive redeployment hypothesis and the functional topography of the brain. *Philos. Psychol.* 20, 143–174. doi: 10.1080/09515080701197163

Anderson, M. L. (2008). "On the grounds of (X)-grounded cognition," in *Handbook of Cognitive Science: An Embodied Approach*, eds P. Calvo and T. Gomila (Amsterdam: Elsevier), 423–435. doi: 10.1016/B978-0-08-046616-3.00021-9

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266. doi: 10.1017/S0140525X10000853

Anderson, M. L. (2014). *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/10111.001.0001

Anderson, M. L. (2016). Précis of after phrenology: neural reuse and the interactive brain. *Behav. Brain Sci.* 39:e120. doi: 10.1017/S0140525X15000631

Anderson, M. L., Kinnison, J., and Pessoa, L. (2013). Describing functional diversity of brain regions and brain networks. *Neuroimage* 73, 50–58. doi: 10.1016/j.neuroimage.2013.01.071

Andrews, P. W., Gangestad, S. W., and Matthews, D. (2002). Adaptationism-how to carry out an exaptationist program. *Behav. Brain Sci.* 25, 489–504. doi: 10.1017/S0140525X02000092

Andriani, P., and Cattani, G. (2016). Exaptation as source of creativity, innovation, and diversity: introduction to the special section. *Indust. Corp. Change* 25, 115–131. doi: 10.1093/icc/dtv053

Barack, D. L. (2017). Cognitive recycling. *Br. J. Philos. Sci.* 70, 239–268. doi: 10.1093/bjps/axx024

Barrett, H. C., and Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychol. Rev.* 113, 628–647. doi: 10.1037/0033-295X.113.3.628

Bellini-Leite, S. C., and Frankish, K. (2021). "Bounded rationality and dual systems," in *Handbook of Bounded Rationality*, ed. R. Viale (London: Routledge), 207–216. doi: 10.4324/9781315658353-13

Bueti, D., and Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1831–1840. doi: 10.1098/rstb.2009.0028

Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L., and Wakefield, J. C. (1998). Adaptations, exaptations, and spandrels. *Am. Psychol.* 53:533. doi: 10.1037/0003-066X.53.5.533

Callebaut, W. (2007). Simon's silent revolution. *Biol. Theory* 2, 76–86. doi: 10.1162/biot.2007.2.1.76

Calvo, P., and Gomila, T. (eds) (2008). *Handbook of Cognitive Science: An Embodied Approach*. Amsterdam: Elsevier.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychol. Rev.* 67, 380–400. doi: 10.1037/h0040373

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford: Clarendon Press. doi: 10.1093/acprof:oso/9780199207077.001.0001

Cattani, G. (2006). Technological pre-adaptation, speciation, and emergence of new technologies: how corning invented and developed fiber optics. *Ind. Corp. Change* 15, 285–318. doi: 10.1093/icc/dtj016

Cattani, G., and Malerba, F. (2021). Evolutionary approaches to innovation, the firm, and the dynamics of industries. *Strategy Sci.* 6, 265–289. doi: 10.1287/stsc.2021.0141

Cattani, G., and Mastrogiorgio, M. (eds) (2021). *New Developments in Evolutionary Innovation: Novelty Creation in a Serendipitous Economy*. Oxford: Oxford University Press. doi: 10.1093/oso/9780198837091.001.0001

Cermak, S. A., and Larkin, D. (2001). *Developmental Coordination Disorder*. Albany, NY: Delmar.

Chapman, P. D., Bradley, S. P., Haught, E. J., Riggs, K. E., Haffar, M. M., Daly, K. C., et al. (2017). Co-option of a motor-to-sensory histaminergic circuit correlates with insect flight biomechanics. *Proc. R. Soc. B Biol. Sci.* 284:20170339. doi: 10.1098/rspb.2017.0339

Chater, N., Felin, T., Funder, D. C., Gigerenzer, G., Koenderink, J. J., Krueger, J. I., et al. (2018). Mind, rationality, and cognition: an interdisciplinary debate. *Psychonom. Bull. Rev.* 25, 793–826. doi: 10.3758/s13423-017-1333-5

Chiel, H., and Beer, R. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body, and environment. *Trends Neurosci.* 20, 553–557. doi: 10.1016/S0166-2236(97)01149-1

Chiou, R. Y. C., Wu, D. H., Tzeng, O. J. L., Hung, D. L., and Chang, E. C. (2012). Relative size of numerical magnitude induces a size-contrast effect on the grip scaling of reach-to-grasp movements. *Cortex* 48, 1043–1051. doi: 10.1016/j.cortex.2011.08.001

Clancey, W. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge, MA: Cambridge University Press.

Clancey, W. J. (1993). Situated Action: a neuropsychological interpretation (Response to Vera and Simon). *Cogn. Sci.* 17, 117–133. doi: 10.1207/s15516709cog1701_7

Clark, A. (1997). "Economic reason: the interplay of individual learning and external structure," in *The Frontiers of the New Institutional Economics*, eds J. Drobak and J. Nye (San Diego, CA: Academic Press), 269–290.

Clark, A. (2008). Pressing the flesh: a tension in the study of the embodied, embedded mind? *Philos. Phenomenol. Res.* 76, 36–59. doi: 10.1111/j.1933-1592.2007.00114.x

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19. doi: 10.1093/analys/58.1.7

Cohen Kadosh, R., and Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behav. Brain Sci.* 32, 313–373. doi: 10.1017/S0140525X09990938

Cook, R., Bird, G., Catmur, C., Press, C., and Heyes, C. (2014). Mirror neurons: from origin to function. *Behav. Brain Sci.* 37, 177–192. doi: 10.1017/S0140525X13000903

Cosmides, L., and Tooby, J. (2006). *Evolutionary Psychology, Moral Heuristics, and the Law*. Oxford: Dahlem University Press. doi: 10.1002/0470018860.s00529

Cosmides, L., and Tooby, J. (2013). Evolutionary psychology: new perspectives on cognition and motivation. *Ann. Rev. Psychol.* 64, 201–229. doi: 10.1146/annurev.psych.121208.131628

Darwin, C. (1872). *The Origin of Species*. London: John Murray. doi: 10.5962/bhl.title.28875

Dawkins, R. (1983). "Universal darwinism," in *Evolution from Molecules to Man*, ed. D. S. Bendall (Cambridge: Cambridge University Press).

Dehaene, S., and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron* 56, 384–398. doi: 10.1016/j.neuron.2007.10.004

Dehaene, S., Dehaene-Lambertz, G., and Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* 21, 355–361. doi: 10.1016/S0166-2236(98)01263-6

Dennett, D. (1997). Darwinian fundamentalism: an exchange. *N. Y. Rev. Books* 44, 13.

Denzau, A. T., and North, D. C. (1994). Shared mental models: Ideologies and institutions. *Kyklos* 47, 3–31. doi: 10.1111/j.1467-6435.1994.tb02246.x

Dew, N., Sarasvathy, S. D., and Venkataraman, S. (2004). The economic implications of exaptation. *J. Evol. Econ.* 14, 69–84. doi: 10.1007/s00191-003-0180-x

Dhami, M. K., Hertwig, R., and Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychol. Bull.* 130, 959–988. doi: 10.1037/0033-2909.130.6.959

Eldredge, N., and Gould, S. J. (1972). "Punctuated equilibria: an alternative to phyletic gradualism," in *Models in Paleobiology*, ed. T. J. M. Schopf (San Francisco, CA: Freeman, Cooper and Co), 82–115. doi: 10.5531/sd.paleo.7

Elqayam, S. (2011). "Grounded rationality: a relativist framework for normative rationality," in *The Science of Reason: A Festschrift in Honour of Jonathan St.B.T. Evans*, eds K. I. Manktelow, D. E. Over, and S. Elqayam (Hove: Psychology Press).

Favela, L. H. (2021). "Fundamental theories in neuroscience: why neural darwinism encompasses neural reuse," in *Neural Mechanisms. Studies in Brain and Mind*, Vol. 17, eds F. Calzavarini and M. Viola (Cham: Springer), 143–162. doi: 10.1007/978-3-030-54092-0_7

Feldman, M. W., Odling-Smee, J., and Laland, K. N. (2017). Why Gupta et al.'s critique of niche construction theory is off target. *J. Genet.* 96, 505–508. doi: 10.1007/s12041-017-0797-4

Felin, T. (2015). A forum on minds and institutions. *J. Instit. Econ.* 11, 523–534. doi: 10.1017/S1744137415000144

Felin, T., Felin, M., Krueger, J. I., and Koenderink, J. (2019). On surprise-hacking. *Perception* 48, 109–114. doi: 10.1177/0301006618822217

Felin, T., Kauffman, S., Mastrogiorgio, A., and Mastrogiorgio, M. (2016). Factor markets, actors, and affordances. *Indus. Corp. Chang.* 25, 133–147. doi: 10.1093/icc/dtv049

Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732

Gallagher, S. (2013). The socially extended mind. *Cogn. Syst. Res.* 25, 4–12. doi: 10.1016/j.cogsys.2013.03.008

Gallagher, S. (2018). "Embodied rationality," in *The Mystery of Rationality*, eds G. Bronner and F. Di Iorio (Cham: Springer), 83–94. doi: 10.1007/978-3-319-94028-1_7

Gallagher, S., and Crisafi, A. (2009). Mental institutions. *Topoi* 28, 45–51. doi: 10.1007/s11245-008-9045-0

Gallagher, S., Mastrogiorgio, A., and Petracca, E. (2019). Economic reasoning and interaction in socially extended market institutions. *Front. Psychol.* 10:1856. doi: 10.3389/fpsyg.2019.01856

Gallese, V. (2008). Mirror neurons and the social nature of language: the neural exploitation hypothesis. *Soc. Neurosci.* 3, 317–333. doi: 10.1080/17470910701563608

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensorimotor system in reason and language. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Gallese, V., Mastrogiorgio, A., Petracca, E., and Viale, R. (2021). "Embodied bounded rationality," in *Handbook of Bounded Rationality*, ed. R. Viale (London: Routledge). doi: 10.4324/9781315658353-26

Gigerenzer, G. (1991). "How to make cognitive illusions disappear: beyond "heuristics and biases," in *European Review of Social Psychology*, Vol. 2, eds W. Stroebe and M. Hewstone (Chichester: Wiley), 83–115. doi: 10.1080/14792779143000033

Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to Kahneman and Tversky. *Psychol. Rev.* 103, 592–596. doi: 10.1037/0033-295X.103.3.592

Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious*. London: Penguin.

Gigerenzer, G. (2008). Why heuristics work. *Perspect. Psychol. Sci.* 3, 20–29. doi: 10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G. (2019). How to explain behavior? *Top. Cogn. Sci.* 12, 1363–1381. doi: 10.1111/tops.12480

Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.* 1, 107–143. doi: 10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Ann. Rev. Psychol.* 62, 451–482. doi: 10.1146/annurev-psych-120709-145346

Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–669. doi: 10.1037/0033-295X.103.4.650

Gigerenzer, G., and Gray, W. D. (2017). A simple heuristic successfully used by humans, animals, and machines: the story of the RAF and Luftwaffe, hawks and ducks, dogs and frisbees, baseball outfielders and sidewinder missiles-oh my! *Top. Cogn. Sci.* 9, 260–263. doi: 10.1111/tops.12269

Gigerenzer, G., and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychol. Rev.* 102, 684–704. doi: 10.1037/0033-295X.102.4.684

Gigerenzer, G., and Murray, D. J. (1987). *Cognition as Intuitive Statistics*. Hillsdale, NJ: Erlbaum.

Gigerenzer, G., and Murray, D. J. (2015). *Cognition as Intuitive Statistics*. Hove: Psychology Press. doi: 10.4324/9781315668796

Gigerenzer, G., and Selten, R. (eds) (2002). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT press. doi: 10.7551/mitpress/1654.001.0001

Giudice, M. D., Manera, V., and Keysers, C. (2009). Programmed to learn? The ontogeny of mirror neurons. *Dev. Sci.* 12, 350–363. doi: 10.1111/j.1467-7687.2008.00783.x

Gode, D. K., and Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *J. Polit. Econ.* 101, 119–137. doi: 10.1086/261868

Gould, S. J. (1982). Darwinism and the expansion of evolutionary theory. *Science* 216, 380–387. doi: 10.1126/science.7041256

Gould, S. J. (1991). Exaptation: a crucial tool for an evolutionary psychology. *J. Soc. Iss.* 47, 43–65. doi: 10.1111/j.1540-4560.1991.tb01822.x

Gould, S. J. (1997). Darwinian fundamentalism. *N. Y. Rev. Books* 44, 34–37.

Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Cambridge, MA: Belknap-Harvard. doi: 10.2307/j.ctvjsf433

Gould, S. J., and Eldredge, N. (1977). Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3, 115–151.

Gould, S. J., and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lon. Ser. B. Biol. Sci.* 205, 581–598. doi: 10.1098/rspb.1979.0086

Gould, S. J., and Vrba, E. (1982). Exaptation - a missing term in the science of form. *Paleobiology* 8, 4–15. doi: 10.1017/S0094837300004310

Gray, W. D., Neth, H., and Schoelles, M. J. (2006). "The functional task environment," in *Attention: From Theory to Practice*, eds A. F. Kramer, D. A. Wiegmann, and A. Kirlik (Oxford: Oxford University Press), 100–118. doi: 10.1093/acprof:oso/9780195305722.003.0008

Greeno, J. (1998). The situativity of knowing, learning, and research. *Am. Psychol.* 53, 5–26. doi: 10.1037/0003-066X.53.1.5

Greeno, J. C., and Moore, J. L. (1993). Situativity and symbols: response to vera and simon. *Cogn. Sci.* 17, 49–59. doi: 10.1207/s15516709cog1701_3

Gupta, M., Prasad, N. G., Dey, S., Joshi, A., and Vidya, T. N. C. (2017a). Niche construction in evolutionary theory: the construction of an academic niche? *J. Genet.* 96, 491–504. doi: 10.1007/s12041-017-0787-6

Gupta, M., Prasad, N. G., Dey, S., Joshi, A., and Vidya, T. N. C. (2017b). Feldman et al. do protest too much, we think. *J. Genet.* 96, 509–511. doi: 10.1007/s12041-017-0796-5

Hamlin, R. P. (2017). The gaze heuristic:" biography of an adaptively rational decision process. *Top. Cogn. Sci.* 9, 264–288. doi: 10.1111/tops.12253

Hertwig, R., and Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: how intelligent inferences look like reasoning errors. *J. Behav. Decis. Making* 12, 275–305.

Hertwig, R., and Pleskac, T. J. (2010). Decisions from experience: why small samples? *Cognition* 115, 225–237. doi: 10.1016/j.cognition.2009.12.009

Hodgson, G. M. (2004). *The Evolution of Institutional Economics*. London: Routledge. doi: 10.4324/9780203300350

Hodgson, G. M. (2005). Generalizing Darwinism to social evolution: some early attempts". *J. Econ. Iss.* 39, 899–914. doi: 10.1080/00213624.2005.11506859

Höfer, S., Raisch, J., Toussaint, M., and Brock, O. (2018). No free lunch in ball catching: a comparison of Cartesian and angular representations for control. *PLoS One* 13:e0197803. doi: 10.1371/journal.pone.0197803

Humphrey, N. (1976). "The social function of intellect," in *Growing Points in Ethology*, eds P. P. G. Bateson and R. A. Hinde (Cambridge: Cambridge University Press), 303–317.

Hutchinson, J. M., and Gigerenzer, G. (2005). Simple heuristics and rules of thumb: where psychologists and behavioural biologists might meet. *Behav. Process.* 69, 97–124. doi: 10.1016/j.beproc.2005.02.019

Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134

Jolly, A. (1966). Lemur social behavior and primate intelligence. *Science* 153, 501–506. doi: 10.1126/science.153.3735.501

Kaaronen, R. O. (2020). Mycological rationality: heuristics, perception and decision-making in mushroom foraging. *Judg. Dec. Mak.* 15, 630–647. doi: 10.31234/osf.io/7g8er

Kahneman, D. (2011). *Thinking, Fast and Slow*. Basingstoke: Macmillan.

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511809477

Kahneman, D., and Tversky, A. (1996). On the reality of cognitive illusions: a reply to Gigerenzer's critique. *Psychol. Rev.* 103, 582–591. doi: 10.1037/0033-295X.103.3.582

Kalant, H., Pinker, S., and Kalow, W. (1997). Evolutionary psychology: an exchange. *Exchange* 44, 55–58.

Karelaia, N., and Hogarth, R. M. (2008). Determinants of linear judgment: a meta-analysis of lens model studies. *Psychol. Bull.* 134, 404–426. doi: 10.1037/0033-2909.134.3.404

Kauffman, S. A. (2000). *Investigations*. Oxford: Oxford University Press.

Kauffman, S. A. (2008). *Reinventing the Sacred: A New View of Science, Reason, and Religion*. New York, NY: Basic Books.

Kiverstein, J., and Clark, A. (2009). Introduction: mind embodied, embedded, enacted: one church or many? *Topoi* 28, 1–7. doi: 10.1007/s11245-008-9041-4

Lakoff, G., and Núñez, R. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York, NY: Basic Books.

Laland, K., and Seed, A. (2021). Understanding human cognitive uniqueness. *Annu. Rev. Psychol.* 72, 689–716. doi: 10.1146/annurev-psych-062220-051256

Laland, K. N., and Brown, G. R. (2006). Niche construction, human behavior, and the adaptive-lag hypothesis. *Evol. Anthropol.* 15, 95–104. doi: 10.1002/evan.20093

Laland, K. N., Kendal, J. R., and Brown, G. R. (2007). The niche construction perspective: Implications for evolution and human behaviour. *J. Evol. Psychol.* 5, 51–66. doi: 10.1556/JEP.2007.1003

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., et al. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. *Proc. R. Soc. B Biol. Sci.* 282:20151019. doi: 10.1098/rspb.2015.1019

Lewontin, R. C. (1970). The units of selection. *Ann. Rev. Ecol. Syst.* 1, 1–18. doi: 10.1146/annurev.es.01.110170.000245

Lewontin, R. C. (1983). "Gene, organism and environment," in *Evolution from Molecules to Men*, ed. D. S. Bendall (Cambridge: Cambridge University Press).

Lloyd, E. A. (1999). Evolutionary psychology: the burdens of proof. *Biol. Philos.* 14, 211–233. doi: 10.1023/A:1006638501739

Lloyd, E. A., and Gould, S. J. (2017). Exaptation revisited: changes imposed by evolutionary psychologists and behavioral biologists. *Biol. Theory* 12, 50–65. doi: 10.1007/s13752-016-0258-y

Longo, G., Montévil, M., and Kauffman, S. (2012). "No entailing laws, but enablement in the evolution of the biosphere," in *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation* (New York, NY), 1379–1392. doi: 10.1145/2330784.2330946

Luan, S., Schooler, L. J., and Gigerenzer, G. (2014). From perception to preference and on to inference: an approach–avoidance analysis of thresholds. *Psychol. Rev.* 121, 501–525. doi: 10.1037/a0037025

Marcus, G. F. (2008). *Kluge: The Haphazard Construction of the Human Mind*. Boston, MA: Houghton Mifflin.

Marewski, J. N., and Gigerenzer, G. (2012). Heuristic decision making in medicine. *Dial. Clin. Neurosci.* 14:77. doi: 10.31887/DCNS.2012.14.1/jmarewski

Mastrogiorgio, A. (2011). "The embodied dimension of rationality: a hypothesis," in *Proceedings of the IAREP/SABE/ICABEEP Conference 2011* (England: University of Exeter),

Mastrogiorgio, A. (2015). Commentary: cognitive reflection versus calculation in decision making. *Front. Psychol.* 6:936. doi: 10.3389/fpsyg.2015.00532

Mastrogiorgio, A., and Petracca, E. (2012). "Setting the ground for a theory of embodied rationality," in *Proceedings of the IAREP*, Vol. 2012, eds A. Gasiorowska and T. Zaleskiewicz (Wrocklaw), 201.

Mastrogiorgio, A., and Petracca, E. (2014). Numerals as triggers of system 1 and system 2 in the 'bat and ball' problem. *Mind Soc.* 13, 135–148. doi: 10.1007/s11299-014-0138-8

Mastrogiorgio, A., and Petracca, E. (2015). Razionalità incarnata. *Sist. Intell.* 27, 481–504.

Mastrogiorgio, A., and Petracca, E. (2016). "Embodying rationality," in *Model-Based Reasoning in Science and Technology*, eds L. Magnani and C. Casadio (Cham: Springer), 219–237. doi: 10.1007/978-3-319-38983-7_12

Menary, R. (2010). "Cognitive integration and the extended mind," in *The Extended Mind*, ed. R. Menary (Cambridge, MA: The MIT Press). doi: 10.7551/mitpress/9780262014038.001.0001

Navarrete, G., and Santamaría, C. (2011). Ecological rationality and evolution: the mind really works that way? *Front. Psychol.* 2:251. doi: 10.3389/fpsyg.2011.00251

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

Newell, A., and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. ACM* 19, 113–126. doi: 10.1145/360018.360022

Newen, A., De Bruin, L., and Gallagher, S. (2018). *The Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780198735410.001.0001

Odling Smee, J., Laland, K., and Feldman, M. (2003). *Niche Construction: The Neglected Process in Evolution*. Princeton, NJ: Princeton University Press, 488.

Osiurak, F., and Heinke, D. (2018). Looking for intoolligence: a unified framework for the cognitive study of human tool use and technology. *Am. Psychol.* 73:169. doi: 10.1037/amp0000162

Osiurak, F., and Reynaud, E. (2020). The elephant in the room: what matters cognitively in cumulative technological culture. *Behav. Brain Sci.* 43:e156. doi: 10.1017/S0140525X19003236

Parkinson, C., and Wheatley, T. (2015). The repurposed social brain. *Trends Cogn. Sci.* 19, 133–141. doi: 10.1016/j.tics.2015.01.003

Penner-Wilger, M., and Anderson, M. L. (2008). "An alternative view of the relation between finger gnosis and math ability: Redeployment of finger representations for the representation of number," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, eds V. Sloutsky, B. Love, and K. McRae (Austin, TX: Cognitive Science Society).

Petracca, E. (2021). Embodying bounded rationality: from embodied bounded rationality to embodied rationality. *Front. Psychol.* 12:710607. doi: 10.3389/fpsyg.2021.710607

Petracca, E., and Gallagher, S. (2020). Economic cognitive institutions. *J. Inst. Econ.* 16, 747–765. doi: 10.1017/S1744137420000144

Pievani, T., and Serrelli, E. (2011). Exaptation in human evolution: how to test adaptive vs exaptive evolutionary hypotheses. *J. Anthropol. Sci.* 89, 9–23.

Pigliucci, M., and Müller, G. B. (eds) (2010). *Evolution - the Extended Synthesis*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/9780262513678.001.0001

Pinker, S. (2002). *The Blank Slate*. New York, NY: Viking Press.
Pittendrigh, C. S. (1958). "Adaptation, natural selection, and behavior," in *Behavior and Evolution*, eds A. Roe and G. G. Simpson (New Haven, CT: Yale University Press), 390–416.
Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* 10, 59–63. doi: 10.1016/j.tics.2005.12.004
Prinz, J. J. (2006). Is the mind really modular. *Contemp. Debates Cogn. Sci.* 14, 22–36.
Raab, M., and Gigerenzer, G. (2005). "Intelligence as smart heuristics," in *Cognition and Intelligence*, eds R. J. Sternberg and J. E. Pretz (Cambridge: Cambridge University Press), 188–207.
Ranzini, M., Lugli, L., Anelli, F., Carbone, R., Nicoletti, R., and Borghi, A. M. (2011). Graspable objects shape number processing. *Front. Hum. Neurosci.* 5:147. doi: 10.3389/fnhum.2011.00147
Rathkopf, C. (2021). "Neural reuse and the nature of evolutionary constraints," in *Neural Mechanisms. Studies in Brain and Mind*, Vol. 17, eds F. Calzavarini and M. Viola (Cham: Springer), 191–208. doi: 10.1007/978-3-030-54092-0_9
Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. (2002). "From mirror neurons to imitation, facts, and speculations," in *The Imitative mind: Development, Evolution, and Brain Bases*, eds A. N. Meltzoff and W. Prinz (Cambridge: Cambridge University Press), 247–266. doi: 10.1017/CBO9780511489969.015
Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141. doi: 10.1016/0926-6410(95)00038-0
Roli, A., and Kauffman, S. A. (2020). Emergence of organisms. *Entropy* 22:1163. doi: 10.3390/e22101163
Samuels, R. (1998). Evolutionary psychology and the massive modularity hypothesis. *Br. J. Philos. Sci.* 49, 575–602. doi: 10.1093/bjps/49.4.575
Samuels, R., Stich, S., and Bishop, M. (2004). "Ending the rationality wars: how to make disputes about human rationality disappear," in *Common Sense, Reasoning, and Rationality*, ed. E. Renee (New York, NY: Oxford University Press), 236–268. doi: 10.1093/0195147669.003.0011
Samuels, R., Stich, S., Tremoulet, P. D. (1999). "Rethinking rationality: from bleak implications to darwinian modules," in *What is Cognitive Science?*, eds E. Lepore and Z. Pylyshyn (Oxford: Blackwell), 74–120. doi: 10.1007/978-94-017-1070-1_3
Sanabria, F., and Killeen, P. R. (2005). All thumbs? *Behav. process.* 69, 143–145. doi: 10.1016/j.beproc.2005.02.015
Sato, M., Cattaneo, L., Rizzolatti, G., and Gallese, V. (2007). Numbers within our hands: Modulation of corticospinal excitability of hand muscles during numerical judgment. *J. Cogn. Neurosci.e* 19, 684–693. doi: 10.1162/jocn.2007.19.4.684
Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychol. Review* 63:129. doi: 10.1037/h0042769
Simon, H. A. (1990). Invariants of human behavior. *Annu. Rev. Psychol.* 41, 1–20. doi: 10.1146/annurev.ps.41.020190.000245
Solé, R. V., Ferrer-Cancho, R., Montoya, J. M., and Valverde, S. (2002). Selection, tinkering, and emergence in complex networks. *Complexity* 8, 20–33. doi: 10.1002/cplx.10055
Spellman, B. A., and Schnall, S. (2009). Embodied rationality. *Queens Law J.* 35, 116–117. doi: 10.2139/ssrn.1404020
Stanovich, K. (2011). *Rationality and the Reflective Mind*. Oxford: Oxford University. doi: 10.1093/acprof:oso/9780195341140.001.0001
Sterelny, K. (2007). Social intelligence, human intelligence and niche construction. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 719–730. doi: 10.1098/rstb.2006.2006
Sterelny, K. (2010). Minds: extended or scaffolded? *Phenomenol. Cogn. Sci.* 9, 465–481. doi: 10.1007/s11097-010-9174-y
Todd, P. M., and Gigerenzer, G. (2000). Précis of" Simple heuristics that make us smart". *Behav. Brain Sci.* 23, 727–741. doi: 10.1017/S0140525X00003447
Tooby, J., and Cosmides, L. (1992). "The psychological foundations of culture," in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, eds J. Barkow, L. Cosmides, and J. Tooby (New York, NY: Oxford University Press), 19–136.
Tooby, J., and Cosmides, L. (2005). "Conceptual foundations of evolutionary psychology," in *The Handbook of Evolutionary Psychology*, ed. D. Buss (Hoboken, NJ: Wiley), 5–67. doi: 10.1002/9780470939376.ch1
Tooby, J., and Cosmides, L. (2007). Evolutionary psychology, ecological rationality, and the unification of the behavioral sciences. *Behav. Brain Sci.* 30, 42–43. doi: 10.1017/S0140525X07000854
Tramacere, A., and Ferrari, P. F. (2016). Faces in the mirror, from the neuroscience of mimicry to the emergence of mentalizing. *J. Anthropol. Sci.* 94, 1–14.
Tramacere, A., Pievani, T., and Ferrari, P. F. (2017). Mirror neurons in the tree of life: mosaic evolution, plasticity and exaptation of sensorimotor matching responses. *Biol. Rev.* 92, 1819–1841. doi: 10.1111/brv.12310
Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
Tversky, A., and Kahneman, D. (1983). Extension versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90, 293–315. doi: 10.1037/0033-295X.90.4.293
Uchiyama, R., Spicer, R., and Muthukrishna, M. (2020). Cultural evolution of genetic heritability. *Behav. Brain Sci.* 21, 1–147. doi: 10.1101/2020.06.23.167676
Vera, A. H., and Simon, H. A. (1993). Situated action: a symbolic interpretation. *Cogn. Sci.* 17, 7–48. doi: 10.1207/s15516709cog1701_2
Vrba, E. S., and Gould, S. J. (1986). The hierarchical expansion of sorting and selection: sorting and selection cannot be equated. *Paleobiology* 12, 217–228. doi: 10.1017/S0094837300013671
Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002
Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton, NJ: Princeton University Press.
Williams, G. C. (1992). Gaia, nature worship and biocentric fallacies. *Q. Rev. Biol.* 67, 479–486. doi: 10.1086/417796
Wilson, M. (2002). Six views of embodied cognition. *Psychonom. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322
Wittgenstein, L. (1953). *Philosophical Investigations*. New York, NY: Macmillan Publishing Company.

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.

Copyright © 2022 Mastrogiorgio, Felin, Kauffman and Mastrogiorgio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 21 March 2022
doi: 10.3389/fpsyg.2022.805386

Dual Process Theory: Embodied and Predictive; Symbolic and Classical

*Samuel C. Bellini-Leite**

Utilization Systems Department, Minas Gerais State University, Belo Horizonte, Brazil

Dual Process Theory is currently a popular theory for explaining why we show bounded rationality in reasoning and decision-making tasks. This theory proposes there must be a sharp distinction in thinking to explain two clusters of correlational features. One cluster describes a fast and intuitive process (Type 1), while the other describes a slow and reflective one (Type 2). A problem for this theory is identifying a common principle that binds these features together, explaining why they form a unity, the unity problem. To solve it, a hypothesis is developed combining embodied predictive processing with symbolic classical approaches. The hypothesis, simplified, states that Type 1 processes are bound together because they rely on embodied predictive processing whereas Type 2 processes form a unity because they are accomplished by symbolic classical cognition. To show that this is likely the case, the features of Dual Process Theory are discussed in relation to these frameworks.

Keywords: dual process theory, embodied cognition, bounded rationality, predictive processing, cognitive science

OPEN ACCESS

Edited by:
Riccardo Viale,
University of Milano-Bicocca, Italy

Reviewed by:
Marius Usher,
Tel Aviv University, Israel
Denis Perrin,
Université Grenoble Alpes, France

***Correspondence:**
Samuel C. Bellini-Leite
samuel.leite@uemg.br

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *30 October 2021*
Accepted: *18 February 2022*
Published: *21 March 2022*

Citation:
Bellini-Leite SC (2022) Dual Process Theory: Embodied and Predictive; Symbolic and Classical. Front. Psychol. 13:805386. doi: 10.3389/fpsyg.2022.805386

INTRODUCTION

Embodied cognition has been proposed as an alternative to symbolic processing since it started to grow in the 90s. Although it is true that embodied cognition contrasts with traditional cognitive science, the possibility that these frameworks might explain different kinds of processes in cognition is overlooked. In the same sense that different framework in physics such as quantum mechanics, general relativity and even the traditional classical mechanics co-exist, each explaining parts of our world, it is likely that 4E cognition, traditional cognitive science, connectionism, and predictive processing can co-exist if we understand to which domains of cognition these apply (Bellini-Leite, 2017). A theory of everything in cognition should most likely attempt to unify parts of these proposals rather than to keep only one.

Evidence in the reasoning and rationality literature has consistently pointed to the idea that human rationality is bounded by proximal stimuli and cognitive limitations. This has led to the interpretation that humans do not have a perfect logical or probabilistic problem-solving system but rather diverse heuristics, algorithms or simple mechanisms that are used to deal with environmental challenges. These conclusions come from experiments which show how people respond in puzzling ways to certain questions. But the reason certain systems are bounded and how they are bounded should vary greatly depending on which systems these are. Thus, we need to consider divisions in cognition as well to understand bounded rationality (Bellini-Leite and Frankish, 2020).

One currently popular way to divide types of cognitive processes is Dual Process Theory (DPT). This theory, proposing there are two distinct processes, Type 1 (T1) and Type 2 (T2), underlying higher-order thinking has recently received much attention for explaining the evidence in reasoning, judgment and decision-making tasks. DPT claims there must be a sharp distinction between two clusters of correlational features. One cluster describes a fast and intuitive process, while the other describes a slow and reflective one (Evans, 2008; Kahneman, 2011; Evans and Stanovich, 2013). Some T2 core features are heavy working memory load, explicitness, low capacity, high effort and slowness, while T1 central features are weak loading on working memory, implicitness, high capacity, low effort, and speed.

However, Samuels (2009) notes that even if one considers the evidence to be convincing and the dichotomy of processes T1 and T2, along with their property clusters (termed S1 and S2), well placed, we still have a basic research question open, which he calls the unity problem:

"though positing mechanisms is a standard strategy for explaining the existence of property clusters, it does not, by itself, constitute a satisfactory explanation. Rather one needs to specify those features of the proposed mechanisms that account for such clustering effects. In the present case, we need to specify those characteristics of type-1 systems that yield S1-exhibiting processes, and those properties of type-2 systems that yield S2-exhibit-ing processes. Again, this does not strike me as a serious objection so much as a challenge for future research—one that requires a more detailed account of the systems responsible for type-1 and type-2 processes." (Samuels, 2009, p. 141).

The unity problem should not be confused with the reference problem (Samuels, 2009). The reference problem of DPT is the problem of determining what the theory is about, to which a possible answer would be "about distinct systems" or "different minds" or "modes" (see Bellini-Leite, 2018). After answering the reference problem, the unity problem remains, we need to determine why these two chosen structures (types, systems, minds, or modes) each form a unity with individual properties, or what the mechanisms that explain this unity are.

The current manuscript attempts to advance in the unity problem by showing how T1 features align with predictive processing and how T2 features align with symbolic processing. Sloman (1996) has done a similar job with the theories of the 90s. However, his project was not developed along the years. Since there have been a multitude of related dual process theories (see Evans, 2008) with different features proposed to explain different areas of cognition, Evans and Stanovich (2013) had to review what the main features for the case of reasoning, judgment and decision making are. Further development in terms of fast or slow responses have also been proposed by Kahneman (2011) and De Neys (2017). Previous attempts at approaching the unity problem like Epstein et al. (1996) and Sloman's (1996), therefore, refer to different theories altogether. The view that there is an "associative" system 1 and a "rule-based" system 2 is somewhat out of line with the developments both of current DPT and current cognitive architectures, like predictive processing. Moreover, there are newly discovered characteristics specific to predictive processing that explain T1 features more than an associative account does. Hopefully these characteristics will be made clear along the argument.

Perhaps a weak spot of the current proposal is that for it to stand, two other hypotheses need to be true:

(1) Predictive processing is aligned with embodied cognition.
(2) Current formulations of DPT adequately explain reasoning, judgment, and decision-making.

Although I will attempt to explain and defend these two hypotheses along the manuscript, I cannot make a full case for each of them here. Hypothesis 1 is defended mainly by Clark (2013a, 2015, 2016) and although hypothesis 2 stems from the reasoning, judgment and decision-making literature starting from the 60s, the current formulation of the theory is what needs to hold (Schneider and Chein, 2003; Kahneman, 2011; Evans and Stanovich, 2013), with some emphasis given to speed, explicitness and implicitness, autonomy, and working memory.

The manuscript is organized to reflect how these features of DPT can be best captured by each of the two considered cognitive architectures. But before getting into the argument, I start by summarizing how Clark (2016) has argued that predictive processing is embodied. Then, I explain which features of DPT will be considered. I then lay out the general hypothesis for how predictive processing and symbolic accounts of cognition could go together to explain human reasoning. Finally, I go on to argue in a few sections that this hypothesis is plausible by showing how it explains the different features of T1 and T2 processing accordingly.

HOW PREDICTIVE PROCESSING IS ALIGNED WITH EMBODIED COGNITION

Although any cognitive proposal speaking of representations and brain circuits were previously considered to be distanced from embodied cognition, Andy Clark (2013a, 2015, 2016) has recently published extensively on how predictive processing can go along with or even enrich embodied, situated, and extended accounts. Philosophers have displayed worries that Andy Clark, by adopting predictive processing, had moved to a different camp.

Predictive processing suggests the brain is in a active cycle of predicting what will perturb it in a proximal and distal future. Instead of being understood as reading input from the world, the predictive brain uses statistics to anticipate input before they arrive. These predictions are based on expectations (or a statistical generative model) which foresees the most likely outcome of stimuli.

These models suggest the brain is formed by a hierarchy of processing (comprising higher and lower levels) where multiple layers of neurons are organized to compose a network with two major streams of information flow. On the top-down flow, each higher layer attempts to predict the workings of the one underneath it. The bottom-up flow conveys error correction on previously attempted predictions to each layer above. If predictions of a given event are on track, lower sensory stimulation is attenuated. On the other hand, if predictions are

misleading, sensory stimulation flags the difference between what was predicted and what was perceived so that the system tries to overcome such gap. This, prediction error minimization, Clark (2013a) claims, is the brain's major goal.

Clark (2013a) notes an interesting shift the predictive processing approach suggests. It proposes that the forward flow consists not so much of all the features that were detected to be passed onward to higher layers but only the error necessary to correct and update models. Instead of conveying all information from the environment, rather, it provides a natural funnel which guarantees processing economy by focusing on newsworthy information in the form of error correction. Predictions flow downward at each layer and error correction escalates upward showing faults to be corrected for future models. Thus, lower layers bring novelty since they detect the most recent error correction to propagate upward, but the higher layers have error correction coming from various other strands of the network. That is, the higher layers have models corrected from various sources while the lower layers will have tokens of newest corrections to be made, that is why at any given time there is not one generative model but various co-evolving models and also why there is a bidirectional flow of information.

Prediction error is also related to the concept of surprisal. Predictions are based on models which are a form of subpersonal expectation. When these expectations are not met, prediction error flags them with surprisal. In order to predict, the brain is always attempting to find a match from higher expectations to the next information reported from the bottom. Surprisal occurs, therefore, when there is a mismatch between expectation and the information conveyed by error signaling. The goal of the system at every second is to minimize surprisal. To reach such goal it must constantly update its models in order to correspond to novelty. Having tuned predictions enables the system to keep surprisal at the lowest level possible.

One of the issues in considering predictive processing as an embodied framework is its intensive use of representations to explain cognition. Embodied and situated cognition had as one of its central tenets that cognitive science had lost itself in the use of cognitive representations, and that the world itself could serve as its best model.[1] When Clark (2013a) then claims that for every aspect of cognition the brain keeps statistical models of reality at first this seems like a huge departure from situated approaches. But it is not so. First, these representations are nothing like symbolic stand-ins, they are not mirrors of reality, and there is not an inner token for each outer stimuli. In predictive processing, these statistical models keep information only of organism-relevant stimuli and events, generating predictions that enable the organism to select affordances (see Gibson, 1979). The word 'model' might also sound misleading here. A model airplane is a replica of a real airplane. However, a statistic model bares a sort of morphism relation to some content, but it does not replicate the content. Further, Clark argues these statistical models do not address an organism neutral world nor even all the aspects that could be relevant to the organism. Unlike classical models, these representations are not stored in blocks and do not cause overload resulting in computational explosion, rather, Clark argues these models have been mathematically studied and found to be extremely feasible and have been applied cheaply to computer simulations. Also, Clark notices there is a sense in which the world can be its best model even if models are guiding perception, no contradiction included. The reason is that these models are not replacements for the world, instead they enable the agent to use the best of what is available in the world. If you follow this trend, the world is not its best model in a literal sense, because (unless you have very specific sensors like insects) the world actually has a majority of irrelevant information for a given agent, just think of a loud, noisy city. There is a sense in which the world is bombarding us with bad information and noise. The true sense of the expression "the world is its best model" is actually preserved by Clark. That is, that our prediction mechanisms should be at each millisecond corrected by errors in the environment, thus the environment really is what shapes us, but we need to let the right information shape us, not any irrelevant information from the environment. Generative models actually permit us to be tuned to the human-relevant environment.

Another issue is that of the implied metaphysics. If our systems only get information (error) relative to predictions, does that imply indirect perception? Clark's (2016) answer is yes and no, or "non-indirect perception." The worry of critics of indirect perception is that we might be locked from the true world itself. The point, once again, is both that we need the mechanisms to engage in the relevant world and that the world itself, as in free from agent intentional perception, is senseless. When we go to the stadium, predictive processing is what enables us to see a soccer game instead of physical objects colliding. Therefore, Clark argues perception cannot be direct since it is mediated by expectations, but no further worry needs to be pursued about "losing the world." This is because predictions allow us to see the part of the world that is relevant to humans, without these models, if we could perceive at all, a random part of a scene would be as relevant as a face.

Finally, there is the embodied coupling of perception and action. This is achieved in predictive processing because actions are a consequence of external and proprioceptive perception and because action reduces prediction error by directing what sort of stimuli perturbs the sensory system. Therefore, to solve a jigsaw puzzle we need to actively engage the objects with our hands, rotating, moving and organizing them, and in every such attempt, action is framing the sort of stimuli that perception will receive, choosing what "shots" of the world are taken. This interplay between action, body and world is what solves a tough jigsaw puzzle, one cannot succeed just by staring at it and thinking. Clark (2016) shows how embodied proposals of the mind can assume diverse shapes. His version might not be very representative of the movement, however, if embodied proposals of the mind are to be relevant to cognitive science, then these must adopt or develop models of cognition like Clark does with predictive processing.

[1] Although it is important to note that embodied cognition comes in various forms. For instance, Glenberg (1999) prefers to speak of embodied models.

FEATURES OF DUAL PROCESS THEORY

Dual process theories come in various shapes. If we simply put all dual process theories that have been proposed together we arrive at a multi-theoretical cluster of attributes for each type of processing (Evans, 2008), thus the correlational features for T1 processes would be: unconscious, implicit, automatic, low effort, rapid, high capacity, default, holistic, perceptual, evolutionary old, follows evolutionary rationality, shared with animals, non-verbal, modular, associative, domain-specific, contextualized, pragmatic, parallel, stereotypical, independent of general intelligence, independent of working memory. In this multi-theoretical cluster version, the correlational features for T2 processes would be: conscious, explicit, controlled, high effort, slow, low capacity, inhibitory, analytic, reflective, evolutionarily recent, follows individual rationality, uniquely human, linked to language, fluid intelligence, rule based, domain general, abstract, logical, sequential, egalitarian, heritable, linked to general intelligence, limited by working memory capacity. Evans (2008) noted that positing all these features as defining characteristics of these types of processing is troublesome, because these characteristics will not always stand.

It is quite improbable that such a strong co-occurring requirement meets reality. Because even if, say, only six dichotomies are advanced, there are still 64 possible combinations of these features that need always co-occur. If DPT were proposing such an alignment assumption for all these features (see Stanovich and Toplak, 2012) then only one of these possible 64 combinations of features would be enough to falsify the theory. Suppose these dichotomies were: conscious/unconscious, explicit/implicit, controlled/automatic, serial/parallel, slow/fast, resource dependent/resource free. Each process that lacked one element of these aligned features would serve as evidence to falsify DPT. For example, a process that was conscious, explicit, controlled but parallel would be evidence for falsification, even considering that most features of such process were rather aligned than unaligned.

Critics have mentioned how DPT features are not well defined (Keren, 2013). However, one can reformulate this theory to account for new evidence. We just have to be aware that if this happens repeatedly, we should start losing our interest in DPT (see Lakatos et al., 1979). The correct way to go about this is to try to consider which would be the crucial features of dual process theories of reasoning such as Schneider and Chein (2003), Kahneman (2011), and Evans and Stanovich (2013) have attempted. This should be at least a combination of features which various theorists of this research field or similar research fields could agree on. By assuming the alignment assumption at least for defining features, the theory gains in predictive power and rigor. Therefore, the more defining features one assumes, the stronger are the empirical consequences; it will predict more but also be more easily false. At least for defining features, predetermined scientific predictions must be possible, or else these features are not truly defining.

Based on the weight placed on these features in the works Schneider and Chein (2003), Kahneman (2011), and Evans and Stanovich (2013) we will focus on five main dual process distinctions: working memory use, explicit and implicit representations, automaticity, and speed.

HYPOTHESIS

To solve the unity problem, I propose a hypothesis to combine embodied predictive processing with symbolic classic approaches. The hypothesis, simplified, states that T1 features form a unity because they rely on embodied predictive processing whereas T2 processes form a unity because they are accomplished by symbolic classical cognition.

Daniel Kahneman (2002, p. 450) wrote that "From its earliest days, the research that Tversky and I conducted was guided by the idea that intuitive judgments occupy a position [. . .] between the automatic operations of perception and the deliberate operations of reasoning." Kahneman and Frederick (2002, p. 50) claimed that intuitive thinking is "perception-like" and that "intuitive prediction is an operation of System 1." Further, that "The boundary between perception and judgment is fuzzy and permeable: the perception of a stranger as menacing is inseparable from a prediction of future harm." Kahneman et al. (1982) have been speaking of "intuitive predictions" for a long time. What I hold is the link between perception and intuition obtains because T1 judgments are embodied predictions. These authors have been noticing that intuition is somewhat like perception and have used the term prediction as what intuition does, but they have not argued for a framework for T1 processes.

Perception clearly has input functions, but what is interesting for DPT of reasoning and decision making is that T1 processes have an output function, in the sense that they generate answers to problems. The predictive processing approach gives a clear output form to perception, by emphasizing its generative character. Thus, a strong claim I want to hold is that T1 processing answers (or output functions) are predictions.

The pivotal role of expectations for determining T1 predictions have gone mostly unnoticed even though task construal in the reasoning and judgment paradigm has been mostly a task of manipulating subject's expectations. The argument for how this occurs in reasoning is that T1 processes take information over prior occurrences and over the current set of states (likelihood) and yields a fast prediction (posterior). If the time constraint is rigid, these predictions will generate actions (inner mental responses or, if too rigid, movements). If the system has time, then these predictions will be available for T2 evaluation. Thus, T2 processes receive T1 predictions as input to analyze and possibly override. That is why manipulating subject's expectations in a task causes their T1 answers to vary accordingly and requires T2 effort to override them.

According to the current hypothesis, T1 processes deal with content encoded in the form of probability density functions, which means there is no symbol and no definite content, but values, means and standard deviation influenced by previous movements and previous world contingencies. Manipulating prior information biases the distribution into one or another direction, closer to or further from a certain value. These functions are not stored in a memory bank but distributed from

the responsible brain regions over to external organs and body parts through neural connections. The values in the distribution do not represent objects directly and discretely, they refer to distinct aspects of the input when perceptual systems are dealing with such objects. This is in line with T1 processes being easily biased when working with references to similar properties, like similar numbers, objects, rhymes or pet names; very often the incorrect value is picked from a distribution. This is also in line with claims of embodied proposals that the world is not represented in symbols.

Finally, T1 processes are subpersonal (see Frankish, 2004, 2009) and their predictions are made by the same systems which process perception. A clear example is that a judgment (a prediction) about facial expressions is related to the FFA (see Egner's et al., 2010). The idea is that perception is not passive but already comes with predictions, and when in problem solving, such prediction is precisely the T1 answer. I do not want to claim that T1 processes are purely perceptual (if in contrast to cognitive), only that such predictions stem from perceptual processes. Kahneman's (2011) example of judgments of angry facial expressions shows how this is expected of DPT. Kahneman (2002) and Kahneman and Frederick (2002) have also argued that the list of features of T1 processing is shared with perception mechanisms. What I propose to do is examine central T1 features to show that it is shared because both (or at least part of) perception and T1 processes work in the manner described by predictive processing, which is also in-line with the claims of embodied cognition that there is no sharp link between perception and reasoning.

It is interesting to note that Clark's (2016, p. 257) embodied version of predictive processing is described accordingly: "Fast, automatic, over-learnt behaviors are especially good candidates for control by models taking a more heuristic form. The role of context-reflecting precision assignments is then to select and enable the low-cost procedural model that has proven able to support the target behavior. Such low-cost models [. . .] will in many cases rely upon the self-structuring of our own information flows, exploiting patterns of circular causal commerce (between perceptual inputs and motor actions) to deliver task-relevant information 'just in time' for use."

Another way to put it, which fits neatly with the framework developed here is: "we need only note that very low-precision prediction errors will have little or no influence upon ongoing processing and will fail to recruit or nuance higher level representations." (Clark, 2016, p. 148) That is, if the task is overlearned and errors are weighted as low, systems will act without further recruiting. This can be understood as a hypothesis for automaticity, which has been used so much in psychology but without an explanation for why it differed from controlled processing.

The general idea I want to hold for T2 processing is that it works like a classical machine for reasoning, such as the General Problem Solver (GPS, Newell and Simon, 1963). The GPS was one of the first attempts to mimic human reasoning. Its purpose was to respond to logical problems like humans would. Of course, human thought is different in various ways from those first machines; but T2 processes are somewhat alike. However, this classical machine only makes sense in the brain if it exists in the wider setup of a predictive processing network generating T1 responses.[2] Thus, like Newell's (1980) physical symbol system, when facing a reasoning problem, T2 processing opens a problem space containing an expression that designates the initial problem (how it was digitized or interpreted) and an expression that designates a solution, which was produced by a probabilistic prediction (T1 processing). Having the initial expression and the predicted expression in the problem space, T2 processing then uses its move generators to attempt to reduce differences between them and sometimes finds different solutions in such path or illuminates something that previously had not come about. Move generators (or operators in the GPS) are mechanisms that apply rules, which might be fed from different sources, such as logic, mathematics or philosophy (say Occam's razor). These generators are likely to be flexible, in that they can change depending on the problem. Thus, although the basic structure is that of a logical machine that works on symbolic expressions it could be set up to apply paraconsistent rules, for instance. This is possible because although it does not work with contradictory expressions it could work with expressions that designate contradictory expressions. Therefore, it is free to work out any sort of principle to solve tasks, exhibiting the property known as universality in computation.

I want to make it clear that I am taking "classical architecture" and "predictive processing" both as whole packages. Computations have universal features, classical architectures could work with representations of probabilities and predictive processing could be realized by a serial machine. But this is out of their standards. To claim that I am taking the whole package means that I am taking features of classical architecture and predictive processing that usually come together in all levels. Therefore, I am speaking of a classical architecture in the form of a serial physical symbol system performing heuristic search such as a GPS (Newell and Simon, 1963, 1976; Newell, 1980) which are responsible for T2 processes and embodied prediction as a hypothesis about how networks in the brain form a system with the body that encodes probabilistic representations of stimuli which are used to infer properties of objects in the world, and act upon them being responsible for T1 processing (Clark, 2013a, 2016; Hohwy, 2013).

Some caveats are in order. We should not want to suppose that there are two processes for the mind as whole, since that would be too strong of a hypothesis and evidence from any cognitive function would serve to falsify it. Therefore, it is important to restrict this hypothesis first to the scope of reasoning, judgment and decision making. Also, a huge list of features have been ascribed to DPT (see Evans, 2008) and it might be the case that some do not follow the current hypothesis. Although I have not identified such features that would not work at all with such hypothesis, Evans (2008) argues that this group of features cannot work coherently together, so some must be off track. Decoupling is an important feature which was not mentioned here, but that is because it requires extensive work, and the manuscript is limited

[2]This is also the case for meaning in the sense of Harnad (1990). In this framework, classical symbols reference instances of predictive processing exchanges with the world.

by space. Interestingly, if this hypothesis stands to empirical tests and there is further reason to believe it, then it could even help expose those features from Evans (2008) which were off track.

This is the general hypothesis. None of what is claimed so far is novel in itself, just in the interpretation of how these claims could work together. To show that this interpretation is likely true, I will proceed by showing how central T1 features are best captured by predictive processing and how central T2 features are best captured by classical architectures.

IMPLICIT AND EXPLICIT FEATURES

Although the "implicit" and "explicit" distinction is vastly used in the literature in the sense of access, this is also the use of "consciousness." When it comes to the implicit and explicit distinction what is unique and coherent (even with the word) is the representational format (see Bellini-Leite, 2021). If we want a difference between the explicit and implicit features in DPT we need to have different representational formats for each type of process.

Predictive processing has a unique representation format, content is encoded in probability density functions. These functions these functions do not disambiguate items discretely, rather, they gather multiple occurrences of events and possibilities from models ranging from various areas of the cortex, body and world contingencies to generate probability. This is most likely the (usually unexplained) meaning of an implicit format in cognitive psychology, one that encodes probability of previous occurrences of movements and world contingencies and not representations by means of symbols. This implicit format is not the type of format T2 reasoning can work with, T2 processes need symbolic, unit-like objects to reason over, and that is the meaning of an explicit representational format: disambiguated stand-ins for a unified object.

A representation is explicit when it has a graspable representational format. By this I mean that subjects seem to grasp such content with ease and they verbally report having done so. This contrasts with fuzzy content which one does not know how to speak of or even think clearly about. It seems we can be conscious both of fuzzy and disambiguated content.

Classical architectures can have more fixed access to the content it deals with than predictive processing networks because of differences in symbolic representations and probability density distributions. Probability density distributions are responsible for much of what gives predictive processing its explanatory success. Representing information in such fashion allows for statistical processing of previous input and for generative guesses for future outcomes involving diverse elements distributed between the cortex and the world. There is a problem with this representation, however, which is keeping a probabilistic take on states of objects, since it includes too much. Having this probabilistic state usually allows embodied agents to act more rapidly, but there are times when we need precise, definite, properly discrete information about an object. In such times, only one answer is valued and related ones should not interfere. To account for this, Clark (2016) speaks of single peak probability distribution functions, representations where each distribution must have a single best explanation. Thus, instead of having various related peaks indicating possible outcomes of movements and world contingencies, only one is enforced. "One fundamental reason that our brains appear only to entertain unimodal (single peak) posterior beliefs may thus be that—at the end of the day—these beliefs are in the game of informing action and behavior, and we can only do one thing at one time." (Clark, 2016, p. 188).

Now, what happens when you have a single peak probability density function is that it acts like a discrete symbolic representation. That is, all other possible states are denied in favor of a single active state. When this is the case, advantages of embodied prediction of using statistical encoding and generative models over the multitude of possible body-world relations are lost and some other form of computing needs to take place. When using single-peak probability density functions you lose the effects of having various related instances as possible outcomes to gain feasibility, you lose effective predictive processing.

Clark (2016) admits that sometimes values in a density function need to be reduced to only one. However, what goes by unnoticed is that this is precisely the effect of turning it into a symbolic representation. This eliminates uncertainty, and possibly is related to subjects being able to grasp the content. You can grasp something that is clearly defined but you cannot easily grasp the meaning of something like values in a probability density function. They are fuzzy because they cannot be simply well defined. It is precisely their fuzziness that allows for context-sensitivity and fluid embodied cognition.

The reason classic symbols are graspable seems to be because working memory can store them and use them in symbolic manipulation. Working memory cannot store all values of a probability density function or manipulate the dynamic workings of a complex relation between movements and world contingencies. But when this whole dynamic is referenced by a single symbol, this symbol can then be treated as a constituent in an expression. When that occurs, the classical architecture can work with compositionality (see Fodor and Pylyshyn, 1988).

As Fodor and Pylyshyn (1988) have explained,[3] the point for compositionality in making content graspable is that manipulations of these expressions can then be easily tracked. Rules and semantic content become related to the inner structure of the computation. Then, when taking some content as a symbolic object, it becomes identifiable in multiple expressions preserving its identity. In contrast, values in a density function might lose their identity, in fact, we should want that to happen if context is to shape their identity.

Even the steps in processing can become symbols themselves by being stored as expressions to be used in metacognition. Therefore, when we are reasoning in a syllogism, we can keep premises in working memory and also the steps used to extract one from the other. Of course, these are fleeting, but also, the way to make them less fleeting is by reducing uncertainty and naming a step or a premise by a letter or a simple symbol, say MP. So it seems plausible that representations in classical

[3]Please note that there are also issues of using the classical interpretation for T1 processing which have been stressed in Bellini-Leite and Frankish (2020).

architectures should make both content and steps of processing more graspable because of ease in determining their identity, reducing uncertainty. Therefore, if the current hypothesis holds, we should want to speak of explicit representations as symbolic and implicit ones as distributed, probabilistic, and multi-valued.

AUTOMATICITY VERSUS WORKING MEMORY

Automaticity concerns overlearned skills, and overlearned skills here can be understood as skills over tasks that became predictable. Let us use the classic example of learning how to drive a non-automatic car to see how predictive processing relates to automaticity. When we first sit behind the driver's wheel, even if we have knowledge on what must be done, our systems cannot coordinate all such knowledge in order to be useful (and safe). When we train ourselves the correct order of using gears, wheel turning and pedals, we are tuning our predictive processing systems to the usual occurrences of car handling. Of course, before driving, our systems cannot have useful priors on the matter. By letting our system engage with the stimuli necessary for driving we tune it to that particular context, that is, we learn embodied/predictive routines. For instance, when in cliffs, our systems need to predict the exact moment to press the clutch at the correct strength to manage the cliff. But not only this, our systems need to predict more precisely when another car is stopping in front of us. They need to predict the order of gears and when they will be necessary, also when the car is being misused through auditory clues.

Various cues are used to predict near-future occurrences. The system needs to know, for various states, that if it is in a given state, another given state is the most probable to follow. Once the system learns various important cues that lead to efficient predictions, it can handle most driving abilities automatically. Thus, an experienced driver will incur in far less surprisal instances than a novice driver. In fact, the higher surprisals which will come by are in the form of unpredictable changes in the environment, such as an animal crossing the road. In contrast, the surprisal which will mostly concern the novice is in terms of actions to handle the machine, so an animal can go by unnoticed. If our systems have no useful priors for driving, they need to rely on effortful controlled skills to train predictions systems, but these effortful controlled skills cannot be predictive processing skills themselves.

Unlike driving, daydreaming seems to be turning attention and effort to oneself and forgetting the world for a while. What seems to happen to attention and working memory in predictable situations is that it turns inward, it starts to generate novelty or monitor inner performance. This is observable in habituation, a phenomenon much known by psychologists where exposure to repeated stimuli decreases attention paid to it. Working memory is an online and ever-ready mechanism for dealing with further uncertainties and unpredictable information. It seems to be that the more predictable a given state is, the less working memory resources systems will consume in processing it. Working memory is needed when predictive processing fails.

The literature in predictive processing does not necessarily shun working memory, but just to illustrate how important this concept is to such framework, it is interesting to see how it is mentioned only once in Clark's (2016) book and absent from Hohwy's (2013) book and other work in predictive processing. Working memory is mentioned 119 times in Frankish and Evans' (2009) review of DPT. In other words, it is probably not a very central tenet of predictive processing. And there is every reason for working memory not to be a relevant tenet of predictive processing. This is precisely because stronger load on working memory concerns cases where the information that needs processing is unpredictable, or is not well accommodated by any statistical judgment, in fact, if the general prediction by statistics schema fails deeply to account for some relevant data, then it seems plausible that another type of processing should be applied. When predictions are working, then, working memory is mostly dispensable.

Working memory is not a feature of how predictive networks work. In contrast, a working memory is a necessary component of a classical architecture, both structurally and functionally. Thus, I argue that it is unlikely that predictive processing can do away completely with models of classical processing as proponents usually hold.

In a Von Neumann (1945) architecture there is a primary storage for holding what to do and what is done, which is basic for the functioning of the machine. More importantly, in a physical symbol system, the model proposed by Newell (1980) for classical cognitive science, a similar component that stores operators and expressions which are being used at a given moment is necessary. In Newell's (1980, p. 159) words "This organization implies a requirement for working memory in the control to hold the symbols for the operator and data as they are selected and brought together." and "[. . .] working memory is an invariant feature of symbol systems."

A working memory in cognitive psychology is usually taken to be a system with executive functions and not only a storage. As Baddeley (1992, p. 557) explains "Although concurrent storage and processing may be one aspect of working memory, it is almost certainly not the only feature." In fact, it is such executive functions which pushed the need for the concept of a working memory instead of just a short-term storage. Baddeley (1992, p. 556) explains that "This definition has evolved from the concept of unitary short-term memory system. Working memory has been found to require the simultaneous storage and processing of information." Instead of being just a short-term storage, the model also includes "an attentional controller and the central executive, supplemented by two subsidiary slave systems" (Baddeley, 1992, p. 556). These slave systems are storages for different types of content, such as phonological or visual. More important for present purposes are the "attentional controller" and "the central executive."

It seems these claims on the processing abilities of working memory are not as clear as what has been said of its storage function. For instance, Baddeley (1992) claimed that the attentional controller was an additional component, but he also claims "the central executive [. . .] is assumed to be an attentional-controlling system." We understand executive functions are

equivalent to the application of operators in Newell's (1980) architecture or to the functioning of a processing unit of a Von Neumann architecture which carries out logical or arithmetic procedures. As for the attentional controller, it is not directly related to attention as in the psychological concept, but to "attention" as in a Turing machine which can only focus on certain elements each moment. This function would also be something like the control unit of the Von Neumann architecture which mediates the flow of processing by providing timing and control signals. With the argument that T2 processing depends on working memory, what is meant is that that a temporary storage is needed but also other mechanisms which mediate symbol processing, or that something like the physical symbol architecture of Newell (1980). Certain operators must be applied to elements of this storage and there must be a control of which expressions are being used at a given moment.

There are two choices here, one is to say that the concept of the working memory refers to Newell's (1980) physical symbol architecture as a whole, or that it is the storage component of such architecture. Since the literature (Baddeley, 1992) sustains the importance of executive functions which differentiates working memory from the concept of short-term memory, the first choice seems more plausible: that working memory is not only a memory, but a system which has very similar (if not the same) properties to that of Newell's (1980).

Newell's (1980) architecture maintains properties of a Von Neumann architecture which maintains (or instantiates) properties of Turing Machines. By transitivity (and if the hypothesis is on track) there should also be some similarity between working memory and Turing machines. First, it is enlightening to notice that Turing started to think about his machine by trying to mimic what he was doing in his own abstract thought, such as the processes he was executing when doing mathematics. Thus, since we must process in working memory what we are thinking consciously and with effort, which clearly was the type of thought he had to engage in for his work, what he probably was doing then was an inspection of the functioning of his own working memory. If this supposition is the case, it would also be no surprise to find similarities of working memory and a Turing machine.

Consider this part of Turing's (1936, p. 250) intuitive argument: "The behavior of the computer at any moment is determined by the symbols which he is observing, and his 'state of mind' at that moment. We may suppose that there is a bound B to the number of symbols or squares which the computer can observe at one moment. If he wishes to observe more, he must use successive observations. We will also suppose that the number of states of mind which need be taken into account is finite. The reasons for this are of the same character as those which restrict the number of symbols. If we admitted an infinity of states of mind, some of them will be 'arbitrarily close' and will be confused. Again, the restriction is not one which seriously affects computation, since the use of more complicated states of mind can be avoided by writing more symbols on the tape."

This description is like that of working memory in various ways. We can see that clearly by switching the term "computer" with "working memory" in this quotation. By doing so, every claim continues to be true. If fact, he could just as equally be describing working memory:

(1) The behavior of working memory at any moment is determined by the symbols which he is observing, and his "state of mind" at that moment. (2) We may suppose that there is a bound B to the number of symbols or squares which working memory can observe at one moment. (3) If working memory wishes to observe more, it must use successive observations. (4) We will also suppose that the number of states of mind which need be taken into account is finite. (5) More complicated states of mind can be avoided by writing more symbols on the storage components of working memory.

This paraphrasing in Turing's words would not work were we to use "predictive processing" or "T1 processes." The statements would then be false. It seems like Newell's (1980) architecture is adequate in many ways to serve as a model of working memory whereas predictive processing is not.

T2 processes are those that load heavily on working memory, and thus, are likely executed by a system like Newell's architecture. On the other hand, of course working memory processes could only be restating what T1 processes have already arrived at. This possibility is show, for instance, by the computerized version of the Wason selection task (Evans, 1996). Also, it is allowed by definition that T1 processes might load weakly in working memory. A possible option is that for us to consider a token process as T2, conclusions to such problem must be reached only after the use of such distinct computational methods of Newell's architecture. That is, something must be found in heuristic search (see Newell and Simon, 1976) which was not found in predictive processing in order for a process to be considered T2.

A stronger hypothesis is that human working memory is literally a classical architecture simulated by the brain, or a component of such, and also that its executive functions are literally the application of operators as in Newell's symbol systems. This would be a problem if the whole mind was said to work in this fashion. But in this case it is only T2 processes that are realized by such architecture, which are a very limited class of mental functions. A weaker hypothesis would be that T2 processes have similar features to that of classical architectures, but there is no metaphysical commitment implied. Either one does the job of solving the unity problem for the working memory feature.

SPEED

Time is valuable for the effectiveness of T2 processes. As we know, the first computers ever invented were much slower than the ones we have today. Thus, having the best hardware for processing in a given way is tantamount to fast processing. In contrast, the brain and the body are a network of cells, so simulating a classical architecture is not what is natural of it.

That we organize our goals explicitly and that we investigate possibilities better than other animals seems to be true. It also seems to be true generally that we are better at T2 processing than other animals are. For instance, no other animal knows

what mathematics is, and are not able to explore consequences of axioms (although, of course, they can know about quantities). So it seems to be true that T2 processes are an unnatural function of the mammal brain. If we follow the hypothesis that T2 processing is the result of operations of simulated classical architecture in the brain, then it would make sense to assume that such simulated architecture does not have the appropriate hardware conditions to perform with the speed of computers built just for such functions.

Following the hypothesis, we should want to claim that classical architectures are slower than predictive processing architectures. We do not have computers with hardware in the forms of networks, much less ones that compute probabilistically in such hardware. We only have simulations. Anyhow, we do have reason to believe that networks are faster. As Fodor and Pylyshyn (1988, p. 35) comment: "in the time it takes people to carry out many of the tasks at which they are fluent (like recognizing a word or a picture, either of which may require considerably less than a second) a serial neurally instantiated program would only be able to carry out about 100 instructions many thousands—or even millions—of instructions in present-day computers (if they can be done at all)."

Of course, by defending classical architectures, Fodor and Pylyshyn (1988, p. 39) go on to argue that these are issues of the implementation level. In fact, that any speed issue should be so. "The moral is that the absolute speed of a process is a property par excellence of its implementation." If this is the case, then apparently, we have two reasons to think that T2 processes in the current developing framework would be slower. First because network processing will tend to be faster in comparison and second because, as physiology teaches us, the brain does not have the appropriate hardware for the implementation of a fast classical architecture. However, although Fodor and Pylyshyn (1988) were correct that implementation relates to speed, they were wrong in claiming that speed is determined solely by implementation. Using explicit steps over discrete symbols implies certainty over speed. Even in speech we can note how we avoid communicating every explicit step of our thoughts but rather leave open implicit assumptions that are never spoken, in order to maximize speed.

In contrast to favoring certainty over speed, to defend predictive processing's speed, Clark (2016, p. 250) claims "Cheap, fast, world-exploiting action, rather than the pursuit of truth, optimality, or deductive inference, is now the key organizing principle." Surely, a cognitive architecture that attempts to predict incoming information surely must have a recipe for being faster than others. A predictive processing architecture can act faster because any cue captured from the world is readily met with predictions (even if bets) concerning a lot more than the cue itself shows. The predictive processor is always taking certain bets about what the current state of the world implies, losing accuracy in compensation for speed. So it fits nicely with the idea that T1 processing needs to abandon certainty and accuracy for speed, an idea previously developed as quick and dirty heuristics (see Gigerenzer, 1996). Predictions are also quick and dirty and perhaps in a way that makes these properties even more ubiquitous since it spans even perceptual details and not only judgments. Thus, when watching a white scene in a movie, there might be guesses that there are no black and brown pixels in some areas of the screen, even if there are. The quick and dirty guessing thus extends far beyond what traditional frugality theorists (i.e., Gigerenzer, 1996) had been considering.

Another property that allows for fast processing is predictive coding (Rao and Ballard, 1999). By predictive coding we mean specifically the property of these system to consider, from the world, only stimuli which result in greater prediction error. Thus, some stimuli are considered in real-time perception already as irrelevant for the adaptive use of the organism. Precision weighing (see Clark, 2013b) quickly determines the size or effect of the prediction error determining if it is eliminated or if it needs to further propagate to other areas. Focusing on prediction-relevant stimuli only permits the agent to quickly decide courses of action and to select amongst possible affordances (see Gibson, 1979). T1 processing can thus be understood as quick predictions emerging from the system's first considerations of these errors.

As Clark explains embodied prediction, the agent is always tuned to environmental cues which can quickly help the system decide between affordances. The predictive architecture provides means for quicker selection, "allowing time-pressed animals to partially 'pre-compute' multiple possible actions, any one of which can then be selected and deployed at short notice and with minimal further processing." (Clark, 2016, p. 180). In the cases studied by DPT, mostly of people taking reasoning and decision-making tests, this quickness of action comes in the form not of body movements but of simplistic hypothesis quickly springing to mind. Such hypothesis come to mind quickly because of the probabilistic relations they bear with the input. So we can even start to ponder about the basis of accessibility, which worries Kahneman (2002, p. 456) "much is known about the determinants of accessibility, but there is no general theoretical account of accessibility and no prospect of one emerging soon." Accessible content could be understood as the higher values in probability density distributions of a generative model related to the range of possible responses to a given task. The more given values have been used to reduce prediction error in the (evolutionary and developmental) past the more the content will be accessible.

CONCLUSION

I have argued that many T1 core features are necessary features of a predictive processing architecture, whereas classical architectures cannot be done away with and its mechanisms are functionally presupposed in T2 processes. Taken together, various reasons were given for this hypothesis to hold in relation to representational format, automaticity, working memory and speed. This endeavor is meant to solve the unity problem as posed by Samuels (2009). It is of central importance to understand why there are two property clusters of processing features for reasoning and decision making and DPT needs further theoretical development to defend it from recent attacks (see Osman, 2004, 2013; Keren and Schul, 2009;

Kruglanski and Gigerenzer, 2011; Keren, 2013; Kruglanski, 2013; Melnikoff and Bargh, 2018).

For the future, we need other associated projects to test this hypothesis. From psychology we need to see if evidence does hold for T1 answers as stemming from predictive processing and T2 as following a classical architecture. From artificial intelligence we need to see that such a hybrid is useful and feasible. Neuroscience should be able to detect different types of related mechanisms in classical reasoning, judgment and decision-making tasks, not too much in brain region but most likely in action potentials. Altogether, this is a hypothesis that needs to be investigated, rather than taken as correct. Although the arguments hold, only empirical evidence will show if it is true or false.

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

The author confirms being the sole contributor of this work, and has approved it for publication.

FUNDING

This work was originally funded by CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*).

ACKNOWLEDGMENTS

Other researchers have contributed to this manuscript, and I am very grateful, specifically, André Abath, Keith Frankish, Ernesto Perini F. M. Santos, Marco Aurélio Sousa Alves, Richard Samuels, and Daniel M. R. Silva.

REFERENCES

Baddeley, A. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359

Bellini-Leite, S. C. (2017). "The revisionist strategy in cognitive science," in *Cognitive science: Recent Advances And Recurring Problems*, eds F. Adams, O. Pessoa, and J. E. Kogler (Wilmington, DE: Vernon press), 265–276.

Bellini-Leite, S. C. (2018). Dual process theory: systems, types, minds, modes, kinds or metaphors? A critical review. *Rev. Philos. Psychol.* 9, 213–225. doi: 10.1007/s13164-017-0376-x

Bellini-Leite, S. C. (2021). How sentience relates to dual process distinctions of consciousness. *J. Conscious. Stud.* 28, 121–129.

Bellini-Leite, S. C., and Frankish, K. (2020). "Bounded rationality and dual systems," in *Routledge Handbook of Bounded Rationality* ed. R. Viale (London: Routledge), 207–216. doi: 10.4324/9781315658353-13

Clark, A. (2013a). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Clark, A. (2013b). Are we predictive engines? Perils, prospects, and the puzzle of the porous perceiver. *Behav. Brain Sci.* 36, 233–253. doi: 10.1017/s0140525x12002440

Clark, A. (2015). "Embodied Prediction," in *Open MIND*, Vol. 7, eds T. Metzinger and J. M. Windt (Frankfurt: MIND Group).

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, And The Embodied Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780190217013.001.0001

De Neys, W. (ed.) (2017). *Dual Process Theory 2.0*. New York, NY: Routledge. doi: 10.4324/9781315204550

Egner, T., Monti, J., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *J. Neurosci.* 30, 16601–16608. doi: 10.1523/JNEUROSCI.2770-10.2010

Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *J. Pers. Soc. Psychol.* 71, 390–405. doi: 10.1037/0022-3514.71.2.390

Evans, J. S. B. (1996). Deciding before you think: relevance and reasoning in the selection task. *Br. J. Psychol.* 87, 223–240.

Evans, J. (2008). Dual-processing accounts of reasoning. judgment, and social cognition. *Ann. Rev. Psychol.* 59, 255–278. doi: 10.1146/annurev.psych.59.103006.093629

Evans, J., and Stanovich, K. (2013). Dual-Process theories of higher cognition: advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi: 10.1177/1745691612460685

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71. doi: 10.1016/0010-0277(88)90031-5

Frankish, K. (2004). *Supermind and Supramind*. Cambridge: Cambridge UniversityPress. doi: 10.1017/CBO9780511487507

Frankish, K. (2009). "Systems and levels: Dual-system theories and the personal–subpersonal distinction," in *Two Minds: Dual Process and Beyond*, eds J. Evans and K. Frankish (Oxford: Oxford University Press), 89–107. doi: 10.1093/acprof:oso/9780199230167.003.0004

Frankish, K., and Evans, J. (2009). "The duality of mind: An historical perspective," in *Two Minds: Dual Processes And Beyond*, eds J. Evans and K. Frankish (Oxford: Oxford University Press), 1–29. doi: 10.1093/acprof:oso/9780199230167.003.0001

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception: Classic Edition*. Hove: Psychology Press.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to kahneman and tversky. *Psychol. Rev.* 103, 592–596. doi: 10.1037/0033-295X.103.3.592

Glenberg, A. (1999). "Why mental models must be embodied," in *Advances In Psychology*, Vol. 128, ed. G. Rickheit (North-Holland: Elsevier), 77–90. doi: 10.1016/S0166-4115(99)80048-X

Harnad, S. (1990). The symbol grounding problem. *Phys. D Nonlinear Phenom.* 42, 335–346. doi: 10.1016/0167-2789(90)90087-6

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001

Kahneman, D., Slovic, S. P., Slovic, P., and Tversky, A. (eds). (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kahneman, D. (2002). Maps of bounded rationality: a perspective on intuitive judgment and choice. *Nobel Prize Lect.* 8, 449–489.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Strauss, Giroux.

Kahneman, D., and Frederick, S. (2002). "Representativeness revisited: attribute substitution in intuitive judgment," in *Heuristics and Biases: The Psychology of Intuitive Judgment*, eds T. Gilovich, D. Griffin, and D. Kahneman (Cambridge: Cambridge University Press), 103–119. doi: 10.1017/CBO9780511808098.004

Keren, G. (2013). A tale of two systems: a scientific advance or a theoretical stone soup? Commentary on evans & stanovich (2013). *Perspect. Psychol. Sci.* 8, 257–262. doi: 10.1177/1745691613483474

Keren, G., and Schul, Y. (2009). Two is not always better than one: a critical evaluation of two-system theories. *Perspect. Psychol. Sci.* 4, 533–550. doi: 10.1111/j.1745-6924.2009.01164.x

Kruglanski, A. W. (2013). Only one? The default interventionist perspective as a unimodel-Commentary on Evans & Stanovich (2013). *Perspect. Psychol. Sci.* 8, 242–247. doi: 10.1177/1745691613483477

Kruglanski, A. W., and Gigerenzer, G. (2011). Intuitive and deliberative judgments are based on common principles. *Psychol. Rev.* 118, 97–109. doi: 10.1037/a0020762

Lakatos, I., Worrall, J., and Currie, G. (1979). The methodology of scientific research programmes: philosophical papers. *Br. J. Philos. Sci.* 30.

Melnikoff, D. E., and Bargh, J. A. (2018). The mythical number two. *Trends Cogn. Sci.* 22, 280–293. doi: 10.1016/j.tics.2018.02.001

Newell, A. (1980). Physical symbol systems. *Cogn. Sci.* 4, 135–183. doi: 10.1207/s15516709cog0402_2

Newell, A., and Simon, H. (1963). "GPS: a program that simulates human thought," in *Computers & Thought*, eds E. Feigenbaum and J. Feldman (New York, NY: McGraw-Hill Book Company), 279–293.

Newell, A., and Simon, H. (1976). Computer science as empirical inquiry: symbols and search. *ACM Commun.* 19, 113–126. doi: 10.1145/360018.360022

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychon. Bull. Rev.* 11, 988–1010. doi: 10.3758/BF03196730

Osman, M. (2013). A case study: dual-process theories of higher cognition–commentary on evans & stanovich (2013). *Perspect. Psychol. Sci.* 8, 248–252. doi: 10.1177/1745691613483475

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Samuels, R. (2009). "The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds," in *Two Minds: Dual Process and Beyond*, eds J. Evans and K. Frankish (Oxford: Oxford University Press), 129–146. doi: 10.1093/acprof:oso/9780199230167.003.0006

Schneider, W., and Chein, J. (2003). Controlled and automatic processing: behavior, theory, and biological processing. *Cogn. Sci.* 27, 525–559. doi: 10.1207/s15516709cog2703_8

Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychol. Bull.* 119, 3–22. doi: 10.1037/0033-2909.119.1.3

Stanovich, K. E., and Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind Society* 11, 3–13. doi: 10.1007/s11299-011-0093-6

Turing, A. (1936). On computable numbers with an application to the entscheidungs problem. *Proc. London Math. Soc. 2* 42, 544–546.

Von Neumann, J. (1945). The first draft report on the EDVAC. *IEEE Ann. Hist. Comput.* 15, 27–75.1. doi: 10.1109/85.238389

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.

Copyright © 2022 Bellini-Leite. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fpsyg.2022.807261

A Generative View of Rationality and Growing Awareness†

Teppo Felin[1,2]* *and Jan Koenderink*[3,4]

[1] Jon M. Huntsman School of Business, Utah State University, Logan, UT, United States, [2] Saïd Business School, University of Oxford, Oxford, United Kingdom, [3] Department of Experimental Psychology, Katholieke Universiteit Leuven, Leuven, Belgium, [4] Department of Experimental Psychology, Utrecht University, Utrecht, Netherlands

In this paper we contrast bounded and ecological rationality with a proposed alternative, generative rationality. Ecological approaches to rationality build on the idea of humans as "intuitive statisticians" while we argue for a more generative conception of humans as "probing organisms." We first highlight how ecological rationality's focus on cues and statistics is problematic for two reasons: (a) the problem of cue salience, and (b) the problem of cue uncertainty. We highlight these problems by revisiting the statistical and cue-based logic that underlies ecological rationality, which originate from the *misapplication* of concepts in psychophysics (e.g., signal detection, just-noticeable-differences). We then work through the most popular experimental task in the ecological rationality literature—the city size task—to illustrate how psychophysical assumptions have informally been linked to ecological rationality. After highlighting these problems, we contrast ecological rationality with a proposed alternative, generative rationality. Generative rationality builds on biology—in contrast to ecological rationality's focus on statistics. We argue that in uncertain environments cues are rarely given or available for statistical processing. Therefore we focus on the psychogenesis of awareness rather than psychophysics of cues. For any agent or organism, environments "teem" with indefinite cues, meanings and potential objects, the salience or relevance of which is scarcely obvious based on their statistical or physical properties. We focus on organism-specificity and the organism-directed probing that shapes awareness and perception. Cues in teeming environments are noticed when they serve as cues-for-something, requiring what might be called a "cue-to-clue" transformation. In this sense, awareness toward a cue or cues is actively "grown." We thus argue that perception might more productively be seen as the *presentation* of cues and objects rather than their *representation*. This generative approach not only applies to relatively mundane organism (including human) interactions with their environments—as well as organism-object relationships and their embodied nature—but also has significant implications for understanding the emergence of novelty in economic settings. We conclude with a discussion of how our arguments link with—but modify—Herbert Simon's popular "scissors" metaphor, as it applies to bounded rationality and its implications for decision making in uncertain, teeming environments.

Keywords: perception, cognition, ecological rationality, psychophysics, biology, uncertainty, decision making, behavioral economics

OPEN ACCESS

Edited by:
Shaun Gallagher,
University of Memphis, United States

Reviewed by:
Sergei Gepshtein,
Salk Institute for Biological Studies, United States
Elisabet Tubau,
University of Barcelona, Spain

***Correspondence:**
Teppo Felin
teppo.felin@sbs.ox.ac.uk

†*We appreciate comments from Andrea van Doorn, Cecilia Heyes, Colin Mayer, Denis Noble, Dennis Snower, Emma Felin, George Ellis, Joachim Krueger, Paul Collier, and Ruth Chang*

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *01 November 2021*
Accepted: *16 February 2022*
Published: *07 April 2022*

Citation:
Felin T and Koenderink J (2022) A Generative View of Rationality and Growing Awareness. Front. Psychol. 13:807261. doi: 10.3389/fpsyg.2022.807261

INTRODUCTION

Recent theories of bounded and ecological rationality focus on the structural and statistical properties of environments. Humans are seen as *intuitive statisticians* who process their surroundings by relying on a "statistical toolbox" of heuristics (Peterson and Beach, 1967; Gigerenzer and Murray, 1987; Cosmides and Tooby, 1996; Goldstein et al., 2001; Scheibehenne et al., 2013; Meder and Gigerenzer, 2014; Gigerenzer and Marewski, 2015; Gigerenzer, 2020).

Over the past decades, the concept of a *cue* has become foundational to this literature (for a review, see Gigerenzer and Gaissmaier, 2011; also see Gigerenzer and Goldstein, 1996; Karelaia and Hogarth, 2008; Marewski et al., 2010). Cues are essentially seen as data or "pieces of information in the environment" (Kozyreva and Hertwig, 2021, p. 1526). Cues represent the data and information that needs to be processed to attain rational judgments and outcomes (Gigerenzer, 2020; Hertwig et al., 2021). This focus on cues has lent itself to applying (or as we argue, misapplying) a whole host of assumptions and methods from psychophysics and statistics to understand and study rationality. The methods used to highlight the idea of humans as intuitive statisticians include various approaches such as random sampling, signal detection, stimulus thresholds, lens model statistics, just-noticeable-differences, Neyman–Pearson statistics, representative design, and Bayesian inference (e.g., Dhami et al., 2004; Hogarth, 2005; Pleskac, 2007; Karelaia and Hogarth, 2008; Hertwig and Pleskac, 2010; Todd and Gigerenzer, 2012; Luan et al., 2014; Gershman et al., 2015; Gigerenzer and Marewski, 2015; Feldman, 2017; Rahnev and Denison, 2018; Szollosi and Newell, 2020).

In this paper we argue for a generative approach to rationality, one that focuses on humans as probing organisms rather than intuitive statisticians. While the ecological rationality literature is strongly anchored on statistics, we build on biology. In the paper we first discuss two problems with the ecological rationality literature's focus on cues and humans as intuitive statisticians: (a) the problem of cue salience, and (b) the problem of cue uncertainty. The emphasis on the physical and statistical aspects of cues—as data to be processed—misses the fact that the relevant cues may lack these qualities. The focus on statistically or physically measurable factors—concepts imported but misapplied from psychophysics: size, intensity, frequency, repetition and so forth (cf. Gigerenzer and Gaissmaier, 2011)—treats cues as predefined or given. In the paper we work through the most popular and frequently discussed experiment in ecological rationality—the city size task—and highlight how psychophysical intuition has been extended to the context of rationality in problematic fashion. We argue that "ready-made" conceptions of environments cannot deal with the question of how cues emerge in the first place, as illustrated by situations where relevant or critical cues are small, non-obvious or hidden.[1] This problem is exacerbated in real-world "teeming" environments, which differ wildly from the environments used in experimental tasks. We revisit the foundations of these arguments—linking to early work by Fechner and others—and the problem of how one might "grow" a cue.

In response to existing work, we develop a generative alternative to rationality, an approach that addresses the aforementioned problems of cue noticeability, relevance and novelty. We argue that environments are organism-specific and that organism-directed search plays a critical role in shaping cue salience. In real-world situations and tasks—particularly in teeming environments—the relevant cues and environmental structure are rarely if ever predefined, given or obvious. Rather, cues are noticed when they serve as cues-for-something (Koenderink, 2011, 2012; cf. Chater et al., 2018)—that is, clues or evidence. In situations of judgment and rationality, noticing the relevant cues has more to do with organism-specific, generative factors rather than bottom-up statistical ones (like thresholds or signal detection). We discuss the need to understand what might be called the cue-to-clue transformation, that is, how organism-specific, top-down factors play a role in transforming "raw" optical structure and latent or dormant cues into clues-for-something. In essence, we provide an alternative theory of noticing—a generative approach to understanding salience, cue "growth" and detectability. We link these arguments to bounded rationality and decision making in uncertain environments, and conclude with a reconceptualization of Herbert Simon's popular "scissors" metaphor.

ECOLOGICAL RATIONALITY, CUES AND STATISTICS: A BRIEF REVIEW

The concept of an environmental *cue* is a foundational unit of analysis within the bounded and ecological rationality literatures (Gigerenzer and Goldstein, 1996; Karelaia and Hogarth, 2008; Luan et al., 2019; Kozyreva and Hertwig, 2021). These literatures build on the premise that environments, as Todd and Gigerenzer (2020, p. 15; also see Hertwig et al., 2021) recently summarize this argument,

> "...can be characterized by distributions of cues and cue values (how many there are, what range of values they can take, etc.), cue validities (how often a cue indicates appropriate decisions), redundancies (inter-cue correlations), and discrimination rates (how often a particular cue distinguishes between alternatives, regardless of its accuracy)."[2]

The focus on cues—as information and data (Todd and Gigerenzer, 2007; Luan et al., 2011)—has enabled scholars to statistically measure and specify the properties and structure of environments. The literature argues that "ecological, or environmental, structures are *statistical* and other descriptive

[1]We certainly recognize that each of these three descriptors—small, non-obvious, and hidden—suggest ontologically different properties. Size (for example, something being comparatively "small"), might be said to essentially deal with physics and physical properties, "hidden" points to perception and Gestalt, and "non-obvious" deals with apperceptive processes. We simply highlight that the generalized emphasis on salience due to psychophysical factors misses critical cues that seemingly have none of these qualities (cf. Koenderink, 2012).

[2]The focus on cues and environmental structure are ubiquitous in the ecological rationality literature. For cue-based definitions of environments, see Todd and Gigerenzer (2012), Hertwig et al. (2021), and Kozyreva and Hertwig (2021).

properties that reflect patterns of *information distribution* in an ecology" (Kozyreva and Hertwig, 2021, p. 13, emphasis added). By focusing on cues, scholars have essentially sought to dimensionalize and quantify environments in various ways, by measuring factors such as the number of cues, or their redundancy, addition, growth, distribution, ordering, correlation, integration, combination, weighting and so forth (for a review, see Gigerenzer and Goldstein, 2011; also see Hutchinson and Gigerenzer, 2005; Chater et al., 2018). Importantly, cues are seen as an *a priori,* statistical property of the environment (Hertwig et al., 2021; Kozyreva and Hertwig, 2021).[3]

Ecological rationality starts with the premise—given its roots in bounded rationality (Gigerenzer and Selten, 2001)—that humans are not able to omnisciently or exhaustively capture, process and compute environmental cues, due to human limitations in "computational capabilities" (Simon, 1990; also see Gigerenzer, 2000; Kahneman, 2003; Lieder and Griffiths, 2020). The ecological rationality literature thus builds on the bounded rationality literature which recognizes that exhaustive or perfect representation is not possible (Simon, 1956). Given the lack of time and computational power, humans face varied trade-offs, including the trade-offs between satisficing versus optimality, good enough versus best, and accuracy versus effort (e.g., Gigerenzer and Brighton, 2009).

Given that omniscient processing and rationality is not feasible, the ecological approach to rationality points to (and offers) varied statistical shortcuts for making rational decisions—a so-called "statistical toolbox" of heuristics. Humans are seen as intuitive statisticians who utilize this statistical toolbox to simplify the process of understanding their environments to make rational decisions (Gigerenzer, 1992; cf. Cosmides and Tooby, 2013). This approach begins with the idea that rationality is best achieved by first, as discussed above, understanding the statistical *structure* of environments (Gigerenzer and Gaissmaier, 2011). And this structure, then, needs to be matched with the right shortcut, or statistical tool and heuristic. In other words, the goal is to "[analyze] the information-processing mechanism of the heuristic, the information structures of the environment, and the match between the two" (Todd and Gigerenzer, 2012, p. 5). To illustrate, in some situations it's rational for an agent to randomly sample environmental cues and thus attain a locally optimal choice (Dhami et al., 2004). That is, rather than needing to engage in exhaustive or complete sampling of environmental cues, data and information, scholars have pointed out how in many situations it's rational to sample on a more delimited basis (Hertwig and Pleskac, 2010). The so-called "less-is-more" heuristic suggests that sampling on a more delimited basis can be just as efficient as "perfect" rationality, which wastes cognitive resources (e.g., Katsikopoulos et al., 2010). Heuristics, then, are said to allow organisms and humans to attend to and sample cues on a more delimited and less costly basis, attaining decisions that not only are good enough but perhaps even equivalent to omniscience or unbounded forms of rationality (Todd and Gigerenzer, 2000).

The ecological rationality literature has developed a growing, statistical toolbox of heuristics. This statistical toolbox now includes tools such as random sampling, signal detection, stimulus thresholds, lens model statistics, just-noticeable-difference, Neyman–Pearson statistics, representative design, and Bayesian inference (e.g., Dhami et al., 2004; Hogarth, 2005; Pleskac, 2007; Karelaia and Hogarth, 2008; Hertwig and Pleskac, 2010; Todd and Gigerenzer, 2012; Luan et al., 2014; Pleskac and Hertwig, 2014; Gershman et al., 2015; Gigerenzer and Marewski, 2015; Feldman, 2017; Rahnev and Denison, 2018; Szollosi and Newell, 2020). And these statistical tools can directly be mapped onto various named heuristics (see Todd and Gigerenzer, 2012). The overall focus on humans as "intuitive statisticians" has been a central pillar of this literature for a number of decades (see Gigerenzer, 1992; Cosmides and Tooby, 1996). And this idea of course is echoed in earlier work as well. For example, Peterson and Beach (1967, p. 43) argued "experiments that have compared human inferences with those of statistical man show that the normative model provides a good first approximation for a psychological theory of inference." This conception of the human statistician has enthusiastically been endorsed in ongoing work (Meder and Gigerenzer, 2014, p. 130; Hertwig et al., 2018).

Before proceeding, we might note that ecological approaches explicitly argue that these statistical tools and heuristics are the result of long-run, evolutionary adaptations to changing environments. As put by Gigerenzer (2008, p. 20), "the adaptive toolbox is a Darwinian-inspired theory that conceives of the mind as a modular system that is composed of heuristics, their building blocks, and evolved capacities." Ecological rationality sees the human mind as composed of varied evolved statistical modules, including modules like Bayesian inference, signal detection, and so forth (see Figure 1, Gigerenzer, 1992, p. 336). Ecological rationality builds on a broader program of research in evolutionary psychology, where "the brain is a computer. . .designed by natural selection"—and, "if you want to describe its operation in a way that captures its evolved function, you need to think of it as composed of programs that process information" (Cosmides and Tooby, 2013, p. 203). This emphasis on computation and statistical processing provides the ongoing foundation for the ecological rationality literature (Gigerenzer, 2020), as well as generalized models of cognition and rationality (e.g., Gershman et al., 2015; Lieder and Griffiths, 2020).

CUES AND ENVIRONMENTS: TWO PROBLEMS

While the notion of humans as intuitive statisticians—and the statistical toolbox of heuristics—has offered useful insights, this literature is overly-reliant on the assumption that environments can be statistically captured, or that the relevant cues can be predefined. As we will show, approaches that treat cues as given

[3]The focus on cues and their physical characteristics is equally important in other literatures within psychology. For example, person-situation research focuses heavily on cues by arguing that "the situation consists of objectively quantifiable stimuli called cues" (Rauthmann and Sherman, 2020, p. 473). Cues in this literature are similarly defined as the "physical or objective elements that comprise the environment," and again, the literature further argues that "they [the cues] can be objectively measured and quantified" Rauthmann et al. (2014, p. 679; cf. Todd and Gigerenzer, 2020).

and environments as, essentially, "ready-made," have not fully come to terms with where cues come from in the first place and the "teeming" nature of real decision environments. To illustrate these points, we discuss how ecological approaches to rationality suffer from two specific problems: (a) the problem of cue salience, and (b) the problem of cue uncertainty. We discuss these two problems by revisiting existing experimental work and by linking the foundations of ecological rationality to psychophysics. Thereafter we propose an alternative, "generative" approach to rationality.

Note that our criticisms here are *not* meant to offer a wholesale challenge to the contributions of the ecological rationality literature. Instead, our efforts might be seen as setting boundaries for the generality of ecological approaches that focus on cues and the idea of humans as "intuitive statisticians." More importantly, our discussion of these problems is meant to provide a jumping-off point and rationale for developing an alternative approach to rationality, one that is focused on organism-specific and directed, generative factors which are essential for understanding rationality in uncertain environments.

The Problem of Cue Salience

One way to recast ecological rationality is to point out how its underlying "theory of noticing" is focused on the quantitative or statistical properties of cues—factors such as the *amount, intensity and distribution* of cues (see Gigerenzer and Gaissmaier, 2011). This is perhaps most evident in the emphasis on "*stimulus detection* as intuitive statistics" (Gigerenzer, 1992). Stimulus detection of course implies knowing what in fact counts as a stimulus. Importantly, in the existing literature the specific mechanism of detecting the stimulus is focused on the amount or "size" of a particular cue. To put this informally, a predefined and given cue is perceived or recognized when there is "lots" of it.[4] As recently summarized by Kozyreva and Hertwig (2021, p. 1531), "sample size itself becomes an important environmental structure." In essence, the underlying theory of noticing—in the simplest of terms (though we add nuance below)—is that noticing is dependent on the proverbial loudness, amount or size of cues: factors that can be physically and statistically measured.

This focus on the statistical and physical aspects of cues—sometimes called inputs, stimuli or data (Gigerenzer, 2020)—builds on a long historical tradition in psychology. The foundations of this work were laid by scholars such as Ernst Weber and Gustav Fechner in psychophysics (Boring, 1942; Wixted, 2020). We revisit the central elements of this work. Doing so is important because these building blocks of psychophysics are the *de facto* foundation of the ecological rationality literature (e.g., Gigerenzer, 1992, 2020; Luan et al., 2011, 2014). In other words, roughly the same mechanisms of salience—the underlying theory of noticing—are employed in both literatures. This underlying foundation of signal detection, just-noticeable-differences and stimulus thresholds was essential for the early work in ecological rationality (Gigerenzer, 1991, 1992) and continues to be centerstage to this day (see Karelaia and Hogarth, 2008; Luan et al., 2011, 2019; Gigerenzer, 2020). However, we argue that these psychophysical foundations have been wrongly applied in the context of ecological rationality.

The goal of early work in psychophysics was to experimentally study if and when humans notice—and become aware of—a *given, prespecified* cue or stimulus (Boring, 1942). In the earliest formal experiments, Gustav Fechner introduced human subjects to a single stimulus—an auditory, haptic or visual one—and proceeded to see *when* the focal stimulus became salient. Fechner's approach was to gradually, in small increments, increase the amount of the focal cue and then to see when subjects noticed and became aware of it. His underlying approach, as he put it in his classic *Elements of Psychophysics*, was to start from "zero" and then to essentially "grow" awareness toward particular physical cues and stimuli. As Fechner (1860, p. 58) put it, stimuli "might be seen as incrementally grown from zero" (in the original German: "aus positiven Zuwüchsen von Null an erwachsen angesehen werden" – our translation).

This early work in psychophysics sought to provide a scientific basis for psychology, a way to rigorously quantify and statistically measure physical stimuli and cues in environments. One aim of this approach was to make psychology more like the hard sciences, like physics, where the amounts and quantities of cues or stimuli served the equivalent of mass and force. Awareness was essentially seen as a function of the metaphorical mass of something—the amount, intensity and frequency of the cue. Fechner's work became the basis of signal detection theory, a ubiquitously important theory that offered a statistical and quantitative basis for how increased intensities or amounts of stimuli were the central variable of interest for understanding perception and awareness (see Link, 1994; Wixted, 2020). This work also became the basis of theories of signal detectability (TannerJr., and Swets, 1954; also see Peterson and Birdsall, 1953), which have also had a strong influence on ecological rationality (e.g., Gigerenzer, 2000).[5]

This logic continues to pervade behavioral economics more broadly, where salience is seen as the "the property of a stimulus that draws attention bottom up" (Bordalo et al., 2021, p. 6). Or as put by Kahneman (2003, p. 1453), "the impressions that become accessible in any particular situation are mainly determined, of course, by the *actual* properties of the object of judgment," and "physical salience [of objects and environments] determines accessibility." Thus the emphasis is on predefined cues and whether humans appropriately process them based on their physical and statistical characteristics.

Early work in psychophysics—specifically the work of Ernst Weber—also looked at when humans noticed *comparative* differences between two cues or stimuli (Weber, 1834; for a review, see Boring, 1942; Algom, 2021). Here the premise again was to start from zero: a "zero" difference between two cues (e.g., optical stimuli, lifted weights, or sounds), and then to

[4] Arguably, the most common reaction of a biological organism to structures where there is "lots" of a cue is probably to ignore these cues, since they can be taken for granted. In this sense, the most useful cues are necessarily rare.

[5] Our goal by no means is to dismiss Fechner's important and voluminous work. We merely point out how the underlying logic of Fechner's *Elements of Psychophysics*—where cues are taken as givens (and salience is a function of statistical or physical qualities)—has been problematically applied in the context of ecological rationality.

incrementally increase the brightness, weight, or loudness of one of the stimuli to see when the comparative difference was noticed. As summarized by Gigerenzer (1992, p. 339), "detection occurs only if the effect a stimulus has on the nervous system exceeds a certain threshold value, the 'absolute threshold.' Detecting a difference (discrimination) between two stimuli occurs if the excitation from one exceeds that of the other by an amount greater than a 'differential threshold'." This logic provides much of the foundational intuition behind ecological rationality. Scholars have debated whether absolute or relative differences matter more within the context of judgment and decision making (e.g., Hau et al., 2010; Hertwig and Pleskac, 2010). But the underlying foundations of Weber's pioneering work—concepts such as just-noticeable-differences—continue to be center stage in ecological rationality literature (e.g., Pleskac and Busemeyer, 2010; Luan et al., 2014; Gigerenzer, 2020).

Importantly, Weber and Fechner's work on stimulus comparison and difference detection had been extended into the domain of judgment and decision making earlier, by scholars like Thurstone (1927). Thurstone developed his so-called laws of comparative judgment and discrimination, and these Thurstonian notions were in turn further extended by decision theorist Duncan Luce into axioms of choice and decision making, with a strong focus on the representation of signals and environments (Luce, 1963, 1977; also see Dawes and Corrigan, 1974). This work is also central to ecological rationality, particularly arguments about the representational nature of perception and rationality (e.g., Gigerenzer, 1991; Juslin and Olsson, 1997; Luan et al., 2014). But as we will discuss, these psychophysical foundations have been misapplied by the ecological rationality literature.

Now, these psychophysical foundations are clearly important for understanding certain aspects of perception.[6] However, the central question here is whether the underlying statistical architecture of psychophysics—focused on noticing a stimulus as a function of its "amount" (such as frequency, intensity, size)—is sufficiently general for handling varied questions and situations of rationality. For example, how might we account for situations where the relevant cues have *none* of the traditional statistical or physical characteristics of salience? Also, the underlying logic psychophysics was to introduce *one* stimulus, and to identify when it was salient (based on amount), or to compare the relative salience between two stimuli (from a baseline of zero). But most environmental settings "teem" with indefinite cues and stimuli. As we highlight below, the focus on the amount—whether absolute or relative—does not generalize to situations where amounts simply are not relevant. A more central question—particularly in environments that teem with indefinite cues and stimuli—is how one might become aware of the *relevant* cues, amongst varied potential distractions and noise. The logic of incrementally growing or increasing the intensity of a given cue—or comparing two cues—does not translate to these types of settings.

The default starting point or initial condition of psychophysics might, in effect, be seen as a proverbial dark or silent room, where the intensity of a focal stimulus is gradually increased, dialed up and "grown"—to establish threshold levels of awareness or signal detection. While this of course is important (and certainly relevant for situations of visual or auditory impairment), and allows for scientifically clean and controlled conditions for explaining a highly particular form of awareness (when organisms "notice" something, or do not), it scarcely mimics many of the complex situations and teeming environments that humans and other organisms encounter and find themselves in. The idealized starting point of a metaphorical dark or silent room of psychophysics might instead be replaced by a different metaphor. A better default metaphor might be captured by a human standing midday at Times Square in New York, encountering indefinite visual and auditory stimuli, bombarded by innumerable sounds and sights. This real-life "Wimmelbild" better captures the problem faced by a decision maker in an uncertain environment. This teeming visual scene, like any other, is full of "signals" and "affordances" (Krebs and Dawkins, 1984; Koenderink, 2012) which cannot be accounted for by any kind of generic focus on the physical or statistical aspects of the scene.[7]

Now, while it has not meaningfully been integrated into the ecological rationality literature, there is of course a larger literature in the domain of perception that has wrestled with how humans process cues and information in "busy," multisensory environments. This literature has focused on such questions as how we might bind, combine or separate particular cues and sensory inputs in visual scenes and environments (e.g., Treisman and Gelade, 1980; Landy et al., 1995; Noppeney, 2021; Wolfe, 2021). While this literature is important, it also builds on the aforementioned psychophysical premise where cues and features are given, and salience is driven by physical or statistical factors, specifically the *relationships* amongst the cues (for example: the spatial distance of cues, cue similarity or difference). This research presents experimental subjects with varied arrays of visual cues or scenes and looks at how and whether humans process them veridically. While this work certainly has its place (particularly in contexts of establishing sensory deficiencies), it builds on an "all-seeing" conception of perception (Koenderink, 2014; Hoffman et al., 2015; Felin et al., 2017). Thus we think different perceptual foundations are needed for understanding judgment and rationality in uncertain environments (cf. Chater et al., 2018).

The so-called "cocktail party effect" or cocktail party problem (Cherry, 1953; Shinn-Cunningham, 2008) offers a somewhat better instantiation of the types of teeming environments

[6]We have focused on the early foundations of psychophysics, though arguably a broader conception of "modern" psychophysics would include many important contributions and additions (Kingdom and Prins, 2016; also see Lu and Dosher, 2013). Our goal is not to review this very large literature. Rather, we simply seek to point out how some of the key aspects of psychophysics (the emphasis on the statistical and physical nature of cues) have been misapplied in the context of ecological rationality.

[7]For example, standing at Times Square we might observe cues (or signals) like people walking from a certain direction with shopping bags and thus make inferences about a certain shop in that direction. Or to offer another example, the presence of a "yellow" car might signal a taxi. Visual scenes thus abound with varied signals, affordances and meanings that cannot be accounted for through a strict psychophysical lens that is focused on a purely statistical or physical reading of the environment.

encountered by organisms, humans included. Despite its obvious relevance, the cocktail problem surprisingly has not been cited or addressed in the ecological rationality literature. The cocktail problem is the very relatable problem of how one focuses on a particular conversation or auditory stimulus in a noise-filled environment filled with distractions. In this type of teeming situation, cue salience is *not* given by any form of statistical aspects of the cues themselves (e.g., how loud a stimulus is). We might of course highlight which cues are, in a relative psychophysical sense louder and thus seemingly more salient than others. But here the question is rather about selecting and picking out a relevant conversation or cue. In these situations, salience is given by deliberate, top-down mechanisms on the part of subjects. This literature thus focuses on factors such as motivation and interest as drivers of cue salience, in an environment filled with other cues and distractions. Another parallel is the literature focused on "motivated perception," "motivated seeing" and "wishful seeing" (Bruner and Goodman, 1947; Balcetis and Dunning, 2006; Leong et al., 2019). However, these literatures have largely focused on the biased or self-delusional nature of hoping or wanting to see and find something (a form of confirmation bias), rather than rationality-related considerations and concerns.

In all, we might summarize ecological rationality as follows. Ecological rationality treats the world as a dataset to be processed, where the cues and data are given. The role of the human, as intuitive statistician, is to efficiently process these cues using heuristics and associated statistical tools. However, what is lost in these abstractions is the often messy and critical process of deciding what represents a cue in the first place, or how a potentially "small" or hidden (but relevant) cue might somehow be identified or detected. As we discuss (see section below: "Humans as Probing Organisms"), in many situations of judgment and decision making, the relevant cues are scarcely obvious. And importantly, critical cues often do not have any of the traditional psychophysical characteristics of being loud, intense or large. Thus some alternative mechanisms for generating salience are needed.

The Problem of Cue Uncertainty

While the literature on ecological rationality emphasizes that it is squarely focused on decision making in the context of *uncertainty*, yet the most common experiments and tasks are relatively straightforward, even mundane. But as we illustrate next, it's hard to know how the key experiments and examples of ecological rationality actually generalize to novel situations and real-world environments that teem with more radical forms of uncertainty.

To illustrate this problem, consider the most popular experiment and example used by scholars of ecological rationality, the city size comparison task. The city size task is a useful example as it is the focal experiment of the most highly cited academic article in the ecological rationality literature (Gigerenzer and Goldstein, 1996) and also extensively discussed in highly cited books (e.g., Gigerenzer and Todd, 1999). Furthermore, variants of the city size experiment have been done across numerous different contexts over the past three decades, published in various top psychology and cognitive science outlets (e.g., Gigerenzer et al., 1991; Goldstein and Gigerenzer, 2002; Chater et al., 2003; Schooler and Hertwig, 2005; Pohl, 2006; Richter and Späth, 2006; Dougherty et al., 2008; Gigerenzer and Brighton, 2009; Marewski et al., 2010; Hoffrage, 2011; Pachur et al., 2011; Heck and Erdfelder, 2017; Filevich et al., 2019). The city size experiment has also been highlighted as an example of different heuristics, including the recognition heuristic, as well as the less-is-more, tally, and take-the-best heuristics (Goldstein and Gigerenzer, 2008). In all, the city size experiment appears to be the most popular experiment in the ecological rationality literature. Thus it serves as a useful example for us to make our point, namely, that it's hard to see how the arguments about ecological rationality generalize to decision situations and environments that actually feature uncertainty. Furthermore, the city size experiment offers a practical example of how the basic logic of psychophysics—and the associated statistical toolbox—has been imported and translated into the domain of ecological rationality.

In a prototypical city size experiment, subjects are presented with pairs of cities and asked to estimate which of the two cities has a larger population. Subjects might be asked whether, say, Milan versus Modena has more inhabitants (Volz et al., 2006)—or whether Hamburg versus Cologne (Gigerenzer and Goldstein, 1996), Detroit versus Milwaukee (Neth and Gigerenzer, 2015) or San Diego versus San Antonio (Chase et al., 1998) has a larger population. In some experiments subjects are asked to compare cities in their country of residence—or sometimes in a foreign country, or both (see Chater et al., 2003; Pohl, 2006; Richter and Späth, 2006; Gigerenzer and Brighton, 2009; Marewski et al., 2010). Though there are any number of variants to the experiment, the most basic version of the experiment is one where subjects are given a city pairing and simply asked to guess or pick the more populous city. The upshot is that, in a relatively high percentage of instances (higher than chance), the guesses and picks of experimental subjects turn out to be correct.

The popular city size experiment is said to be an example of—amongst other things—the "recognition heuristic." The recognition heuristic is relatively intuitive and simple, defined as follows: "If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion" (Schooler and Hertwig, 2005, pp. 611–612; Volz and Gigerenzer, 2012, p. 3; first defined in these particular terms in Goldstein and Gigerenzer, 2002, p. 76). To put the recognition heuristic in the context of the city size experiment, the idea is that while experimental subjects might not actually know which (say German) city has a larger population, the process of recognizing the name of one of the comparison cities can serve as a useful shortcut or heuristic for making the correct choice. If an American experimental subject is asked whether, say, the city of Munich or Cologne has a larger population, they might draw on other cues and information to enable them to pick the larger city. For example, an experimental subject might have visited Germany and thus be more likely to have flown into Munich, since it is Germany's second largest airport for international flights. Or an experimental subject might be aware of other facts about Munich—for example, that

Oktoberfest is based in Munich. Or they might be aware of the popular German soccer team Bayern Munich.

The idea behind the recognition heuristic is that these cues or "ancillary" bits of information can serve as additional information for recognizing Munich, and therefore arriving at the correct decision about its size relative to Cologne. In some of the experiments, subjects are given some form of additional or related cues, or primed to focus on certain ones (e.g., Gigerenzer and Goldstein, 1996), and in others they are simply given the pair-wise city comparisons and asked to choose the city with the larger population (Pohl, 2006; Todd and Gigerenzer, 2012). But the key point that scholars of ecological rationality hope to make with the city size and related experiments is that a subject's informational recall and memory essentially serve as a shortcut to amass and tally cues to increase the probability that they arrive at the correct decision. While the city size experiment has largely been used to highlight the recognition heuristic, the same experiment has also been used as an example of a host of other heuristics, including heuristics like take-the-best, less-is-more and tally (weighted and unweighted) (see Todd and Gigerenzer, 2012).

The city size experiment and its variants are highly informative as they show how scholars in ecological rationality essentially borrow and translate the logic of psychophysics and cues into the context of heuristics and decision making (Gigerenzer and Gaissmaier, 2011). Cues are treated synonymously with varied, discrete bits of information about the cities. These cues, then, are the metaphorical equivalent of psychophysical "growing a cue from zero"—where information accumulates toward the correct judgment. Again, if a subject is given the task of deciding whether Munich versus Cologne has a larger population, simply knowing about Bayern Munich represents one cue or bit of information that favors its selection. And knowing that Munich hosts the Oktoberfest might serve as another, and so forth. This allows scholars to apply psychophysics-type intuition where the cues are *tallied, weighted, sequenced, and ordered* (and so forth) in different ways (cf. Karelaia and Hogarth, 2008). In other instances, knowing (or being given) some additional facts about a given city is treated in probabilistic fashion (called "probabilistic mental models"), where increased information about a particular city increases one's confidence that it will have a larger population (Gigerenzer et al., 1991).

The full logic of the city size argument, linking psychophysical cues with bounded and ecological rationality, is simulated and worked out by Gigerenzer and Goldstein (1996) in their article titled "Reasoning the fast and frugal way: Models of bounded rationality." They create a computer simulation that features a set of competing heuristics (or algorithms) for estimating the population size of 83 German cities (i.e., all cities in Germany with more than 100,000 inhabitants). The set of cues used to engage in this task includes nine binary (yes/no) bits of information about each city—for example, whether the city has a soccer team in the Bundesliga, whether the city has a university, or whether the city has an intercity train. This same logic has been applied to many other decision environments (for a summary of 27 different ones, see Todd and Gigerenzer, 2012, pp. 203–206). And these findings have not just been simulated, but variants of this approach have been studied with experimental subjects (Goldstein and Gigerenzer, 2002; Dieckmann and Rieskamp, 2007).

Ecological rationality's focus on the city size experiment—and similar tasks—tells us a lot about the approach. It reduces judgment and decision making to a type of signal detection and statistical processing. This is further evident in, for example, applications of the cue-based logic of Brunswik to the city size problem (e.g., Gigerenzer et al., 1991; Hoffrage and Hertwig, 2006). The idea of Bayesian inference is also featured prominently in the city size task and heuristics literature more broadly (Chase et al., 1998; Goldstein and Gigerenzer, 2002; Martignon and Hoffrage, 2002), given the obvious links to signal detection. The idea is that humans don't exhaustively process information, but they use a statistical toolbox—sampling to make their choices. This logic has been applied and extended to many other tasks of comparison and estimation, such as mammal lifespans, car accident rates, the number of species on Galapagos Islands, homelessness, and car mileage (see Todd and Gigerenzer, 2012, pp. 203–206; for a metareview, see Karelaia and Hogarth, 2008).

Now, in principle there is no problem with highlighting how humans might use varied cognitive shortcuts and tricks to enable them to arrive at correct answers about such questions as which city has a larger population, or who (say) won a particular historical match at Wimbledon (Todd and Gigerenzer, 2007). It seems very plausible that humans use shortcuts like this, using ancillary cues and information as a guide. The recognition and associated heuristics undoubtedly can prove useful in the types of situations and experiments constructed by the experimenter.

But our concern is that the most popular examples and experiments of ecological rationality—like the city size experiment—seem to scarcely generalize to other settings, situations and tasks where the relevant cues are not given, and where the right answer simply cannot be looked up. This is a problem, because the focus of ecological rationality is supposed to explicitly be on "situations of uncertainty where an optimal solution is unknown" (Gigerenzer, 2020, p. 1362). The city size and related experiments scarcely are an example of an uncertain situation. While these experiments are highly prominent in the ecological rationality literature, it's extremely hard to see how they might tell us something meaningful about judgment and decisions in truly uncertain, teeming environments.

HUMANS AS PROBING ORGANISMS: A GENERATIVE APPROACH

Next we develop an alternative, generative approach to rationality, in response to some of the aforementioned problems we have identified with ecological rationality. Our generative alternative argues that humans might best be seen as probing organisms rather than intuitive statisticians. While ecological rationality builds on statistics, we build on and extend biological arguments and develop a more generative form of rationality.

We should note that in juxtaposing the aforementioned discussion of humans as intuitive statisticians with our generative alternative, we certainly do not want to offer a wholesale challenge to existing, ecological arguments. The two approaches have their respective benefits, *depending on the task or problem at hand*. We recognize that the logic of intuitive statistics can be applicable to certain settings and for specific types of tasks, where the relevant cues are given and varied forms of statistical processing indeed might be useful. But our proposed alternative might be seen as establishing some much-needed boundaries and contingencies for ecological rationality and related arguments. And more importantly, we hope to highlight how our generative alternative offers a more viable (though admittedly tentative) and biologically grounded option for judgment and decision making, especially in uncertain environments.

Organism-Specific, Teeming Environments

Rather than seeking to first, *a priori*, dimensionalize or quantify environments—based on the redundancy, sample size or the variability of cues (see Gigerenzer and Gaissmaier, 2011, p. 457)—the generative approach starts with the premise that environments are organism-specific. As put by Goldstein (1963, p. 88), "environment first arises from the world only when there is an ordered organism." From our perspective there is no *a priori* environment or environmental structure to be accounted for in the first place—whether statistically or otherwise—without first understanding the organism in question (cf. Schrödinger, 1944; Riedl, 1984; Uexküll, 2010). What an organism is aware of, what becomes salient to it, and what it sees, is organism-dependent. While this might sound like an obvious statement, this organism-dependence—including its downstream consequences for rationality—has not been recognized, as we will illustrate.[8]

Organism-specificity means that an organism's physiology and nature are central to understanding what its environment is (Tinbergen, 1963; Uexküll, 2010). As put by the biologist Uexküll (2010, p. 117), each organism exists in its own surroundings (what he called "Umwelt"), where certain species-specific things are visible and salient to it: "every animal is surrounded with different things, the dog is surrounded by dog things and the dragonfly is surrounded by dragonfly things." At the most basic level, organism-specificity means that organism perception is given by what the organism's visual and sensory organs enable it to see. Sensory organs provide the enabling and constraining mechanism for what the organism can see in its environment, allowing the organism to perceive certain things it encounters, but not others. Certain stimuli, cues, colors, objects are inherently salient to particular organisms. For example, humans can see the visual electromagnetic spectrum between 700 and 400 nm, while bees can detect light between 600 and 300 nm, which includes ultraviolet light (between 400 and 300 nm – not visible to the "naked" eye). Visual scenes and environments therefore look fundamentally different to different species (Cronin et al., 2014; Marshall and Arikawa, 2014). Importantly, this visual heterogeneity applies not only to colors and the electromagnetic spectrum but also to the set of *objects* that are salient and evident to a given species.[9]

As Caves et al. (2019) recently emphasize, treating environments the same across species is a common problem in the sciences, creating significant biases in how we talk about perception, judgment and environments. By treating the environment in homogeneous fashion, we succumb to faulty assumptions like assuming that animals are "doing the math" (or behaving "as if" they did the math: cf. Gigerenzer, 2021), or assuming that different organisms segment cues and stimuli in the very same ways that humans do. These biases have extended into the judgment and decision making literature where scholars have, for example, compared bee cognition with human cognition, suggesting that humans in many instances are less rational than certain animals (Stanovich, 2013). Or in other instances scholars have compared human perception with the "biased" and non-veridical perception of, say, a house fly (Marr, 1982, p. 34; cf. Hoffman et al., 2015, p. 1481). From our perspective, there is no "biased" nor veridical perception of an environment, where one view somehow is more veridical or more/less biased than another. These types of claims succumb to an "all-seeing" view of perception, a view that remains pervasive even though it is untenable (Koenderink, 2014; also see Felin et al., 2017). The problem is that we assume that disparate organisms perceive, or should perceive, the same cues and stimuli in the same way in a given environment—that there is a form of global optimality or omniscience. But this is scarcely the case. Environments are as heterogeneous as the organisms in them.

Now, so far we've emphasized visual heterogeneity across species, highlighting different forms of perception and the indefinite, teeming nature of any environment. But what about visual heterogeneity "within" species? Or put differently, what does any given organism, a human included, see at any particular moment? This moment-by-moment visual heterogeneity *within* a given species or organism is critical for our arguments, as visual metaphors and arguments are the foundation of much of the rationality literature (see Simon, 1956; Kahneman, 2003; Chater et al., 2018). The critical question is, if visual scenes and environments teem with potential objects and things—far beyond any ability to capture them all—then what is salient and visible

[8]We should recognize that while ecological rationality focuses on "statistical properties of the environment that exist *independent* of a person's knowledge" (Kozyreva and Hertwig, 2021, p. 1519, emphasis added), existing work has *rhetorically* (though not substantively) recognized organism-dependence. To illustrate, in his foundational 1956 paper "rational choice and the structure of the environment," Herbert Simon mentions that "we are *not* interested in describing some physically objective world in its totality, but only those aspects of the totality that have relevance as the 'life space' of the organism considered" (Simon, 1956, p. 130, emphasis added). However, the underlying models of search and bounded rationality are organism-independent and general (see Simon, 1980, 1990; for a review, see Felin et al., 2017; also see Chater et al., 2018).

[9]Our organism-centric, biologically informed approach here argues that some measure of generativity is needed to account for the ongoing novelty and heterogeneity we observe all around us, whether in nature or in economic settings. Organism-environment interactions are not just a one-way street, where organisms adapt to their environments over time. Organisms also actively shape their environments. Organisms "are agentive and thus capable of initiating activity by themselves" (Longo et al., 2015, p. 5; cf. Noble, 2015).

to an organism at any given moment? Here the answer is not about what a given organism *can* see (as enabled by the organism's sensory organs, discussed above), nor is it about any form of *ex ante* physical salience (as suggested by psychophysics and ecological rationality). Instead, our focus is on what an organism *might* become aware of at any given moment, amongst indefinite environmental possibilities.

The biologist Uexküll's (2010) notion of a Suchbild (German for "search image") offers a powerful way to think about moment-by-moment awareness. It suggests that, at the simplest level, organism perception is directed toward what it is looking for, whether it be foraging for food or looking for shelter. This Suchbild might be innate (like in the case of the frog looking for flies to eat) or cognitive (in the case of a humans, say, looking for their car keys). Salience is created by the image that the organism has in mind, the object or thing it is searching for (also see Tønnessen, 2018). Organisms fixate on certain visual features or objects—features and objects that essentially serve as the "answers" to their queries (cf. Felin and Kauffman, 2021). What is seen in the environment are the plausible answers or solutions to the organism's search image. For example, when hunting and foraging for crickets, frogs are highly attuned to movement, perceiving motion (of a certain type) rather than perceiving the cricket itself (Ewert, 2004).

In the context of human perception, search images can be seen as a form of question-answer probing that guides visual awareness in our everyday life (Koenderink, 2012; Felin et al., 2017). For example, if I have lost my house keys, I scan my surroundings with a key search image in mind, looking for objects or stimuli that have key-like features. The search image allows me to ignore any number of other items and objects in my surroundings—even ones with psychophysically salient characteristics (like size)—and to focus on the task of finding my keys. Visual salience, then, is given by what I am looking *for*, offering a simplistic example of the intentional nature of perception.

Notice that this perspective suggests that perception is a form of active *presentation* rather than representation. That is, the organism plays a critical role in actively presenting certain stimuli or objects, rather than representing them (or the environment more broadly). As put by Brentano (1982/1985, pp. 78–79; also see Albertazzi, 2015), "by presentation I do not mean what is presented, but rather *the act* of presentation."[10] The sought-after object becomes salient, presenting itself to us through the process of active probing and search by the organism.

Our key point here is that *visual search is not just organism-specific but also task-, problem-, and object-specific.* That is, our moment-by-moment awareness happens in generative fashion and is structured by what we are looking for and "doing"—or asked to do—at any given moment.[11] This generative and presentational lens on perception means that any appeals to notions of human perceptual "blindness" or bias—a common point of emphasis in the rationality literature (see Kahneman, 2011; Felin et al., 2017)—simply do not make any sense. This fundamentally changes how scholars of rationality should think about perception, particularly as perceptual and psychophysical arguments are at the very heart of rationality (Kahneman, 2003; for a review, see Chater et al., 2018). For example, Kahneman (2011, pp. 23–24) extends the core argument of the inattentional blindness literature (see Simons and Chabris, 1999) into the domain of judgment and rationality and argues that humans are "blind to the obvious." But the reason humans "miss" things in their visual scenes—things that should be obvious (based on the logic of psychophysics)—is not because they are blind, but rather because they are engaged in tasks which direct their awareness toward other things (Felin et al., 2019). This points toward a "presentational" view of perception, where what presents itself are the cues or objects that we are looking for (or asked to look for), rather than a representational view that focuses on those cues or objects that have certain (*a priori*) psychophysical features or characteristics [what Kahneman (2003) calls "natural assessments" such as the size, distance or loudness of cues and objects].

A better way, then, to think about the organism-environment relationship—so fundamental to the bounded rationality literature (Simon, 1956, 1990; Gigerenzer and Gaissmaier, 2011)—might be to speak of a more fine-grained organism-object relationship instead. That is, moment-by-moment organism awareness is about specific objects that are situation- or task-relevant. The broader notion or word "environment" thus unwittingly creates a black box that needs to be unpacked. Awareness is *about* something specific in the environment (Brentano, 1982/1985; also see Brentano, 1995/1874), rather than about the environment as a whole. Psychophysical efforts seek to understand environments by treating them like data, pixels and dots—cues and statistical properties—and therefore miss this type of specificity and the indefinite potential objects that might be salient. To offer a simple metaphor, psychophysical and bottom-up approaches to environments treat it like an urn of cues and information, one that cannot exhaustively be sampled due to costs or computational limitations (Ellsberg, 1961; Edwards et al., 1963; see Brandstätter et al., 2006; Gigerenzer, 2021). The environment might be represented with an urn of, say, 10,000 red and black balls. And truth is then represented by a full knowledge of the relative proportion of the two different colors. Our task might be to somehow estimate this truth by sampling from the urn on a more limited basis, in heuristic fashion, given the costs associated with counting all of the balls. This urn-like conception of the environment allows ecological rationality to presume a quantifiable reality, matching heuristic and statistical techniques with that reality, and to compare varied heuristic techniques against an omniscient ideal. This type of simplification, of treating the environment like an urn (or set of cues and data points), has enabled the literature to focus on

[10] As noted by Albertazzi et al. (2010, p. 8), "the central idea in Brentano's work, that of perception as presentation, has been entirely missing from cognitive science and has only recently been introduced into contemporary dialogue." For further discussion of the critically important, phenomenological aspects of vision (including associated neural mechanisms), see Koenderink (2012).

[11] The language of "top-down" is occasionally used in the context of the bounded and ecological rationality literatures (e.g., see Todd and Brighton, 2016). However, these appeals to top-down mechanisms still emphasize *predefined* cues, while our specific emphasis is on emergent cues and their psychogenesis (Koenderink, 2012).

various statistical and probabilistic approaches to understanding environments (cf. Savage, 1950).

However, this urn-like, atomistic treatment and idea of sampling environments reduces environments to bottom-up cues and data. This is the metaphorical equivalent of assuming that one might understand a painting by adding up its constituent "dots" or pigments of color. To briefly extend the metaphor, consider Seurat's painting *La Grande Jatte*, which consists of an estimated 220,000 dots (Goldstein, 2019). The problem is that no form of bottom–up sampling or quantification of these dots will communicate the same information as the top-down reading of the painting. The only thing we might learn from sampling the dots is how much of each color was used in the painting, but little else. But this is precisely how environments are metaphorically treated by ecological rationality (and literatures on scene statistics). This type of statistical analysis tells us nothing about the individual objects or subject-matter of the painting itself.[12] The key point here is that: a bottom–up conception of environments doesn't translate or scale to the real world in any meaningful way, except in limited circumstances.

Rather than speak of the broad organism-environment relationship, our focus is on the situation-relevant objects or cues within it. Perception is necessarily directed toward some object—for example, something we might be looking for—rather than the environment as a whole (or some disaggregated notion of the environment). Perceiving is *about* and *for* something specific, an object the organism is interested in. To offer an example, consider the work of Yarbus (1967). It offers a powerful example of how the search-for-something—like an answer to a question—shapes what presents itself and becomes salient and visible. Yarbus studied what he called the "perception of complex objects," specifically by tracking the eye movements of experimental subjects, in an attempt to understand what humans perceive when encountering a teeming visual scene with disparate stimuli. For example, he tracked the eye movements of subjects viewing the artist Ilya Repin's painting *The Unexpected Visitor*. Yarbus highlighted how a battery of prompts and questions that he posed shaped the stimuli and objects that were salient to experimental subjects. For example, he asked subjects to "estimate the material circumstances of the family in the picture," or to "give the ages of the people," or to "surmise what the family had been doing before the arrival of the unexpected visitor," or to "estimate how long the unexpected visitor had been away from the family." The upshot of this work is that it highlights how questions provide a type of search image for which answers are sought in visual scenes, presenting and creating salience for certain objects, cues and things at the "expense" of other things.

Notice how there is no single question that can somehow elicit all the feasible cues, objects and stimuli from a visual scene, whether we're talking about Repin's painting *Unexpected Visitor* or any other scene or environment. A generic prompt or request to simply "observe" or "describe the scene" might of course yield varied answers about the number of people in the picture, perhaps their ages, and so forth (or perhaps "typical" foci in human perception, like faces). But there's no way to meaningfully exhaust visual scenes and environments. While some fields of psychology and cognitive science insist that this is possible, we argue that this simply is not the case (for a debate and discussion, see Chater et al., 2018). And importantly, as the Yarbus example highlights, there's no way to speak of any form of psychophysical salience independent of the top-down questions and prompts that direct awareness. The salient things don't inherently "shout" their importance, as assumed by psychophysics. Object obviousness is driven by the questions, interests or tasks specified by the organism or agent in question (Koenderink, 2012).

This underlying generative logic, as we discuss next, suggests a rather significant shift in how we think about perception, with important implications for the judgment, rationality and decision making literatures as well. While it might seem obvious that, say, questions direct awareness and salience, this logic remains radically under-appreciated and is counter to the key drivers of salience from the perspective of ecological rationality, where salience is said to be given by cue characteristics, environmental structure and statistics. And while there are mentions of "top-down" perception in the bounded and ecological rationality literatures, the focus remains on the perception of *predefined* cues. Thus we next revisit the idea of "growing" awareness and cues, and we highlight how dormant cues—not readily evident or obvious—might be identified and transformed into evidence or put differently, clues-for-something.

Growing Awareness Toward (Relevant) Cues

As discussed above, psychophysics "grows" awareness toward cues based on their statistical or physical characteristics, such as intensity, frequency or size (Fechner, 1860). In its simplest form, the experimentalist essentially increases or "dials up" a specific stimulus, until awareness is reached. The focus on the amount-of-something as the critical ingredient (or mechanism) of perceptual salience is also the background logic behind "stimulus detection as intuitive statistics" (Gigerenzer, 1992), and the basis of the ongoing extensions of the logic of signal detection and size (whether sample or cue size) into the domain of ecological rationality (Gigerenzer, 2020; Kozyreva and Hertwig, 2021). To summarize (and oversimplify): psychophysics-based approaches argue that cue detectability is a function of how loud, big or intense a cue or stimulus is.

But what about situations where a critical cue has none of these salience-generating physical characteristics or statistical properties? What is the mechanism of salience in these situations? How might we detect something that is quiet, small and scarcely obvious but nonetheless highly relevant? Put differently, how is something that is hidden—or barely detectable—nonetheless detected? Is there a way of amplifying or "growing" awareness toward these types of cues? We address these questions next.

Our emphasis is specifically on the *psychogenesis* of awareness, rather than the psychophysics of perception and attention—a

[12]One exception to this might be the notion of "criterion" that is often mentioned in the context of ecological rationality. However, ecological rationality focuses on how "available cues predict the criterion" (Kozyreva and Hertwig, 2021, p. 1530; also see Hogarth and Karelaia, 2007), while our emphasis instead would be on how a criterion (like a specific question or hypothesis) enables the presentation of relevant cues—a critical distinction.

critical distinction (Koenderink, 2012, 2018; Felin et al., 2017). We essentially propose to offer an alternative, generative way of "growing" awareness toward a cue or "clue."[13] That is, rather than focusing on the intensity or size of a cue to enable its detection, we point to organism-specific, top-down mechanisms of detection. We point out how humans might become aware of "small"—seemingly non-obvious and undetectable—cues even when they have none of the traditional characteristics of salience.

Our approach to growing awareness toward a specific cue might best be introduced by an informal example. Consider Arthur Conan Doyle's fictional detective story *The Adventure of Silver Blaze*. The story features a brief but informative bit of dialogue between the Scotland Yard detective and Sherlock Holmes:

> *Scotland Yard detective*
> Is there any other point to which you would wish to draw my attention?
>
> *Sherlock Holmes*
> To the curious incident of the dog in the night-time.
>
> *Scotland Yard detective*
> The dog did nothing in the night-time.
>
> *Sherlock Holmes*
> That was the curious incident.

The story describes a situation where the protagonists—a Scotland Yard detective and Sherlock Holmes—are engaging in an effort to identify the perpetrator of a crime. The investigators encounter and seek to systematically canvas an environment with innumerable cues and potential clues: people and their motives, a crime scene with innumerable objects (some visible, some not)—any number of *in situ* and *ex situ* variables that may or may not be relevant for solving the case. In short, the environment teems with indefinite, possible and dormant cues and potential clues.

The problem of course is that *anything* could be relevant: the fact that a door or window was left open (or not), the fact that some object is present (or missing) in a particular room, the fact that a chair is two versus three meters from a door, or that focal building in question is 120 miles from London, or that there is (or isn't) a cigar butt on the ground, etc. In short, it's impossible to know what might be relevant. Furthermore, the key clue or piece of information might be small and scarcely obvious. There's no computational or statistical procedure for processing the scene. And important for our arguments, there is no *a priori* environmental structure that we might speak of.

The reason we highlight the above dialogue between Sherlock and the Scotland Yard detective is because it highlights a critical, generalizable point. Namely, one of the critical cues in this particular case (evident in the dialogue)—the dog that *didn't* bark—has *no* physical or statistical properties whatsoever: it is not loud or large, it is not repeated, nor obvious in any meaningful way. There is no way to argue for psychophysical salience nor to point to some form of *a priori* representation. The example of course is fictional. But it nicely illustrates how a relevant cue might not meet any of the traditional characteristics of cue salience or detectability, as specified by psychophysics or ecological rationality. Rather, here we have a situation where the *lack* of an auditory sound—a dog *not* barking—is identified as curious and critical, providing vital information about the crime (in this case, the dog didn't bark and therefore someone familiar with the dog was present at the crime scene).

The point we want to make is that cues do not say or mean anything by themselves. Just like in science, cues and data are meaningless without a theory or some alternative top-down factor, like a hypothesis, question or conjecture.[14] The problem in science is that, as put by physical chemist Polanyi (1957, p. 31), "things are not labeled 'evidence' in nature" (for a recent discussion see Felin et al., 2021). Similarly, environmental cues don't come with labels that say "this is relevant or important" or "this is evidence." Cues—clues for something—are not inherently obvious. Furthermore, the size or amount of cues or samples cannot be equated with relevance or importance either. There is no "scene statistics" for resolving Sherlock's case, just as there are no general statistics for processing visual scenes and environments (Koenderink, 2012). Cues are simply raw material and dormant data, until they are met with a probing organism and the right question. In this sense, cues are *made* visible rather than being inherently visible. Some form of top-down mechanism is needed to generate or grow awareness toward cues, to engage in what might be called a "cue-to-clue transformation."[15]

Related to this transformation, it's interesting to note that in Simon's (1956) pathbreaking paper—"Rational choice and the structure of the environment"—he uses the word "clue" a number of times (while "cues" are the emphasis in the ecological rationality literature: Gigerenzer and Gaissmaier, 2011). Most of the instances of the word "clue" in Simon's article are used in a relatively traditional psychophysical sense, where clues are perceptually seen based on their vicinity ("an organism's vision permits it to see a circular portion"—Simon, 1956, p. 130; cf. Kahneman, 2003). But at the end of the article the word "clue" is parenthetically used in a more investigative and anticipatory sense. Specifically, Simon (1956, p. 136, *emphasis added*) discusses how an organism might search an environment randomly, or alternatively, on the basis of "clues in the environment (either the actual visibility of need-satisfying points or *anticipatory clues*)." It's Simon's parenthetical

[13]The language of "growing awareness" has also been used in the economics literature (see Karni and Vierø, 2013). However, that literature builds on various large and small-world conceptions (cf. Savage, 1950) to model "expanding state spaces" and their implications for economic decision makers. Our approach, instead, is focused on perception. We address how awareness toward novel cues or objects might be endogenously grown, as well as the critical cue-to-clue transformation (building on Koenderink, 2012).

[14]The exploratory and generative process of hypothesizing can be seen as a general biological process, where organisms (of all stripes) engage in this process (Riedl, 1984; cf. Popper, 2013).

[15]We use the language of a "cue-to-clue transformation" to make our point about how awareness toward something/anything requires active probing on the part of the organism. In an important sense, the specification or recognition of *any* cue necessarily requires some mechanism for generating awareness. That is, strictly speaking, any qualifier that we might use in front of the word cue (a *salient* cue, a *relevant* cue, an *important* cue, a *meaningful* cue, a *surprising* cue, etc.) is redundant (Koenderink, 2012). However, we nonetheless use this language to help us explicate our central argument relative to existing ecological approaches.

remark about "*anticipatory* clues" that finds some resonance with our discussion of generative rationality here. That is, an organism's ability to recognize and see something as a clue might be *independent* of proximity (visual proximity or distance being the key mechanism of salience for the bounded rationality literature) or other psychophysical measures of salience (such as size). In other words, cue salience can also emerge independent of distance or independent of other physical characteristics. Our approach here can be seen as an effort to develop the organism-specific factors that enable this type of anticipation and recognition of tentative clues, where the search images, probing, conjectures and hypotheses of organisms—independent of the psychophysical characteristics of the cues (as measured by, say, their vicinity, proximity or size)—can shape judgment and decision making. Thus, again, our approach is firmly focused on the active, *presentational* aspects of rationality, rather than their representational nature.

In a generative sense, awareness toward a cue or cues needs to be actively nurtured—the *relevant* cues need to somehow be identified, presented and made salient, from amongst the meaningless mass of potential and indefinite things within an environment or scene. Returning briefly back to our short Sherlock dialogue, notice how even *after* Sherlock points the dog out to the Scotland Yard detective, the latter still remains puzzled as to why the dog is in any way relevant to the situation, that is, why the dog (cue) represents a "curious incident." This indeed is the problem: any cue could be "curious" and important, or not. But for something to "pop out" and become meaningful, from amongst indefinite potential cues—or put differently, for a cue to count as evidence, for it to signal something—requires a top-down mechanism. In essence, we are saying that there are indefinite varieties of signal detection beyond simply looking at the amount or intensity of a cue or cues. Our generative form of visual "pop out" therefore is fundamentally different from psychophysical approaches to vision and perception (see Wolfe and Horowitz, 2017). Some form of top-down rationale is needed to enable us to recognize a cue in the first place, as the cue does not inherently impose itself onto our awareness, but only becomes salient in response to active probing. Top-down factors or reasons play a critical role in presenting, specifying and selecting the relevant cues—again, independent of the physical qualities of cues. And in the case of our Sherlock example, the top-down imposition of a "plot"—an imagined, hypothesized conjecture or narrative of what happened—directs salience toward certain objects, cues, features and aspects of the environment (Koenderink, 2012). The plot makes the cue salient. Without a top-down plot, there is no reason whatsoever for the non-barking dog to be salient or evident in any way. It's only with the top-down plot that a cue (or clue), such as the dog not barking, can even meaningfully be identified.

To offer a contrast, in our hypothetical Sherlock situation it's hard to point to any of the heuristics from ecological rationality that might similarly resolve the situation. We might, perhaps—*in retrospect*—be able to shoe-horn an explanation that is in line with ecological rationality by saying that the "non-barking-dog"-cue is identified through some mechanism of random or other form of sampling (Though it's hard to imagine how one might, in the first place, become aware of the non-barking dog and its importance). Or we might highlight a growing "tally" of cues that increasingly, in the aggregate, point to a threshold conclusion that a particular individual is the sought-after culprit in the case—the non-barking-dog being one of many cues pointing in this direction. But any heuristics or associated statistics that we might point to are merely an after-the-fact epiphenomenon of a process that is necessarily initiated top-down. Again, cues themselves don't say or mean anything, they aren't somehow inherently evident (based on, say, their physical characteristics). Rather, cues become cues-*for*-something, or clues, in the context of a particular top-down plot. That said, we of course recognize that the plot might be wrong, but it can readily be amended if the relevant cues and evidence cannot be found. Thus we need an *a priori* way of generating awareness toward specific cues, a reason for growing or elevating—and creating salience for—a particular cue based on some top-down factor.

Now, we have of course pointed to a fictional example. But this idea of having a top-down "plot" might be generalized to both mundane, daily experiences as well as more novel ones. To offer an everyday example (linked to the aforementioned example of lost keys): if I have lost my house keys, my visual search for them is guided by a key search image. I know what I am looking for, what my keys look like, and thus I can scan for key-like items in my surroundings. Importantly, this visual "investigation" and search is critically enabled by me having a conjecture or hypothesis (an informal plot, of sorts) about where I might have lost the keys in the first place. I might remember having had the keys two hours earlier, and I might therefore trace my steps and search across the rooms I've occupied during the intervening time period. No form of random sampling or item-by-item inspection makes sense in this situation. Nor does any notion of psychophysical salience. After all, not only are my keys "small" but they might have slipped into the crack of the couch and thus not even be visible. But a top-down plot or hypothesis enables me to find them.

Beyond the mundane search for keys, these top-down factors are also the underlying mechanism behind the emergence of novelty, including in the sciences. Science itself might be seen as an effort to "grow" awareness and salience toward novel objects or unique observations, things that previously were non-obvious and seemingly hidden. Theories serve a top-down plot-like function in enabling us to observe and see a new cue, data point or piece of evidence—or to see something (like an apple falling) in a completely new way. As put by Einstein, "whether you can observe a thing or not depends on the theory which you use. It is the theory which decides what can be observed" (Polanyi, 1971, p. 604). Furthermore, theories might lead us to construct instruments or technologies—such as telescopes or microscopes—and methods for making observations of things that are not evident to the naked eye (for a recent discussion, see Felin et al., 2021). For example, the postulation of gravitational waves led to the construction of detectors to measure them. Cue-first-based, psychophysical approaches do not offer this type of mechanism for observing novel things. Bayesian approaches similarly are unable to address questions of novel observation. This is informally illustrated by the fact that no amount of

watching falling items (like apples) will yield insights about gravity, without first having a conjecture, hunch or theory about what one is looking for and at.

The idea of top-down theories also has critical implications for economic settings, which abound in uncertainty and latent possibility. Biological intuition has traditionally been applied to economic settings at the level of randomness and environmental selection (Penrose, 1952). Or ecological rationality focuses on the long-run evolutionary adaptation of the mind to changing environments (Gigerenzer, 2000; also see Cosmides and Tooby, 2013). What is missing in this work is the organism-directed probing and exploration that also shapes and creates novelty. That is, rather than merely passively adapting to their surroundings, organisms (including economic agents) make novel use of objects around them. Economic environments are inherently "unprestatable," and entrepreneurs and managers can identify novel uses and affordances (Kauffman, 2014). Thus, rather than merely adapting to environments, important exaptive mechanisms also play a role (e.g., La Porta et al., 2020; Cattani and Mastrogiorgio, 2021).

SOME CONCLUDING REMARKS: SCISSORS REVISITED

We believe that our generative view of rationality offers a unique way to think about rationality, with novel implications for future work. To illustrate this, by way of some concluding remarks, and to highlight links to bounded rationality, we briefly revisit Simon's (1990) famous and oft-quoted "scissors" metaphor (e.g., Chase et al., 1998; Gigerenzer and Selten, 2001; Gigerenzer and Gaissmaier, 2011; Puranam et al., 2015; Petracca, 2021). Simon's scissors metaphor is an evocative idea that has been discussed or mentioned in hundreds of articles over the past decades. We highlight how a focus on generativity might offer a useful and different way to think about the two "blades" of the scissors, with attendant implications for judgment and decision making in situations of uncertainty.

Simon's (1990, pp. 7–9) scissors metaphor is the idea that rationality is shaped by two blades, namely, the "structure of the environment" and the "computational capabilities of the actor." In the ecological rationality literature, the two blades are summarized as the "internal and external *constraints*" of judgment and decision making (Kozyreva and Hertwig, 2021, p. 1524, emphasis added). Or to cite Todd and Gigerenzer (2003, p. 143), the scissors metaphor is the overarching idea "that human rationality is bounded by both internal (mental) and external (environmental) constraints" (also see Gigerenzer and Goldstein, 1996; Chater and Oaksford, 1999). This two-pronged, blades approach is also central for ongoing definitions of uncertainty as well. For example, Kozyreva and Hertwig (2021, p. 1525) argue that "uncertainty concerns environmental constraints as well as computational constraints, which both prevent the subject from determining the structure of the environment." In all, the emphasis on both organism-related and environmental *constraints* is ubiquitous and offers a useful contrast to how generative rationality characterizes the two blades.

Rather than focus on constraints (important as they undoubtedly are), our emphasis in this paper has been on the *generative* nature of organisms and the *teeming* nature of environments. Thus our arguments might be seen as a friendly amendment for how we might think about the organism-environment interface—specifically, a call to recognize the novel and emergent aspects of both sides of the organism-environment interface. While the ideals of optimization and constraint are heavily emphasized and juxtaposed in existing work, this has come at the expense of understanding how novelty emerges. Of course, in shifting the emphasis from constraint and boundedness to generativity, we certainly do not mean to suggest—as the examples below will illustrate—that organisms are characterized by some form of omniscience, or that there aren't costs and limits associated with judgment and decision making. Constraints and boundedness are important. However, we do think that the heavy emphasis on the constraints of information processing—and the experiments constructed to point this out—have unnecessarily sidelined the generative nature of organisms and the possibilities presented by teeming environments.[16]

Before offering some examples, it's important to point out that the scissors metaphor was specifically discussed by Simon (1990; cf. Newell and Simon, 1972) in an article that focuses on the "invariants" and similarities between human judgment, computers and general information processing (also see Simon, 1980). The computational logic has readily lent itself to extensions like the idea of humans as "intuitive statisticians" and the importance of the statistical toolkit and environmental structure. But this conception of rationality is highly dependent on the types of tasks, experiments and examples that scholars construct and focus on. Computers undoubtedly perform computational tasks well, indeed, better than humans. But what the computational and statistical analogies miss is the situations, tasks and settings where human judgment readily outperforms any form of computation or statistical processing (cf. Culberson, 1998). This is particularly the case for novel situations and uncertain environments, where environmental structure can't be specified ex ante. For example compared to computers, humans and other living organisms routinely solve the "frame problem," an impossibility for computers (McCarthy and Hayes, 1981), where humans readily discover new uses and affordances that simply aren't computationally pre-statable (Felin et al., 2014; Kauffman, 2014).

To make this point more concrete, and to informally contrast the computational logic with the generative one, consider a

[16]There are some research streams that touch on related issues (though they are not directly focused on perception and rationality). For example, Grandori (2010) discusses how the bounded rationality literature also needs to understand scientific and economic *discovery*. Others have focused on notions such as "creative rationality," and the logic of abduction (e.g., Gooding, 1996; Forest, 2017). Felin and Zenger (2017) look at how economic *theories*—and associated problem formulation and solving—shape perception and the emergence of novelty. More broadly, Viale (2020) recently highlights various literatures that touch on the creative or novel aspects of bounded rationality. Unfortunately we cannot cover all of this work. While all of this work is broadly related, our specific focus is different. Namely, we are focused on the *perceptual* foundations of ecological rationality (as well as our generative alternative). But we certainly see opportunities for future work to carefully make linkages across our arguments and the aforementioned literatures.

simple search problem like the frequently discussed search for a needle in a haystack (see Simon, 1969; Simon, 1978; cf. Baumol, 1979; Winter, 2000). Here we have a quintessential (albeit stylized) search problem, where we are faced with an overwhelming search task. To find the needle, we might engage in some form of "brute force" search, where we select an item randomly, and iterate item-by-item through the objects until we encounter the needle or the item we seek (Culberson, 1998). This type of "exhaustive" search of course is overly costly and prohibitive. Thus we might think about applying heuristics or "search rules" to solve the problem—rules about where to search and when to stop searching (see Gigerenzer and Gaissmaier, 2011, pp. 454–456). Simon for example imagines a haystack where needles of varied sharpness are distributed randomly, and highlights how we might decide to satisfice and end search when we encounter a needle that is "sharp enough" (Simon and Kadane, 1975).

But humans can readily solve these types of search problems—like the needle-haystack problem—in various novel and creative ways. For example, we might postulate that the needle is made of steel and is nickel-plated, and therefore use a powerful magnet to quickly find the needle. Or we might, say, burn the haystack or use some kind of large sieve or leaf blower. Or perhaps some kind of sorting device could even be constructed from the hay itself. Or we might delimit the search by hypothesizing that the needle—due to its relative size and weight—is best found by looking on the ground (Felin and Kauffman, 2021). Thus the brute or exhaustive search option need not be held up as an ideal, as varied shortcuts and solutions can be generated. Notice that this type of creative problem-solving—the hallmark of generativity—is in fact ubiquitous in nature. This type of creative problem solving is not just a human prerogative, but innovative problem solving and tool use is evident across species (Fragaszy and Liu, 2012; Griffin and Guez, 2014; Morand-Ferron et al., 2016; Fragaszy and Mangalam, 2018; Amici et al., 2019).

Thus the hacks and solutions to search, judgment and decision making might involve utilizing tools and objects in our environments in various creative ways, beyond statistical inference or computation. Even ecological rationality's popular city size task (Gigerenzer and Goldstein, 1996)—discussed by us extensively above—can easily be solved by, say, asking someone, or by quickly looking the answer up on the internet. In other words, *in the real world* we use the material resources, affordances and technologies around us in creative ways to come up with solutions (Uexküll, 2010; Gabora, 2019). While the prototypical decision tasks and environments of ecological rationality try to offer a tractable microcosm for helping us understand judgment and decision making, it's hard to see how these decision tasks—like the frequently used city size experiment—generalize to more uncertain settings. For example, the tasks of an entrepreneur or manager are fundamentally different from anything like comparing city sizes: they are highly ambiguous and highly multidimensional. This doesn't mean that judgment should be studied by, say, using inkblots. But the classic literature, for example, on functional fixedness (James, 1890; Duncker, 1945), might offer a basis for exploring judgment decision making and creativity in situations of uncertainty.

In all, the existing literature—within the domain of bounded and ecological rationality—should recognize the affordances, uses and functions of the material world. With our focus on the "generative" nature of rationality we hope to emphasize the possibility of these emergent and novel outcomes. The statistical and computational tasks that characterize the extant literature on bounded and ecological rationality are of course important. Undoubtedly representational and statistical approaches have their place. But it's important for scholars to also address the generative (presentational or even "expressive") aspects of perception, as these relate to judgment and decision making. Thus our hope is that this paper—an effort to outline the broad contours of a generative approach to rationality—might offer the basis for future work along these lines.

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

TF wrote the first version of the manuscript, with key ideas and inputs provided by JK. Both authors edited, added to and revised the full manuscript.

REFERENCES

Albertazzi, L. (2015). "Philosophical background: phenomenology," in *The Oxford Handbook of Perceptual Organization*, ed. J. Wagemans (Oxford: Oxford University Press), 21–40.

Albertazzi, L., Van Tonder, G. J., and Vishwanath, D. (eds). (2010). *Perception Beyond Inference: The Information Content of Visual Processes*. Cambridge, MA: MIT Press.

Algom, D. (2021). The Weber–Fechner law: a misnomer that persists but that should go away. *Psychol. Rev.* 128, 757–765. doi: 10.1037/rev0000278

Amici, F., Widdig, A., Lehmann, J., and Majolo, B. (2019). A meta-analysis of interindividual differences in innovation. *Anim. Behav.* 155, 257–268. doi: 10.1016/j.anbehav.2019.07.008

Balcetis, E., and Dunning, D. (2006). See what you want to see: motivational influences on visual perception. *J. Pers. Soc. Psychol.* 91, 612–625. doi: 10.1037/0022-3514.91.4.612

Baumol, W. J. (1979). On the contributions of Herbert A. Simon to economics. *Scand. J. Econ.* 81, 74–82. doi: 10.2307/3439459

Bordalo, P., Gennaioli, N., and Shleifer, A. (2021). *Salience*. National Bureau of Economic Research (working paper #29274). Cambridge, MA: National Bureau of Economic Research.

Boring, E. G. (1942). *Sensation and Perception in the History of Experimental Psychology*. New York, NY: Appleton-Century.

Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The priority heuristic: making choices without trade-offs. *Psychol. Rev.* 113:409. doi: 10.1037/0033-295X.113.2.409

Brentano, F. (1982/1985). *Descriptive Psychology*. London: Routledge.

Brentano, F. (1995/1874). *Psychology From An Empirical Standpoint*. London: Routledge.

Bruner, J. S., and Goodman, C. C. (1947). Value and need as organizing factors in perception. *J. Abnorm. Soc. Psychol.* 42, 33–45. doi: 10.1037/h0058484

Cattani, G., and Mastrogiorgio, M. (eds). (2021). *New Developments in Evolutionary Innovation: Novelty Creation in a Serendipitous Economy*. Oxford: Oxford University Press.

Caves, E. M., Nowicki, S., and Johnsen, S. (2019). Von Uexküll revisited: addressing human biases in the study of animal perception. *Integr. Comp. Biol.* 59, 1451–1462. doi: 10.1093/icb/icz073

Chase, V. M., Hertwig, R., and Gigerenzer, G. (1998). Visions of rationality. *Trends Cogn. Sci.* 2, 206–214. doi: 10.1016/s1364-6613(98)01179-6

Chater, N., Felin, T., Funder, D. C., Gigerenzer, G., Koenderink, J. J., Krueger, J. I., et al. (2018). Mind, rationality, and cognition: an interdisciplinary debate. *Psychon. Bull. Rev.* 25, 793–826. doi: 10.3758/s13423-017-1333-5

Chater, N., and Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cogn. Psychol.* 38, 191–258. doi: 10.1006/cogp.1998.0696

Chater, N., Oaksford, M., Nakisa, R., and Redington, M. (2003). Fast, frugal, and rational: how rational norms explain behavior. *Organ. Behav. Hum. Decis. Process.* 90, 63–86. doi: 10.1016/s0749-5978(02)00508-3

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Cosmides, L., and Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58, 1–73. doi: 10.1016/s0010-0277(00)00098-6

Cosmides, L., and Tooby, J. (2013). Evolutionary psychology: new perspectives on cognition and motivation. *Annu. Rev. Psychol.* 64, 201–229. doi: 10.1146/annurev.psych.121208.131628

Cronin, T. W., Johnsen, S., Marshall, N. J., and Warrant, E. J. (2014). *Visual Ecology*. Princeton, NJ: Princeton University Press.

Culberson, J. C. (1998). On the futility of blind search: an algorithmic view of "no free lunch". *Evol. Comput.* 6, 109–127. doi: 10.1162/evco.1998.6.2.109

Dawes, R. M., and Corrigan, B. (1974). Linear models in decision making. *Psychol. Bull.* 81, 95–105.

Dhami, M. K., Hertwig, R., and Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychol. Bull.* 130, 959–988. doi: 10.1037/0033-2909.130.6.959

Dieckmann, A., and Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Mem. Cogn.* 35, 1801–1813. doi: 10.3758/bf03193511

Dougherty, M. R., Franco-Watkins, A. M., and Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychol. Rev.* 115, 199–213. doi: 10.1037/0033-295X.115.1.199

Duncker, K. (1945). On problem-solving. *Psychol. Monogr.* 58, 1–113.

Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychol. Rev.* 70, 193–214. doi: 10.1037/h0044139

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Q. J. Econ.* 75, 643–669. doi: 10.2307/1884324

Ewert, J. P. (2004). "Motion perception shapes the visual world of amphibians," in *Complex Worlds From Simpler Nervous Systems*, ed. F. R. Prete (Cambridge, MA: MIT Press), 117–160. doi: 10.1242/jeb.167700

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.

Feldman, J. (2017). What are the "true" statistics of the environment? *Cogn. Sci.* 41, 1871–1903. doi: 10.1111/cogs.12444

Felin, T., Felin, M., Krueger, J. I., and Koenderink, J. (2019). On surprise-hacking. *Perception* 48, 109–114. doi: 10.1177/0301006618822217

Felin, T., and Kauffman, S. (2021). "The search function and evolutionary novelty," in *New Developments in Evolutionary Innovation: Novelty Creation in a Serendipitous Economy*, eds G. Cattani, and M. Mastrogiorgio (Oxford: Oxford University Press), 113–143. doi: 10.1093/oso/9780198837091.001.0001

Felin, T., Kauffman, S., Koppl, R., and Longo, G. (2014). Economic opportunity and evolution: beyond landscapes and bounded rationality. *Strateg. Entrep. J.* 8, 269–282. doi: 10.1002/sej.1184

Felin, T., Koenderink, J., and Krueger, J. I. (2017). Rationality, perception, and the all-seeing eye. *Psychon. Bull. Rev.* 24, 1040–1059. doi: 10.3758/s13423-016-1198-z

Felin, T., Koenderink, J., Krueger, J. I., Noble, D., and Ellis, G. F. (2021). The data-hypothesis relationship. *Genome Biol.* 22, 1–4.

Felin, T., and Zenger, T. R. (2017). The theory-based view: economic actors as theorists. *Strategy Sci.* 2, 258–271. doi: 10.1287/stsc.2017.0048

Filevich, E., Horn, S. S., and Kühn, S. (2019). Within-person adaptivity in frugal judgments from memory. *Psychol. Res.* 83, 613–630. doi: 10.1007/s00426-017-0962-7

Forest, J. (2017). *Creative Rationality and Innovation*. New York, NY: John Wiley & Sons.

Fragaszy, D., and Liu, Q. (2012). "Instrumental behavior, problem-solving, and tool use in nonhuman animals," in *Encyclopedia of the Sciences of Learning*, ed. N. M. Seel (New York, NY: Springer), 1579–1582. doi: 10.1007/978-1-4419-1428-6_928

Fragaszy, D. M., and Mangalam, M. (2018). "Tooling," in *Advances in the Study of Behavior*, Vol. 50, eds M. Naguib, L. Barrett, S. D. Healy, J. Podos, L. W. Simmons, and M. Zuk (Amsterdam: Academic Press), 177–241.

Gabora, L. (2019). Creativity: linchpin in the quest for a viable theory of cultural evolution. *Curr. Opin. Behav. Sci.* 27, 77–83.

Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science* 349, 273–278. doi: 10.1126/science.aac6076

Gigerenzer, G. (1991). From tools to theories: a heuristic of discovery in cognitive psychology. *Psychol. Rev.* 98, 254–267.

Gigerenzer, G. (1992). Discovery in cognitive psychology: new tools inspire new theories. *Sci. Context* 5, 329–350.

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality in the Real World*. New York, NY: Oxford University Press.

Gigerenzer, G. (2008). Why heuristics work. *Perspect. Psychol. Sci.* 3, 20–29. doi: 10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G. (2020). How to explain behavior? *Top. Cogn. Sci.* 12, 1363–1381.

Gigerenzer, G. (2021). Axiomatic rationality and ecological rationality. *Synthese* 198, 3547–3564.

Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Topics Cogn. Sci.* 1, 107–143. doi: 10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Annu. Rev. Psychol.* 62, 451–482.

Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–670. doi: 10.1037/0033-295x.103.4.650

Gigerenzer, G., and Goldstein, D. G. (2011). The recognition heuristic: a decade of research. *Judgm. Decis. Mak.* 6, 100–121.

Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychol. Rev.* 98, 506–531. doi: 10.1037/0033-295x.98.4.506

Gigerenzer, G., and Marewski, J. N. (2015). Surrogate science: the idol of a universal method for scientific inference. *J. Manag.* 41, 421–441.

Gigerenzer, G., and Murray, D. J. (1987). *Cognition as Intuitive Statistics*. London: Psychology Press.

Gigerenzer, G., and Selten, R. (eds). (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.

Gigerenzer, G., and Todd, P. M. (1999). "Fast and frugal heuristics: the adaptive toolbox," in *Simple Heuristics That Make Us Smart*, eds G. Gigerenzer, P. M. Todd, and The ABC Research Group (New York, NY: Oxford University Press), 3–34.

Goldstein, D. G., and Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychol. Rev.* 109, 75–90.

Goldstein, D. G., and Gigerenzer, G. (2008). The recognition heuristic and the less-is-more effect. *Handb. Exp. Econ. Results* 1, 987–992. doi: 10.1016/s1574-0722(07)00106-0

Goldstein, D. G., Gigerenzer, G., Hogarth, R. M., Kacelnik, A., Kareev, Y., Klein, G., et al. (2001). "Why and when do simple heuristics work?", in *Bounded Rationality: The Adaptive Toolbox. Dahlem Workshop Report*, eds G. Gigerenzer and R. Selten (Cambridge, MA: MIT Press), 173–190.

Goldstein, J. L. (2019). Seurat's dots: a shot heard 'round the art world—fired by an artist, inspired by a scientist. *Cell* 179, 46–50. doi: 10.1016/j.cell.2019.07.051

Goldstein, K. (1963). *The Organism*. Boston, MA: Beacon Press.

Gooding, D. (1996). Creative rationality: towards an abductive model of scientific change. *Philosofica* 58, 73–102.

Grandori, A. (2010). A rational heuristic model of economic decision making. *Rationality Soc.* 22, 477–504.

Griffin, A. S., and Guez, D. (2014). Innovation and problem solving: a review of common mechanisms. *Behav. Process.* 109, 121–134. doi: 10.1016/j.beproc.2014.08.027

Hau, R., Pleskac, T. J., and Hertwig, R. (2010). Decisions from experience and statistical probabilities: why they trigger different choices than a priori probabilities. *J. Behav. Decis. Mak.* 23, 48–68.

Heck, D. W., and Erdfelder, E. (2017). Linking process and measurement models of recognition-based decisions. *Psychol. Rev.* 124, 442–473. doi: 10.1037/rev0000063

Hertwig, R., Hogarth, R. M., and Lejarraga, T. (2018). Experience and description: exploring two paths to knowledge. *Curr. Dir. Psychol. Sci.* 27, 123–128. doi: 10.1007/s10897-017-0071-1

Hertwig, R., Leuker, C., Pachur, T., Spiliopoulos, L., and Pleskac, T. J. (2021). Studies in ecological rationality. *Top. Cogn. Sci.* doi: 10.1111/tops.12567

Hertwig, R., and Pleskac, T. J. (2010). Decisions from experience: why small samples? *Cognition* 115, 225–237. doi: 10.1016/j.cognition.2009.12.009

Hoffman, D. D., Singh, M., and Prakash, C. (2015). The interface theory of perception. *Psychon. Bull. Rev.* 22, 1480–1506.

Hoffrage, U. (2011). Recognition judgments and the performance of the recognition heuristic depend on the size of the reference class. *Judgm. Decis. Mak.* 6, 43–57.

Hoffrage, U., and Hertwig, R. (2006). "Which world should be represented in representative design?," in *Information Sampling and Adaptive Cognition*, eds K. Fiedler and P. Juslin (New York, NY: Cambridge University Press), 381–408.

Hogarth, R. M. (2005). The challenge of representative design in Psychology and economics. *J. Econ. Methodol.* 12, 253–263.

Hogarth, R. M., and Karelaia, N. (2007). Heuristic and linear models of judgment: matching rules and environments. *Psychol. Rev.* 114:733. doi: 10.1037/0033-295X.114.3.733

Hutchinson, J. M., and Gigerenzer, G. (2005). Simple heuristics and rules of thumb: where psychologists and behavioural biologists might meet. *Behav. Process.* 69, 97–124. doi: 10.1016/j.beproc.2005.02.019

James, W. (1890). *The Principles of Psychology*. New York, NY: Henry Holt & Co.

Juslin, P., and Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: a sampling model of confidence in sensory discrimination. *Psychol. Rev.* 104, 344–366. doi: 10.1037/0033-295x.104.2.344

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475.

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Macmillan.

Karelaia, N., and Hogarth, R. M. (2008). Determinants of linear judgment: a meta-analysis of lens model studies. *Psychol. Bull.* 134, 404–426. doi: 10.1037/0033-2909.134.3.404

Karni, E., and Vierø, M. L. (2013). Reverse bayesianism: a choice-based theory of growing awareness. *Am. Econ. Rev.* 103, 2790–2810.

Katsikopoulos, K. V., Schooler, L. J., and Hertwig, R. (2010). The robust beauty of ordinary information. *Psychol. Rev.* 117, 1259–1280. doi: 10.1037/a0020418

Kauffman, S. A. (2014). Prolegomenon to patterns in evolution. *Biosystems* 123, 3–8. doi: 10.1016/j.biosystems.2014.03.004

Kingdom, F. A. A., and Prins, N. (2016). *Psychophysics: A Practical Introduction*. London: Academic Press.

Koenderink, J. J. (2011). "Vision and information," in *Perception Beyond Inference: The Information Content of Visual Processes*, eds L. Albertazzi, G. Tonder, and D. Vishnawath (Cambridge: MIT Press), 27–58. doi: 10.1155/IJBI/2006/92329

Koenderink, J. J. (2012). Geometry of imaginary spaces. *J. Physiol.* 106, 173–182. doi: 10.1016/j.jphysparis.2011.11.002

Koenderink, J. J. (2014). The all seeing eye? *Perception* 40, 1–6.

Koenderink, J. J. (2018). *The Way of the Eye*. Utrecht: De Clootcrans Press.

Kozyreva, A., and Hertwig, R. (2021). The interpretation of uncertainty in ecological rationality. *Synthese* 198, 1517–1547. doi: 10.1007/s11229-019-02140-w

Krebs, J. R., and Dawkins, R. (1984). "Animal signals: mind reading and manipulation," in *Behavioural Ecology: An Evolutionary Approach*, eds J. R. Krebs and N. B. Davies (Oxford: Blackwell Scientific Publications).

La Porta, C., Zapperi, S., and Pilotti, L. (2020). *Understanding Innovation Through Exaptation*. Cham: Springer.

Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res.* 35, 389–412. doi: 10.1016/0042-6989(94)00176-m

Leong, Y. C., Hughes, B. L., Wang, Y., and Zaki, J. (2019). Neurocomputational mechanisms underlying motivated seeing. *Nat. Hum. Behav.* 3, 962–973. doi: 10.1038/s41562-019-0637-z

Lieder, F., and Griffiths, T. L. (2020). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* 43, 1–60. doi: 10.1017/S0140525X1900061X

Link, S. W. (1994). Rediscovering the past: Gustav Fechner and signal detection theory. *Psychol. Sci.* 5, 335–340. doi: 10.1111/j.1467-9280.1994.tb00282.x

Longo, G., Montévil, M., Sonnenschein, C., and Soto, A. M. (2015). In search of principles for a theory of organisms. *J. Biosci.* 40, 955–968.

Lu, Z. L., and Dosher, B. (2013). *Visual Psychophysics: From Laboratory to Theory*. Cambridge, MA: MIT Press.

Luan, S., Reb, J., and Gigerenzer, G. (2019). Ecological rationality: fast-and-frugal heuristics for managerial decision making under uncertainty. *Acad. Manag. J.* 62, 1735–1759.

Luan, S., Schooler, L. J., and Gigerenzer, G. (2011). A signal-detection analysis of fast-and-frugal trees. *Psychol. Rev.* 118, 316–331. doi: 10.1037/a0022684

Luan, S., Schooler, L. J., and Gigerenzer, G. (2014). From perception to preference and on to inference: an approach–avoidance analysis of thresholds. *Psychol. Rev.* 121, 501–525. doi: 10.1037/a0037025

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychol. Rev.* 70, 61–79. doi: 10.1037/h0039723

Luce, R. D. (1977). The choice axiom after twenty years. *J. Math. Psychol.* 15, 215–233.

Marewski, J. N., Gaissmaier, W., and Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cogn. Process.* 11, 103–121. doi: 10.1007/s10339-009-0337-0

Marr, D. (1982). *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press.

Marshall, J., and Arikawa, K. (2014). Unconventional colour vision. *Curr. Biol.* 24, R1150–R1154.

Martignon, L., and Hoffrage, U. (2002). Fast, frugal, and fit: simple heuristics for paired comparison. *Theory Decis.* 52, 29–71.

McCarthy, J., and Hayes, P. J. (1981). "Some philosophical problems from the standpoint of artificial intelligence," in *Readings in Artificial Intelligence*, eds B. L. Weber and N. J. Nilsson (Burlington: MAMorgan Kaufmann), 431–450. doi: 10.1097/00006123-199604000-00001

Meder, B., and Gigerenzer, G. (2014). "Statistical thinking: No one left behind," in *Probabilistic Thinking*, eds E. Chernoff and B. Sriraman (Dordrecht: Springer), 127–148. doi: 10.1007/978-94-007-7155-0_8

Morand-Ferron, J., Cole, E. F., and Quinn, J. L. (2016). Studying the evolutionary ecology of cognition in the wild: a review of practical and conceptual challenges. *Biol. Rev.* 91, 367–389. doi: 10.1111/brv.12174

Neth, H., and Gigerenzer, G. (2015). "Heuristics: tools for an uncertain world," in *Emerging Trends in the Social and Behavioral Sciences*, eds R. A. Scott and S. M. Kosslyn (New York, NY: Wiley Online Library), 1–18.

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-hall.

Noble, D. (2015). Evolution beyond neo-Darwinism: a new conceptual framework. *J. Exp. Biol.* 218, 7–13.

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annu. Rev. Neurosci.* 44, 449–473. doi: 10.1146/annurev-neuro-100120-085519

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L., and Goldstein, D. G. (2011). The recognition heuristic: a review of theory and tests. *Front. Psychol.* 2:147. doi: 10.3389/fpsyg.2011.00147

Penrose, E. T. (1952). Biological analogies in the theory of the firm. *Am. Econ. Rev.* 42, 804–819.

Peterson, C. R., and Beach, L. R. (1967). Man as an intuitive statistician. *Psychol. Bull.* 68, 29–41. doi: 10.1037/h0024722

Peterson, W. W., and Birdsall, T. G. (1953). *The Theory of Signal Detectability*. Ann Arbor, MI: Michigan University Engineering Research Institute.

Petracca, E. (2021). Embodying bounded rationality: from embodied bounded rationality to embodied rationality. *Front. Psychol.* 12:710607. doi: 10.3389/fpsyg.2021.710607

Pleskac, T. J. (2007). A signal detection analysis of the recognition heuristic. *Psychon. Bull. Rev.* 14, 379–391. doi: 10.3758/bf03194081

Pleskac, T. J., and Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol. Rev.* 117, 864–901. doi: 10.1037/a0019737

Pleskac, T. J., and Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *J. Exp. Psychol. Gen.* 143, 2000–2019. doi: 10.1037/xge0000013

Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *J. Behav. Decis. Mak.* 19, 251–271.

Polanyi, M. (1957). *Personal Knowledge*. Chicago, IL: University of Chicago Press.

Polanyi, M. (1971). Genius in science. *Arch. Philos.* 34, 593–607.

Popper, K. (2013). *All Life is Problem Solving*. New York, NY: Routledge.

Puranam, P., Stieglitz, N., Osman, M., and Pillutla, M. M. (2015). Modelling bounded rationality in organizations: progress and prospects. *Acad. Manag. Ann.* 9, 337–392.

Rahnev, D., and Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behav. Brain Sci.* 41, 1–66.

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., et al. (2014). The situational eight: a taxonomy of major dimensions of situation characteristics. *J. Pers. Soc. Psychol.* 107, 677–701. doi: 10.1037/a0037250

Rauthmann, J. F., and Sherman, R. A. (2020). The situation of situation research: knowns and unknowns. *Curr. Dir. Psychol. Sci.* 29, 473–480. doi: 10.1177/0963721420925546

Richter, T., and Späth, P. (2006). Recognition is used as one cue among others in judgment and decision making. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 150–162. doi: 10.1037/0278-7393.32.1.150

Riedl, R. (1984). *Biology of Knowledge: The Evolutionary Basis of Reason*. New York, NY: Wiley.

Savage, L. J. (1950). *Foundations of Statistics*. New York, NY: Dover Publications.

Scheibehenne, B., Rieskamp, J., and Wagenmakers, E. J. (2013). Testing adaptive toolbox models: a Bayesian hierarchical approach. *Psychol. Rev.* 120, 39–57. doi: 10.1037/a0030777

Schooler, L. J., and Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychol. Rev.* 112, 610–628. doi: 10.1037/0033-295X.112.3.610

Schrödinger, E. (1944). *What is Life?*. London: Macmillan.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci.* 12, 182–186. doi: 10.1016/j.tics.2008.02.003

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychol. Rev.* 63, 129–150. doi: 10.1037/h0042769

Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT press.

Simon, H. A. (1978). On how to decide what to do. *Bell J. Econ.* 9, 494–507.

Simon, H. A. (1980). Cognitive science: the newest science of the artificial. *Cogn. Sci.* 4, 33–46. doi: 10.1016/s0364-0213(81)80003-1

Simon, H. A. (1990). Invariants of human behavior. *Annu. Rev. Psychol.* 41, 1–20. doi: 10.1146/annurev.ps.41.020190.000245

Simon, H. A., and Kadane, J. B. (1975). Optimal problem-solving search: all-or-none solutions. *Artif. Intell.* 6, 235–247. doi: 10.1016/0004-3702(75)90002-8

Simons, D. J., and Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 28, 1059–1074. doi: 10.1068/p281059

Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: cognitive complexity and the axioms of rational choice. *Think. Reason.* 19, 1–26. doi: 10.1080/13546783.2012.713178

Szollosi, A., and Newell, B. R. (2020). People as intuitive scientists: reconsidering statistical explanations of decision making. *Trends Cogn. Sci.* 24, 1008–1018. doi: 10.1016/j.tics.2020.09.005

Tanner, W. P. Jr., and Swets, J. A. (1954). A decision-making theory of visual detection. *Psychol. Rev.* 61, 401–418.

Thurstone, L. L. (1927). Three psychophysical laws. *Psychol. Rev.* 34, 424–442.

Tinbergen, N. (1963). On aims and methods of ethology. *Z. Tierpsychol.* 20, 410–433.

Todd, P. M., and Brighton, H. (2016). Building the theory of ecological rationality. *Minds Mach.* 26, 9–30.

Todd, P. M., and Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behav. Brain Sci.* 23, 727–741. doi: 10.1017/s0140525x00003447

Todd, P. M., and Gigerenzer, G. (2003). Bounding rationality to the world. *J. Econ. Psychol.* 24, 143–165. doi: 10.1016/s0167-4870(02)00200-3

Todd, P. M., and Gigerenzer, G. (2007). Environments that make us smart: ecological rationality. *Curr. Dir. Psychol. Sci.* 16, 167–171.

Todd, P. M., and Gigerenzer, G. (2020). "The ecological rationality of situations: behavior = f(adaptive toolbox, environment)," in *The Oxford Handbook of Psychological Situations*, eds J. F. Rauthmann, R. A. Sherman, and D. C. Funder (New York, NY: Oxford University Press), 143–158.

Todd, P. M., and Gigerenzer, G. E. (2012). *Ecological Rationality: Intelligence in the World*. New York, NY: Oxford University Press.

Tønnessen, M. (2018). The search image as link between sensation, perception and action. *Biosystems* 164, 138–146. doi: 10.1016/j.biosystems.2017.10.016

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136.

Uexküll, J. V. (2010). *A Foray Into the Worlds of Animals and Humans (translated by JD O'Neil)*. Minneapolis, MN: University of Minnesota Press.

Viale, R. (2020). "Why bounded rationality," in *Routledge Handbook of Bounded Rationality*, ed. R. Viale (Minneapolis, MN: Routledge), 1–54.

Volz, K. G., and Gigerenzer, G. (2012). Cognitive processes in decisions under risk are not the same as in decisions under uncertainty. *Front. Neurosci.* 6:105. doi: 10.3389/fnins.2012.00105

Volz, K. G., Schooler, L. J., Schubotz, R. I., Raab, M., Gigerenzer, G., and Von Cramon, D. Y. (2006). Why you think Milan is larger than Modena: neural correlates of the recognition heuristic. *J. Cogn. Neurosci.* 18, 1924–1936. doi: 10.1162/jocn.2006.18.11.1924

Weber, E. H. (1834). *De Pulsu, Resorptione, Auditu et Tactu*. Leipzig: Koehler.

Winter, S. G. (2000). The satisficing principle in capability learning. *Strateg. Manag. J.* 21, 981–996. doi: 10.1002/1097-0266(200010/11)21:10/11<981::aid-smj125>3.0.co;2-4

Wixted, J. T. (2020). The forgotten history of signal detection theory. *J. Exp. Psychol.* 46, 201–230. doi: 10.1037/xlm0000732

Wolfe, J. M. (2021). Guided search 6.0: an updated model of visual search. *Psychon. Bull. Rev.* 28, 1–33. doi: 10.3758/s13423-020-01859-9

Wolfe, J. M., and Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nat. Hum. Behav.* 1, 1–8.

Yarbus, A. (1967). *Eye Movements and Vision*. New York, NY: Plenum Press.

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.

Copyright © 2022 Felin and Koenderink. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fpsyg.2022.815691

Embodied Rationality Through Game Theoretic Glasses: An Empirical Point of Contact

*Sébastien Lerique**

Embodied Cognitive Science Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

The conceptual foundations, features, and scope of the notion of rationality are increasingly being affected by developments in embodied cognitive science. This article starts from the idea of embodied rationality, and aims to develop a frame in which a debate with the classical, possibly bounded, notion of rationality-as-consistency can take place. To this end, I develop a game theoretic description of a real time interaction setup in which participants' behaviors can be used to compare the enactive approach, which underlies embodied rationality, with game theoretic approaches to human interaction. The Perceptual Crossing Paradigm is a minimal interaction interface where two participants each control an avatar on a shared virtual line, and are tasked with cooperatively finding each other among distractor objects. It is well known that the best performance on this task is obtained when both participants let their movements coordinate with the objects they encounter, which they do without any prior knowledge of efficient interaction strategies in the system. A game theoretic model of this paradigm shows that this task can be described as an Assurance game, which allows for comparing game theoretical approaches and the enactive approach on two main fronts. First, accounting for the ability of participants to interactively solve the Assurance game; second, accounting for the evolution of choice landscapes resulting from evolving normative realms in the task. Similarly to the series of paradoxes which have fueled debates in economics in the past century, this analysis aims to serve as an interpretation testbed which can fuel the current debate on rationality.

Keywords: Team Rationality, Perceptual Crossing, Game Theory, Assurance game, Participatory Sense-Making, social awareness, Linguistic Bodies

OPEN ACCESS

Edited by:
Shaun Gallagher,
University of Memphis, United States

Reviewed by:
Marek Pokropski,
University of Warsaw, Poland
Enrico Petracca,
University of Bologna, Italy

***Correspondence:**
Sébastien Lerique
sebastien.lerique@oist.jp

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *15 November 2021*
Accepted: *07 March 2022*
Published: *11 April 2022*

Citation:
Lerique S (2022) Embodied Rationality Through Game Theoretic Glasses: An Empirical Point of Contact. Front. Psychol. 13:815691. doi: 10.3389/fpsyg.2022.815691

1. INTRODUCTION

The conceptual foundations, features, and scope of the notion of rationality are increasingly being affected by developments in embodied cognitive science. This article starts from the idea of Embodied Rationality (Gallagher, 2018; Rolla, 2021) which, among the array of proposals bringing embodied cognition and rationality together, stands out with the following features (Petracca, 2021): (1) it is the most radical, both philosophically and in terms of its departure from Simon's original bounded rationality (Simon, 1956); (2) no empirical studies have yet been developed to support, falsify, or otherwise empirically distinguish it from other approaches—so far, the case for embodied rationality has been made at the conceptual and philosophical levels; (3) it connects with

two issues that render it relevant across most domains in which rationality is currently discussed, namely, the scales of agency, and the dialectical evolution of normative realms.

Indeed, cognitively inspired modifications to the notion of rationality have traditionally entered the debate under the rubric of bounded rationality, separated in two different strands (Ross, 2014). On one side the psychology-driven tradition, which has convincingly shown the inadequacy of modeling an agent as capable of perfect predictions obtained using boundless resources. This tradition inherits from Simon's bounded rationality (Simon, 1956) and Gigerenzer's ecological rationality (Gigerenzer and Selten, 2002), and conceives rationality as successful adaptation to real-world tasks and situations. On the other side the economics tradition, interested in modeling collective behavior such as markets at the aggregate level, discusses rationality in terms of consistency between preferences, decision, and action. While this normative framework was originally developed for the individual level, underlying Rational Choice Theory, and further used by Kahneman and Tversky as the reference against which cognitive distortion, risk aversion and framing effects were evaluated (Kahneman, 2003), Becker (1962) argued early on that models of collective behavior need not make strong assumptions on the rationality or irrationality of individual agents: for results at the collective scale, it makes sense to approximate away from the details of psychological processes which may cancel each other out. Rubinstein (1998)'s seminal work makes a similar move for individual-level bounded rationality, providing a case-by-case evaluation of the relevance and effects of bounded mechanisms in models of collective behavior.

While notions of rationality have long been fragmented and debated, this conceptual divide seems to underpin the surprising idea that no matter the breadth of phenomena observed in psychology and behavioral economics, effects can be abstracted away or selectively added to otherwise unaffected premises of models of collective behavior. Infante et al. (2016), for instance, show that behavioral economists have largely adopted a dualist model of economic agents made of a rational core inside a psychological shell: the preferences of the shell can be revealed by traditional field experiments, but must then be purified in order to reveal the true, stable preferences of the rational core, which can therefore be used in economic models. At this point it is worth noticing the role that underlying metaphors of the mind play in the debate. Petracca (2021) groups the range of bounded rationality approaches into four, increasingly radical notions (Embodied Bounded Rationality, Body Rationality, Extended Rationality, and Embodied Rationality). The first three, which together cover the bounded rationality approaches presented above, remain broadly compatible with the computational metaphor of mind, albeit with increasing constraints[1]. This persistence of computationalist roots is likely to have played a role in the sedimentation of this conceptual divide: a set of models compatible with a qualified computational metaphor of mind, as the majority of bounded rationality approaches seem to be, can more easily be approximated as variations of a general set of premises (the ones underlying rationality-as-consistency), than an epistemologically more varied set of models.

Thus, by explicitly dropping computationalism and representations, and calling for a broader redefinition of rationality, embodied rationality (Gallagher, 2018; Rolla, 2021) provides a genuinely new element in the debate. Indeed, while the lineage of embodied rationality makes it directly relevant to the adaptive tradition of bounded rationality (see again Petracca, 2021), it would be a mistake to consider that modeling aggregate behavior under embodied rationality assumptions can be done similarly to, and with the same abstractions as, models built on classical bounded rationality-as-consistency[2]. Gallagher et al. (2019) and Petracca and Gallagher (2020), for instance, propose of view of markets as economic cognitive institutions, whereby "what is distributed in the market is not only information but also artifacts, routines, practices, social interactions, and affordances" (Petracca and Gallagher, 2020, p. 15), all "resources" which contribute to, and are affected by, the scales at which agency operates, the scaffolding of cognition and of interactions, autonomy, and ultimately becoming.

However, in order to trigger such a broad reevaluation of the abstractions underlying economic models, and advance to a workable understanding of the co-constitution of minds and collective behavior (markets seen as cognitive institutions being a case in point; Rizzello and Turvani, 2000 and Gallagher et al., 2019, p. 16), this argument also needs to be made at the level of models. This article aims to develop a frame in which such a debate can take place, that is, an empirical, model-friendly point of contact between embodied rationality and the classical, possibly bounded, notion of rationality-as-consistency. Similarly to the series of paradoxes which have fueled the debate on rationality in the past century, this point of contact aims to fuel the current debate by serving as an interpretation testbed.

I propose to do this by providing a new, game theoretic description of a well-studied sensorimotor interaction setup known as the Perceptual Crossing Paradigm (PCP). Through a series of lightweight approximations and empirically grounded assumptions, I will show that participants in recent versions of the PCP face a variation of the *Assurance game*. This game can be seen as a team-centered version of the well-known Prisoner's Dilemma (Ross, 2021), and is known for eliciting behaviors which standard Game Theory cannot account for. Instead, accounting for behaviors in the standard Assurance game using the classical notion of rationality-as-consistency requires articulating rationality with different scales of agency, a move which is made possible in two different ways by Team Reasoning and Conditional Game Theory (Sugden, 1993; Bacharach et al., 2006; Stirling, 2012; Hofmeyr and Ross, 2019). Since Embodied Rationality is directly compatible with the standard, enactive account of PCP, the identification of game theoretic structures

[1]In short: Simon is a founding father of cognitivism; Kahneman and Tversky used "the axioms of logic and probability" as their normative benchmark; researchers in ecological rationality, while explicitly reducing reliance on internal models, "subscribe to Simon's computational program" in their understanding of "fast-and-frugal" heuristics (Petracca, 2021, p. 5).

[2]Or, for that matter, models built on the adaptive bounded rationality approaches compatible with a qualified computational metaphor of mind.

in PCP provides a common empirical testbed for Embodied Rationality, Team Reasoning and Conditional Game Theory to compete for accounts of well-established PCP results.

I will discuss how both Team Reasoning and Conditional Game Theory successfully account for the different scales of agency at play in PCP, and will conclude by focusing on a process which has not yet been usefully formalized: the emergence and evolution of normative realms, and the resulting evolution of the strategies landscape. I contend that this use of the PCP, bringing such different approaches to rationality within arm's reach of each other, opens a path for refining our views of rationality in a way that can change the overall division of labor in modeling both individual behavior and collective behavior such as markets.

1.1. Relevant Works

The Perceptual Crossing Paradigm is first introduced by Auvray et al. (2009) as a new approach to the classic TV-mediated mother-infant interaction paradigm of Trevarthen and colleagues (Murray and Trevarthen, 1985, 1986; Nadel et al., 1999; Soussignan et al., 2006). The setup provides a minimal interaction interface where two participants each control an avatar on a shared virtual line. Each participant is given a device to move their avatar (often using a marble computer mouse) and receive haptic feedback through mechanical vibration. Using this device, each participant can move their avatar to explore different objects on the shared virtual line: a static object, the other participant's avatar, and a shadow object that mirrors the other participant's movements at a fixed distance. The line and objects (including the avatars) are invisible, but touching any object on the line (including the other person's avatar) is felt as mechanical vibration on the participant's device. Each participant is tasked with finding the other participant's avatar and clicking on it; the difficulty is therefore to distinguish between the other avatar and its shadow. **Figures 1**, **2** illustrate the experimental setup and virtual space.

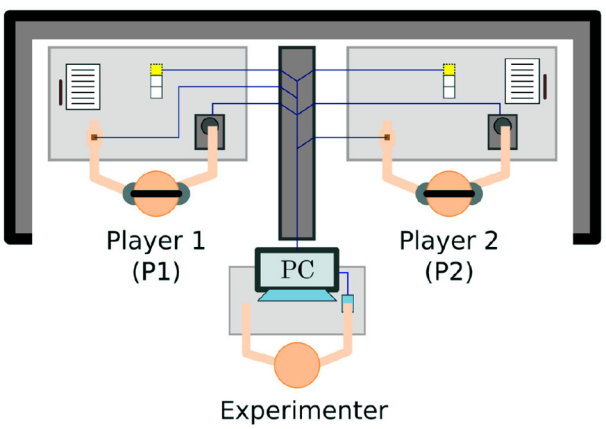


FIGURE 1 | Participants and experimenter in the Perceptual Crossing Paradigm. Copied from Froese et al. (2014a) under CC BY 3.0.

The main interest in the initial version of this setup (Auvray et al., 2009), aside from its simplicity, lies in the fact that participants solve the task collectively but not individually. On one side, participants do not seem able to individually distinguish between avatar and shadow, and on the other the final number of clicks on the other avatar is higher than on the shadow. Success is attributable to the inherent stability of avatar-to-avatar interactions, that is, to a property of the dyadic dynamics that creates more opportunities to click on the avatar than on the shadow. The versatility of the setup has led to a decade of profuse study of the conditions influencing participants' behavior and performance on the task, and in particular the conditions that enable participants to develop a sense of social presence; I review these works in the next section. As a result, the setup has established itself as a major tool for exploring sensorimotor-based interaction dynamics, alongside other setups studying coordinated behavior and cooperation-based performance (Reed et al., 2006; Kelso, 2008; Nordham et al., 2018).

The theoretical understanding of PCP results mainly relies on the developments of Participatory Sense-Making in the context of the Enactive Approach (De Jaegher and Di Paolo, 2007; Froese and Di Paolo, 2011). Participatory Sense-Making describes the enactment of systems of multiple autonomous agents which go from individually regulating their interaction with their environment with respect to their own norms and identity (sense-making), to coordinated regulation,

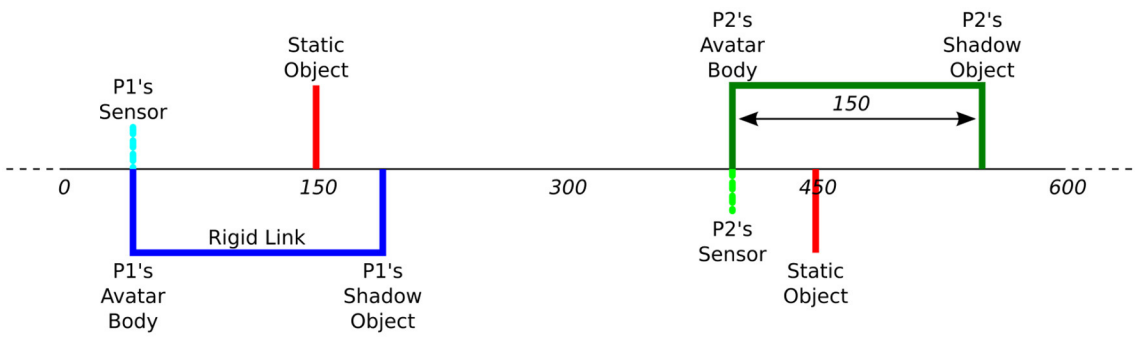


FIGURE 2 | Virtual space in the Perceptual Crossing Paradigm, including the participants' avatars, static objects, and shadow objects. Note that each participant can feel only one static object, as indicated by the positioning of their sensor. Copied from Froese et al. (2014a) under CC BY 3.0.

with other autonomous systems, of their interactions with their environment (participatory sense-making). As a result, "individual sense-making processes are affected and new domains of social sense-making can be generated that were not available to each individual on her own" (De Jaegher and Di Paolo, 2007, p. 497). It is important to note that this notion does not presuppose any social perception or pro-sociality. Auvray et al. (2009)'s initial version of the PCP is therefore a paradigmatic example of Participatory Sense-Making. When such interactions include a social component proper, they transition from Participatory Sense-Making to Social Agency, that is *co-regulation* of agents' interactions with their environment, that is of their sense-making (Di Paolo et al., 2018, p. 145). Social Agency obtains when participants in an interaction not only affect each other's environments (and thereby the conditions of their sense-making), but directly participate in each other's regulation of interaction with the environment, that is in each other's sense-making. As we will see, the more recent versions of PCP which were designed to elicit Social Agency are well-understood with this tooling, and exhibit other characteristic features of the Linguistic Bodies approach (Di Paolo et al., 2018): the existence of partial acts, and the dialectical dynamics of meaning due to evolving tensions between individual and interactive levels of normativity.

Drawing a link between Participatory Sense-Making and Linguistic Bodies on one side, and notions of rationality on the other, may seem challenging at first. On the enactive side, Embodied Rationality and Radically Enactive Rationality (Gallagher, 2018; Rolla, 2021) develop ways of thinking about rationality rooted in bodily performance. Future work will hopefully integrate the idea of rationality under enactivist hermeneutics with the Linguistic Bodies approach, accounting for the emergence of shared realms of rationality similarly to languaging. While such an integration has not yet been fleshed out, from here on I will take Embodied Rationality as the main notion of rationality associated with the enactive approach, and therefore with Participatory Sense-Making and the Linguistic Bodies approach.

Starting from the other side of the crevasse, Game Theory initially seems to be the obvious tool for analysing interdependent dyadic behavior in the rationality-as-consistency framework, and some attempts have been made to apply it to the study of coordinated joint action (Engemann et al., 2012). However, missing in game theory is the capacity to think about rationality at the level of the dyad, as can be seen in its failure to account for empirical results on the Hi-Lo game or the Assurance game (Ross, 2021). For this task, the most promising approaches are, on one side, Team Rationality, and on the other, the extension of Game Theory into Conditional Game Theory.

Robert Sugden seems to have been the first to argue that rationality at the level of a team is worth thinking about in the context of games and economic models. Criticizing approaches such as Schelling's theory of focal points (Schelling, 1960) for introducing external factors instead of expanding the notion of rationality itself, Sugden proposed that some games should be analyzed by asking how people rationally think when considering themselves part of a team (Sugden, 1993, 2003; Bacharach et al., 2006). The approach allows for solutions to standard problems such as the Hi-Lo game and the Assurance game which will reappear throughout this paper. It also dovetails with a broader proposal for a new form of normative economics built on the idea of sets of mutually acceptable market opportunities, instead of individual preferences, thus avoiding the common normative economics pitfall of considering agents mistaken in their unreliable preferences (Sugden, 2018).

A second approach to team phenomena is found in the recent extension of Game Theory into Conditional Game Theory (Stirling, 2012). This approach provides a framework for modeling situations where the preferences of some agents depend on the preferences of other agents. *Preference* conditioning, modeled over an acyclic network of influences between agents, goes beyond the simple interdependence of *choices* that is common in traditional Game Theory. Indeed, an agent's preferences are allowed to change depending on the preferences of influencing agents, after which choices are then made. This formalism also provides solutions to the Hi-Lo and Assurance games, while remaining compatible with traditional Game Theory in non-conditional situations.

While so far the two approaches seem to peacefully coexist (e.g., Lecouteux, 2018; Stirling and Tummolini, 2018; Ross, 2021), Hofmeyr and Ross (2019) correctly note that Conditional Game Theory provides a more general formalism which can also be used in cases of non-aligned groups. However, this observation misses Sugden's broader project of developing a form of normative economics which, by decoupling preferences from opportunities, need not bracket away the results of behavioral economics (Sugden, 2019). The debate is far from over, as the recent extension of Conditional Game Theory to cyclical influence networks (Stirling, 2019) introduces the proposal to the realm of (for now Markovian) stochastic processes. While beyond the scope of this article, this may in turn have relevance for a discussion with the heavily dynamical Linguistic Bodies approach.

How (in)compatible could Embodied Rationality, Team Rationality and Conditional Game Theory be, were they to find empirical applications in which to compare them? What notion of rationality would emerge from a beneficial exchange between these three theories? These are the two questions that I aim to bring into reach by looking at possible game theoretic structures in the PCP. I now start by reviewing previous studies and established results.

2. THE PERCEPTUAL CROSSING PARADIGM

2.1. Experimental Setup

Let us first name the participants in a PCP experiment: Alice and Bob. Recall that for both Alice and Bob, the virtual line is populated with three objects:

1. a static object, whose position is fixed and does not move throughout the whole experiment; there is one static object per participant, and each participant can only feel their own static object;

2. the other avatar, which moves along the line as controlled by the other participant; when Alice and Bob's avatar are touching each other, both Alice and Bob receive haptic feedback;
3. a copy of the other participant's avatar, which is maintained at a fixed distance of their avatar, mirroring its movements; when Alice touches Bob's shadow avatar, Alice receives feedback and Bob does not (and reciprocally when Bob touches Alice's shadow).

The setup can be seen as a toy model for common interaction situations, for instance mutual eye gaze. In this analogy, the avatar-touches-avatar interaction has a similar structure to two people looking mutually at each other in the eyes, whereas the avatar-touches-shadow interaction is analogous to looking at someone who is looking away.

In this setup however, participants are only informed that there is a static object, a moving object, and the other person's avatar. Participants do not know, therefore, that the moving object that is not the other avatar is in fact mirroring the movements of the other participant. The setup is therefore closer to the mother-infant TV-mediated interaction setup introduced by Trevarthen (Murray and Trevarthen, 1985, 1986; Nadel et al., 1999; Soussignan et al., 2006), from which it was originally inspired.

2.2. Success Is Joint

Each participant is then tasked with finding the other participant's avatar in the virtual space, and clicking on it when they believe they have found it. In the original version introduced by Auvray et al. (2009), participants are trained in specific situations, and then have 15 min with short breaks to explore the space and interact, clicking as many times as they see fit.

Initially, the main interest in this setup is the combination of collective success and individual failure in solving the task. On one side, participants do not seem able to individually distinguish between avatar and shadow: the probability that they will click after an encounter with the avatar is not significantly different from the probability of clicking after an interaction with the shadow. Yet the final number of clicks on the other avatar is higher than on the shadow. The reason is that encounters with the other avatar are more frequent, due to a higher stability of the interaction: when the two avatars touch each other, both participants will come back on their steps and oscillate around each other; whereas when an avatar is touching a shadow, the other participant receives no feedback relating to this contact, and will therefore not engage in maintaining the interaction. As Auvray et al. (2009) put it: "If the participants succeeded in the perceptual task, it is essentially because they succeeded in situating their avatars in front of each other." The setup therefore elicits success in a minimal task which can only be explained by the dynamics at the level of the dyad.

2.3. Social Awareness and Turn-Taking

The years following this work then chiefly focused on the question of what this behavior elicits about social cognition (Di Paolo et al., 2008; De Jaegher et al., 2010; Lenay et al., 2011; Auvray and Rohde, 2012; Froese et al., 2012; Lenay, 2012), and what minimal change to the setup could test the strong interpretation according to which social cognition can be partly constituted by social interaction (Michael, 2011; Herschbach, 2012; Michael and Overgaard, 2012; Overgaard and Michael, 2015).

Following Froese and Di Paolo (2011) and Froese et al. (2014a) then introduced a modification to the setup in order to make the task explicitly cooperative. First, participants are asked to cooperate and help each other find their avatars. Second, instead of a single long session in which participants can click without limits, the design is switched to 10–15 1-min long sessions, during which each participant is allowed a single click. Together, the two participants form a team in a tournament, playing against the other pairs of participants passing the experiment. The number of accurate and inaccurate clicks lets experimenters assign a post-experiment score to each team, and declare which pair of participants wins the tournament. Finally, experimenters introduce a questionnaire concerning each participants' clarity of perception of the presence of each other, using the Perceptual Awareness Scale (PAS; Ramsøy and Overgaard, 2004; Sandberg et al., 2010). PAS ratings for each interaction session go from 1 to 4, answering the following question: "Please select a category to describe how clearly you experienced your partner at the time you clicked: (1) No experience, (2) Vague impression, (3) Almost clear experience, (4) Clear experience."

Framing the task as cooperative and making clicks a scarce resource led participants to spontaneously develop a new way of coordinating their behavior, namely turn-taking. Alice would oscillate around Bob while Bob remained static, and the roles would then be repeatedly swapped. This mutually regulated behavior first led participants to more accurately click on each other's avatars. Second, it confirmed the hypothesis that social cognition is partly constituted by social interaction: PAS ratings and turn-taking levels showed that participants developed first-person awareness of each other's presence during coordinated interactions.

2.4. Emergence of Coordination

Later analyses describe the way in which dyadic coordination emerges over successive trials in the form of turn-taking. This learning process is associated with an increase in social awareness as measured by PAS ratings, an increase in the proportion of trials in which both participants make successful clicks (Froese et al., 2014b), and an increase in the time spent with the other participant's avatar instead of the distractors (Hermans et al., 2020).

Inside trials, the emergence of social awareness has been associated with increased movement coordination as measured by cross-correlation and windowed cross-lagged regression between participants' movement time series. Stronger social awareness has also been linked to longer time lags in movement coordination, meaning that trials in which higher social awareness is achieved are likely to see participants coordinating and taking turns on a longer time scale than in trials with lower social awareness (Kojima et al., 2017). The precise dynamics leading a participant to click have also been shown to alternate passive and active stimulation time frames: in the second

preceding a high social awareness click, information flows mostly from the person about to be clicked on, toward the person about to click, and the pattern is reversed with increasing strength up to 10 s after the click. In other words, high social awareness at the moment of a successful click is not an achievement of the person developing the awareness, nor of the other participant on their own; it is again a dyadic achievement (Kojima et al., 2017).

2.5. Shared Acts

The combined results of this research show that the task given to the participants is most successfully solved when both participants enter together in a coordinated shared act: Alice will detect Bob if Bob explores Alice, which he will do if Alice explores Bob, and so on. While the capacity for this kind of shared act emerges gradually over the trials, the end result can be well described in the framework of partial and shared acts developed by the Linguistic Bodies approach (Di Paolo et al., 2018). Let us then take a first step in abstracting out the structure of the interactions that take place in this paradigm. When Alice encounters an object, her exploring it will constitute a partial act of oscillatory stimulation. If Alice is faced with the shadow or the static object, the stimulation she will then receive (or lack thereof) will not allow for stable turn-taking to emerge. If the object is in fact Bob's avatar, Bob may respond to the received stimulation by a stimulation whose characteristics (rhythm, timing, duration) may constitute it as an appropriate response to Alice's partial act. This would lead to stable interaction dynamics where participants take turns in exploring each other, with increasing levels of social awareness. Bob may also, however, not respond appropriately or not respond at all, in which case Alice's partial act will be left unanswered, and the shared act fails.

In this context, the results presented so far indicate that an answer to such a partial act will have higher chances of success if it imitates the stimulation received, allows enough time for the partial act to be made, and allows for stable turn-taking to settle in. At this point in the interaction, both participants' movements strongly depend on each other, shared action is continuously being entertained, and social awareness will emerge.

Recent work has shifted toward investigating how strongly shared this kind of act is or can be (Froese et al., 2020; Hermans et al., 2020), and how variability across people enables it or hinders it (Zapata-Fonseca et al., 2018, 2019). For instance, Hermans et al. (2020) introduce a new measure of subjective experience and show that it is stronger in cases in which both participants click successfully, compared to cases in which neither participant clicks successfully, or only one of them does.

Beyond joint success, Froese et al. (2020) explored the basis for such social awareness. On one side, this could be a simple coordination behavior which allows the pair to enter a region of the dyadic phase space which is otherwise not attainable (*weak genuine intersubjectivity*, in the terms of Froese et al., 2020). On the other side, it could be the result of an event that is in some strong sense shared across the two participants, and merely reflected in their individual experiences of each other (*strong genuine intersubjectivity*). Indeed, in data reanalyzed by Froese et al. (2020), over 21% of the joint success trials show participants clicking within 3 s of each other. In other words, not only do participants develop social awareness of each other, they do so nearly at the same time. Froese et al. (2020) show that short inter-click delays are associated with higher individual and joint success, but only indirectly associated with higher subjective experience (PAS) of the other participant, such that the question of a single experience shared across the two participants is not yet settled.

Taking a step back, and temporarily setting aside the question of the intensity of intersubjectivity, it should now be clear that the structure of opportunities in which participants find themselves is very reminiscent of situations that are well-studied by Game Theory. As we will see, engaging in cooperation also bears a cost for players, and reaching joint success can also be seen as the result of participants navigating an action-dependent cost-benefit landscape, both individually and collectively.

In what follows I will propose a description of the PCP in the language of standard discrete Game Theory, and explore how previous results and open questions are rendered in the Game Theory framing. The shared action structure, in particular, appears at different time scales in the PCP and cannot be explained using traditional Game Theory only. On the other hand, Conditional Game Theory and Team Rationality can both account for the shared action structure of PCP, making this feature a useful contact point with the Linguistic Bodies approach.

3. A GAME-THEORETIC DESCRIPTION OF THE PCP

3.1. Framing the Task

We use the social agency version of the PCP task, as introduced by Froese et al. (2014a), where participants are presented as being part of a team, asked to click on each other *and* help each other succeed in doing so, but are otherwise not informed of any strategy for coordinating or succeeding at the task. Let us now simplify this task so that it can be framed, first, in the language of Decision Theory, and second, in the language of Game Theory. As a participant explores the space with their avatar, each stimulation received signals an encounter with one of the three objects in the space: the static object, the shadow (recall that the participant is not aware of the shadowing behavior), or the other participant's avatar. With no additional knowledge of the task, prior probabilities for an encounter with each of these objects are initially 1/3, and participants need to find their partner given two limited resources: (i) exploration time, and (ii) a single click. Each encounter can then be seen as two parallel decisions under uncertainty: whether or not to engage with the object at hand (if so, spending time to probe it and attempt to determine its nature), and whether or not to click.

We then make two important approximations. First, since a decision to click will formally end the primary task given by the experimenter (viz, clicking on the other), we set aside the click/no-click decision and focus on the decision about whether or not to engage with an encountered object, and if

yes, how[3]. This keeps us free from too complex models where the uncertainties due to the two parallel tasks would interfere with each other, and lets us focus on the dynamics of the exploration-interaction task. At this point, the task can be more simply worded as "detect your partner in the space." Our second approximation concerns the complexity of perceptual mistakes in this latter task. Indeed, a participant can make two types of errors in deciding whether an encountered object is their partner: thinking the object is their partner when it is not (type I error), and thinking it is not when in reality it is (type II error). Taking both these errors into account would require different probabilities for each error, such that decisions would be evaluated using two parallel and possibly conflicting criteria (one for each type of error to minimize). Instead, we set aside type I errors: our model assumes that when a participant believes they have found their partner, they are always right. In other words, a participant will never believe they have found their partner without actually having found them. Note that this in no way reduces the difficulty of the task, as the limited resource of exploration time is still present, and participants must still avoid type II errors: they may fail to perceive their partner if the interaction does not unfold well, or if the partner does not interact. At this point we can reword the task as "*find* your partner in the space," which translates to a single continuous decision under uncertainty, which will now be possible to model: whether or not to engage with an encountered object, and if yes, how.

Finally, we discretize the situation. A perfect description of this task in the game theoretic framework would require us to take into account (i) the fact that the space of available decisions is continuous (rendering it a *continuous game*), (ii) the fact that decisions are continuously taken over time (possibly requiring the theory of *differential games*), and (iii) the long term memory involved in each decision. Such an analysis is beyond the scope of this paper, and we instead discretize each encounter in the following way. First, we reduce the timing of participants' choices to repeated discrete decision moments. Second, we approximate the space of possible strategies (which involve all variations between leaving, sensing, and actively interacting) to two options: (i) leaving or passively sensing (waiting to see if the other object explores my avatar), or (ii) interacting actively. In broad terms, these strategies are equivalent to (i) engage less (and save time for later encounters), (ii) engage more (and invest time).

3.2. Decision Theory Is Insufficient

A first, naive approach to this task would model it as a parametric exploration-exploitation trade-off decision. Given an unknown object, we denote the ordinal costs of interacting with it as 1 and not interacting as 0, and the benefit of interacting as the probability p that this object is the other participant. Naturally, interactions in the immediate past with the object at hand will change the expected probability that this object is the other participant. One way of incorporating this is to estimate the benefit as the posterior probability given past interactions, $b(\text{past}) = p(x = \text{other}|\text{past}) = p(\text{past}|x = \text{other})\frac{p(x=\text{other})}{p(\text{past})}$. This cost-benefit situation is summarized in **Table 1**.

TABLE 1 | Simple framing of the PCP as a parametric decision problem.

	Benefit	Cost
Don't engage	0	0
Engage	b(past)	1

TABLE 2 | Choices faced by Alice, with corresponding payoffs and costs, were all the information available.

	Object of encounter				Cost
	Static	Shadow	Bob engaging less	Bob engaging more	
Engage less	0	0	1	2	0
Engage more	0	0	1	3	1

In this framing, a possible strategy would be similar to the idealized honey bee exploration-exploitation problem[4]: devise a method for exploring the space, and use a criteria to engage in interaction which should be monotonic with respect to the expected benefit of the interaction.

As the results presented in the previous section make clear, however, success is not a matter of individual decisions: what one participant does is constituted by what their partner does, a fact that can be made apparent in the simple approximation of b above. Using the case of turn-taking between Bob and Alice, we know that if Alice engages in a partial act, Bob may respond with more stimulation, such that $p(\text{stimulation received} \in \text{past}|x = \text{Bob})$ will be higher if Alice has engaged in stimulation in the past. In other words, incoming stimulation has a different meaning depending on whether Bob is interacting or not, and Bob will interact differently depending on whether Alice has engaged in interaction in the past or not. While interaction is required for participants to reduce the uncertainty concerning an object encountered, it comes at the cost of time. The main question in this task, then, is when to interact, knowing that the outcome essentially depends on one's partner.

It is clear that this situation is not captured by Decision Theory, in which decisions and payoffs do not depend on the actions of other participants. Here, each participant's payoff depends on what the other participant does, such that a game theoretic description of the situation is warranted.

3.3. Modeling an Encounter

While still not formally representing a game, **Table 2** provides a first representation of the partner-dependent choices faced by a participant, say Alice, if the nature and behavior of the object encountered were known.

The numbers in the table represent the decision cost and ordinal preferences over the outcomes associated with each decision, given the nature of the object encountered, and in

[3]Note that this is indeed an approximation, as participants sometimes explicitly click with uncertainty.

[4]A honey bee must decide whether to exploit a patch of flowers for which it knows the expected payoff, or explore the space to try and find a new patch of flowers which may or may not provide more payoff.

the case of an encounter with Bob, given Bob's strategy. When encountering the static object or the shadow, neither leaving nor engaging with it leads Alice to immediately find Bob, such that all ordinal preferences for the related outcomes are 0. When encountering Bob, the situation depends on Bob's strategy. If he is playing an "engage less" strategy, we consider that there is a slight possibility for Alice to detect Bob. However, we do not tie this possibility to the dynamics of the encounter nor to the time spent in the encounter (since Bob does not engage in it), so the ordinal preferences for the outcome with both strategies in the presence of non-engaging Bob can be set to 1. If Bob is engaging more, there is a higher likelihood of detecting him in both cases, but more so if Alice also engages in the interaction. The ordinal preferences for the outcomes are therefore 2 and 3. Whichever the actual object of encounter, engaging more bears the cost of time, which we initially represent as an ordinal cost of 1, compared to gaining time when not engaging, which we represent as an ordinal cost of 0.

Of course, during a real encounter the nature of the object is unknown, such that benefits and costs need to be combined to represent the choice under uncertainty that participants are faced with. Let us model the probabilities of detecting the other participant given their interaction strategy, and introduce parameters for the dependencies between the probability of each outcome.

First, let us label the "engage less" strategy $\mathcal{L}$, and the "engage more" strategy $\mathcal{M}$. Now, when Alice encounters an object *which in reality* is Bob, the probabilities that Alice detects Bob are as follows[5]:

- $\rho_{\mathcal{LL}}$, if both play $\mathcal{L}$
- $\rho_{\mathcal{LM}}$, if Alice plays $\mathcal{L}$ and Bob plays $\mathcal{M}$
- $\rho_{\mathcal{ML}}$, if Alice plays $\mathcal{M}$ and Bob plays $\mathcal{L}$
- $\rho_{\mathcal{MM}}$, if both play $\mathcal{M}$

We also introduce $\alpha \in [0, 1]$, the variable controlling Bob's strategy choices in the game: α is the probability that Bob plays $\mathcal{M}$, and $1 - \alpha$ the probability for him to play $\mathcal{L}$. Then let $p_{\mathcal{L}}$ be the probability of Alice finding Bob by playing $\mathcal{L}$, during an encounter with an unknown object. Conversely, let $p_{\mathcal{M}}$ be the probability of her finding Bob by playing $\mathcal{M}$ with an unknown object. Since the probability that the unknown object actually *is* Bob is $\frac{1}{3}$, we have:

$$p_{\mathcal{L}}(\alpha) = \frac{1}{3}\left((1-\alpha)\rho_{\mathcal{LL}} + \alpha\rho_{\mathcal{LM}}\right) \quad (1)$$

$$p_{\mathcal{M}}(\alpha) = \frac{1}{3}\left((1-\alpha)\rho_{\mathcal{ML}} + \alpha\rho_{\mathcal{MM}}\right) \quad (2)$$

Now, considering that engaging in interaction requires more time than not engaging in interaction, we are interested in comparing the probabilities of detecting the other participant with different strategies *at constant time cost*. Let us then introduce $\tau \in \mathbb{N}^*$, the ratio of time costs between engaging and not engaging in interaction: if sensing with $\mathcal{L}$ takes 1 s, sensing with $\mathcal{M}$ takes τ seconds[6]. To compare the two strategies at constant time cost, therefore, we look at the probability $P_{\mathcal{L}}$ that Alice will detect Bob by playing $\mathcal{L}$ during τ seconds: Alice can detect Bob during the first second with probability $p_{\mathcal{L}}$, and if not (probability $1 - p_{\mathcal{L}}$), then in a second encounter during the second, or a third encounter in the third second, and so on and so forth. Then the probabilities of Alice detecting Bob using each strategy *at constant time cost* are:

$$\begin{aligned} P_{\mathcal{L}} &= p_{\mathcal{L}} + (1-p_{\mathcal{L}})p_{\mathcal{L}} + (1-p_{\mathcal{L}})^2 p_{\mathcal{L}} + \cdots + (1-p_{\mathcal{L}})^{\tau-1} p_{\mathcal{L}} \\ &= p_{\mathcal{L}} \sum_{i=0}^{\tau-1} (1-p_{\mathcal{L}})^i \end{aligned} \quad (3)$$

$$P_{\mathcal{M}} = p_{\mathcal{M}} \quad (4)$$

Now bringing Equations (1) and (2) into Equations (3) and (4), we obtain the benefit g of playing $\mathcal{M}$ over playing $\mathcal{L}$, at constant time cost, as a function of α and the detection probabilities $\rho_{\mathcal{LL}}$, $\rho_{\mathcal{LM}}$, $\rho_{\mathcal{ML}}$, and $\rho_{\mathcal{MM}}$:

$$\begin{aligned} g(\alpha, \rho_{\mathcal{LL}}, \rho_{\mathcal{LM}}, \rho_{\mathcal{ML}}, \rho_{\mathcal{MM}}) = P_{\mathcal{M}}(\alpha, \rho_{\mathcal{ML}}, \rho_{\mathcal{MM}}) \\ - P_{\mathcal{L}}(\alpha, \rho_{\mathcal{LL}}, \rho_{\mathcal{LM}}) \end{aligned} \quad (5)$$

In order to render the exploration of this system of five variables palatable, let us add some final simplifications:

- let $u = \rho_{\mathcal{LL}} = \rho_{\mathcal{ML}}$ represent the probability of Alice detecting Bob, whichever Alice's strategy, during an encounter with Bob playing $\mathcal{L}$[7]; indeed, we can reasonably consider this probability to not depend on the duration of the interaction, since Bob's $\mathcal{L}$ strategy ensures there is indeed very little interaction, and trying to interact more time with him will not increase the probability of feeling him[8].
- let $w = \frac{\rho_{\mathcal{MM}}}{\rho_{\mathcal{LM}}}$ represent Alice's gain in playing $\mathcal{M}$ compared to $\mathcal{L}$, during an encounter with Bob playing $\mathcal{M}$.

4. GAMES AND STRATEGIES IN THE PCP

4.1. Encounter as an Assurance Game

Equipped with our model, and choosing values for u and τ[9], we can represent the benefit of playing $\mathcal{M}$ vs. $\mathcal{L}$ during an encounter as a function of three variables: $g(\alpha, \rho_{\mathcal{LM}}, w)$. The case $u = 0.04$

[5]Formally, these are the probabilities of avoiding a type II error.

[6]Thus, contrary to the Decision Theory model above, costs in this model have magnitudes instead of being ordinal.

[7]Recall that comparing $\mathcal{L}$ and $\mathcal{M}$ strategies *with an unknown object* and at constant time cost is done using $P_{\mathcal{L}}$ and $P_{\mathcal{M}}$, which will differ even though $\rho_{\mathcal{LL}} = \rho_{\mathcal{ML}}$. Indeed, playing $\mathcal{L}$ allows Alice to encounter several different objects (each of which may be Bob) in the same time cost as when encountering a single object and playing $\mathcal{M}$.

[8]Introducing a factor $\frac{\rho_{\mathcal{ML}}}{\rho_{\mathcal{LL}}} > 1$ does not qualitatively change the results, as long as it remains lower than a value s_{max} derived in the **Supplementary Material**. For $\rho_{\mathcal{LL}} = 0.04$ and $\tau = 3$ as introduced in Section 4, we have $s_{max} \approx 0.98$, that is $\rho_{\mathcal{ML},max} \approx 2.96\rho_{\mathcal{LL}}$. See the **Supplementary Material** for more details.

[9]These variables have the least effect on g. See the **Supplementary Material** for more details.

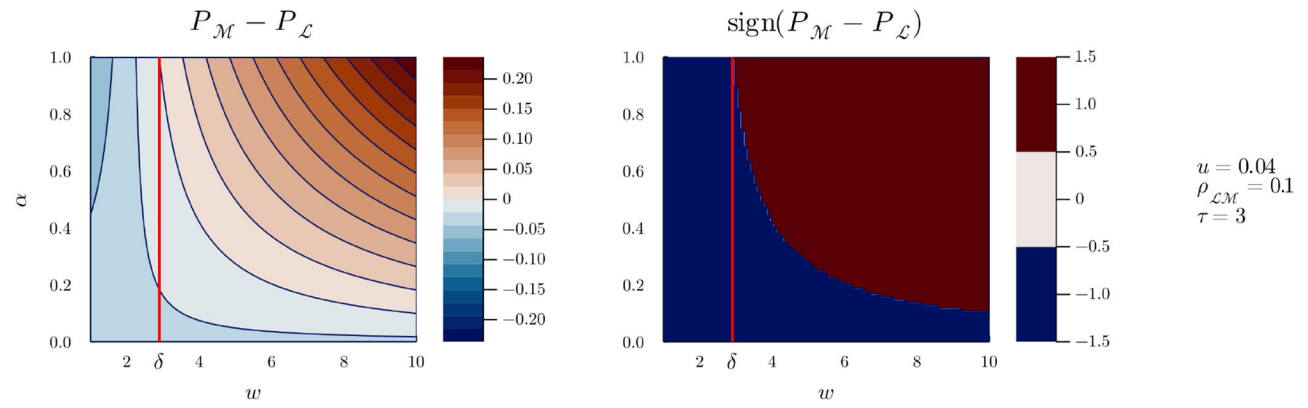


FIGURE 3 | $g(\alpha, w)$ for $u = 0.04$, $\rho_{\mathcal{LM}} = 0.1$, and $\tau = 3$.

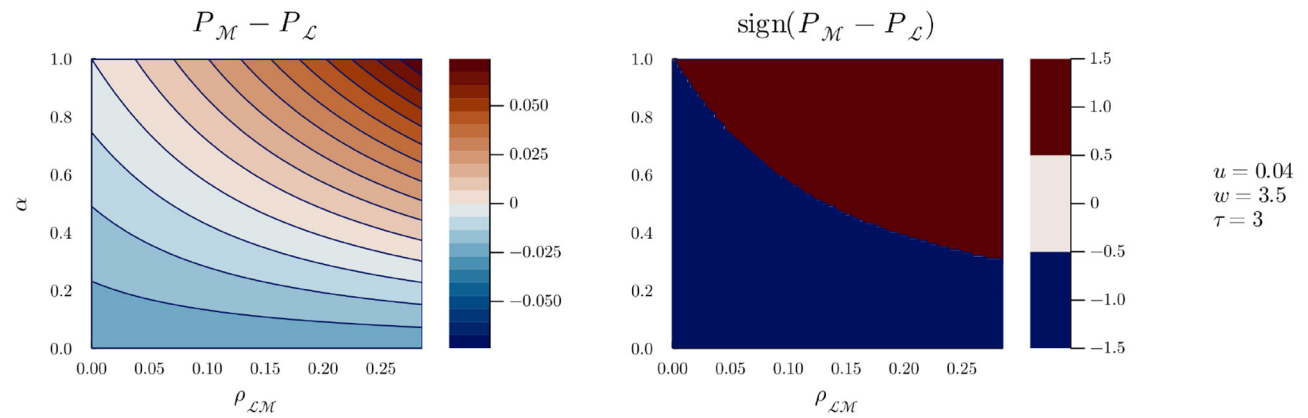


FIGURE 4 | $g(\alpha, \rho_{\mathcal{LM}})$ for $u = 0.04$, $w = 3.5$, and $\tau = 3$.

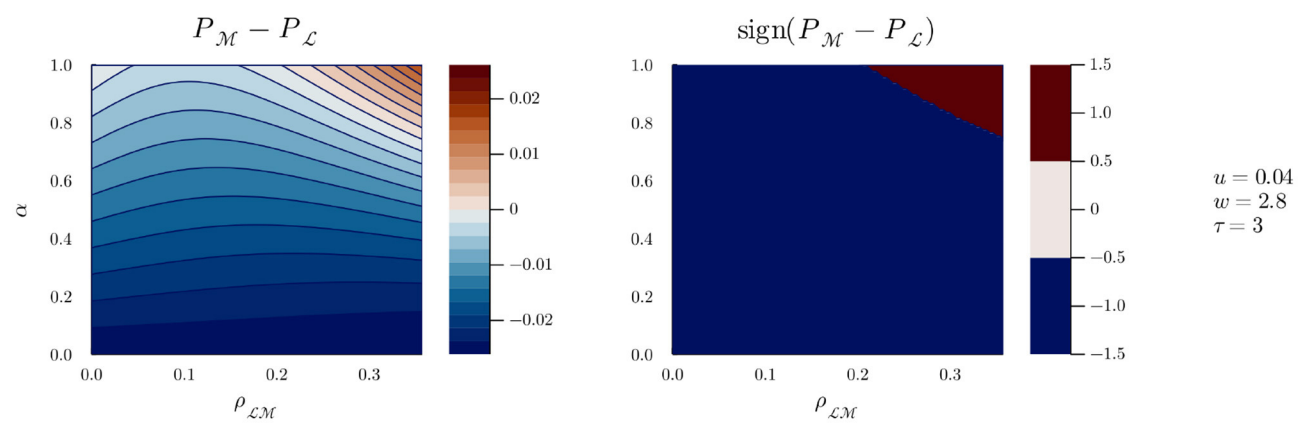


FIGURE 5 | $g(\alpha, \rho_{\mathcal{LM}})$ for $u = 0.04$, $w = 2.8$, and $\tau = 3$.

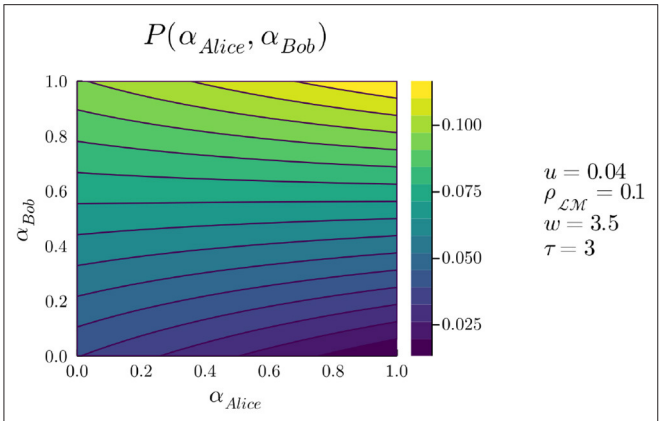


FIGURE 6 | $P(\alpha_{Alice}, \alpha_{Bob})$ for $u = 0.04$, $\rho_{\mathcal{LM}} = 0.1$, $w = 3.5$, and $\tau = 3$.

and $\tau = 3$ (i.e., $\mathcal{M}$ costs three times more time than $\mathcal{L}$), can be seen in **Figure 3**.

The left pane of **Figure 3** shows the values of g at constant $\rho_{\mathcal{LM}} = 0.1$ (i.e., encountering Bob who plays $\mathcal{M}$ leads to a 0.1 probability of detection if Alice plays $\mathcal{L}$), as a function of Bob's interaction strategy (α) and of the gain of playing more interaction if Bob also plays more interaction (w). The right pane represents the sign of g, as a function of the same variables. Colors closer to red (or simply dark red in the right pane) indicate higher values of g, that is, parameters for which Alice is more likely to detect Bob by playing $\mathcal{M}$. Conversely, colors closer to blue (or simply dark blue in the right pane) indicate parameters for which Alice is less likely to detect Bob by playing $\mathcal{M}$, that is she will be better off playing $\mathcal{L}$. It is clear from both panes that for $w < \delta$ with $\tau > \delta \approx 3$[10], Alice is better off always playing $\mathcal{L}$, whereas for $w > \delta$, there is a cutoff value of α above which Alice is better off playing $\mathcal{M}$. As w grows, the cutoff value for α decreases, and Alice is better off playing $\mathcal{M}$ even if Bob has a relatively low α.

If we now pick a value for w, say 3.5, we can inspect the evolution of g as $\rho_{\mathcal{LM}}$ varies, as can be seen in **Figure 4**. A similar pattern can be seen: for all values of $\rho_{\mathcal{LM}}$, there is a cutoff value of α over which Alice is more likely to detect Bob by playing $\mathcal{M}$. The effect of a lower value for w can be seen in **Figure 5**: the range of values of α and $\rho_{\mathcal{LM}}$ for which Alice is better off playing $\mathcal{M}$ is reduced, but does not disappear (it does, however, if w is further reduced). However, we can safely assume that when Bob plays $\mathcal{M}$, the likelihood of detecting him grows with time at least equally whether Alice plays $\mathcal{M}$ or $\mathcal{L}$; in other words, we can assume $w \geq \delta$. And without presupposing any result from previous work, we can further assume that when Alice plays $\mathcal{M}$, she is involved in some active sensing, and the likelihood of detecting Bob (playing $\mathcal{M}$) grows faster with time than when playing $\mathcal{L}$. This is equivalent to stating that $w > \delta$[11], such that there will always exist a value of α above which Alice is better off playing $\mathcal{M}$, and the situation represented in **Figure 5** should not be possible.

Given this knowledge, we can finally look at the expected benefits for Alice depending on the strategies she and Bob play. **Figure 6** represents the probability that Alice will detect Bob at constant time cost, given fixed values for u, $\rho_{\mathcal{LM}}$, w, and τ, as a function of α_{Alice} and α_{Bob}:

$$P(\alpha_{Alice}, \alpha_{Bob}) = \alpha_{Alice} P_{\mathcal{M}}(\alpha_{Bob}) + (1 - \alpha_{Alice}) P_{\mathcal{L}}(\alpha_{Bob}) \quad (6)$$

The plot first reflects what the previous figures indicated, when fixing α_{Bob} and inspecting Alice's options. For low values of α_{Bob}, Alice is better off playing with low α_{Alice}, that is favoring $\mathcal{L}$. Conversely, for high values of α_{Bob}, Alice is better off playing with high α_{Alice}, that is favoring $\mathcal{M}$. These patterns are maintained as long as $w > \delta$, which we have seen is a reasonable assumption. Second, it is also clear that if both players maximize their α values, the likelihood of Alice feeling Bob at constant time cost is much higher than if both players minimize their α. The situation is of course symmetrical for Bob.

Let us now come back to binarized strategy options for both players, in terms of "high α" and "low α". These two options correspond to favoring one or the other of $\mathcal{M}$ and $\mathcal{L}$, though without committing to one or the other entirely. This setting corresponds to dicing **Figure 6** into four quadrants. We see that Alice is worst off (dark blue) when Bob plays low α while Alice plays high α: Alice pays the cost of time by playing $\mathcal{M}$ while not getting any increase in probability of feeling Bob, since he plays $\mathcal{L}$. A slightly better situation (dark green) is obtained when both play low α, that is, while Alice does not have a high probability of feeling Bob due to the two $\mathcal{L}$ strategies, she at least reduces time cost and therefore increases the possibility of feeling Bob in other encounters. A yet better situation (light green) occurs when Bob plays high α and Alice plays low α, and the best probability (light yellow) is obtained when both play high α. The values represented here are probabilities at constant time cost, which can be taken as payoffs for game actions; thus we can represent the ordinal preferences for each outcome, now including all costs incurred (incorporated in the computation of the probabilities), as shown in **Table 3**. The situation is identical for Bob (so symmetrical in the table), and the full game is represented in **Table 4**.

This is not a Prisoner's Dilemma, but an Assurance game. If Bob plays low α (favoring $\mathcal{L}$ most of the time), Alice has the choice between playing high α and wasting time on encounters in which interaction is not reciprocated, or playing low α and

[10] As δ is defined by $g(1, \rho_{\mathcal{LM}}, \delta) = 0$, it is easy to derive that $\delta = \sum_{i=0}^{\tau-1} \left(1 - \frac{\rho_{\mathcal{LM}}}{3}\right)^i$. For $\rho_{\mathcal{LM}} = 0.1$ and $\tau = 3$, we have $\delta \approx 2.901$.

[11] Consider variations of $\rho_{\mathcal{MM}}$, $p_{\mathcal{M}}$, and $P_{\mathcal{M}}$, marked with a tilde, which we use to represent the probability of detecting $\mathcal{M}$-playing Bob by also playing $\mathcal{M}$, but in the same time as when playing $\mathcal{L}$. This lets us ask how probable it is to detect $\mathcal{M}$-playing Bob by interacting with him during the same duration as it would take to not interact. We have $\tilde{\rho}_{\mathcal{MM}} \geq \rho_{\mathcal{LM}}$ and $\tilde{p}_{\mathcal{M}}(1) = \frac{1}{3}\tilde{\rho}_{\mathcal{MM}}$ (recall $\alpha = 1$ since Bob plays $\mathcal{M}$). Now our first assumption in the main text is that $\tilde{P}_{\mathcal{M}}(1)$ follows the same form as $P_{\mathcal{L}}(1)$ or a form that grows faster with τ. In other words, $\tilde{P}_{\mathcal{M}}(1) \geq \tilde{p}_{\mathcal{M}}(1) \sum_{i=0}^{\tau-1} (1 - \tilde{p}_{\mathcal{M}}(1))^i$. Since $p_{\mathcal{M}}(1) = P_{\mathcal{M}}(1) \geq \tilde{P}_{\mathcal{M}}(1)$, we have $p_{\mathcal{M}}(1) \geq \tilde{p}_{\mathcal{M}}(1) \sum_{i=0}^{\tau-1} (1 - \tilde{p}_{\mathcal{M}}(1))^i$. Replacing with Equations (1) and (2) yields $w \geq \delta$. Our second assumption is that $P_{\mathcal{M}}(1) > \tilde{P}_{\mathcal{M}}(1)$, which in that case yields $w > \delta$.

TABLE 3 | Structure of the game faced by Alice.

		Bob	
		Low α	**High α**
Alice	Low α	1	2
	High α	0	3

TABLE 4 | Structure of the game faced by both Alice and Bob.

		Bob	
		Low α	**High α**
Alice	Low α	1, 1	2, 0
	High α	0, 2	3, 3

moving from encounter to encounter, betting on the possibility that in one of them Bob will be detectable. Alice is better off following Bob's strategy: low α. If Bob plays high α, Alice can choose to passively receive Bob's stimulation (low α), or reciprocate interactions (high α) in which case detection is much more likely. She is better off again following Bob's strategy, in this case high α.

As the situation is identical for Bob, it follows that the game has two Nash Equilibriums, which are the two situations in which both players pick the same strategy.

4.2. Repeated Encounters

The model developed here partly sets aside the repeated nature of the game. First, encounters occur repeatedly during a single trial[12], such that at this scale one can see the interaction as a repeated Assurance game which, in our approximation, ends for each player whenever they click.

More importantly, PCP experiments have repeated trials (going from 6 to 15 trials), over which participants learn about the space, the objects, and their interactions. Empirical results indicate players develop a stronger sensitivity and a more effective social interaction repertoire over time. In other words, over repeated trials interactions can become more effective, improving the signal-to-noise ratio and reducing uncertainty, which in the model is mainly represented by an increase in w. This becomes possible if players indeed engage in interactions, that is if they play high α. When played over repeated sessions therefore, the assurance game is reinforced: not only will players be more likely to find each other if they both play high α, but doing so from the start will even further increase the probability of detecting each other whenever they encounter each other, reducing the uncertainties and increasing the payoff associated with high α. Similarly to the session- and encounter-level games, if Alice plays this way but Bob doesn't, Alice will incur the cost of repeatedly playing high α without improvements in interactions (i.e., without increased w). A precise description of these dynamics requires more detailed modeling of the effects and costs related to learning over trials. While this is beyond the scope of this article, it seems likely that a similar structure could come to light, that is, another Assurance game could also describe the interaction at the scale of the experiment.

4.3. Summary of Results

Let us summarize the observations that can be made from this first description of PCP in the language of Game Theory.

We separated the PCP task into decisions about whether or not to click, and decisions about whether or not to interact. In order to focus on the decisions about interactive behavior, we set aside the decision about whether or not to click, and approximated the PCP as a situation in which participants never mistakenly think they are interacting with their partner (ignoring type I errors). In this approximation, the task is to "find" the other participant, type I errors (unsuccessful clicks) are ignored, and we focus on the relationship between type II errors (missed opportunity of a successful click) and interactions with an object that is not the other participant. This lets us concentrate on the structure of the "interact now or later" game which participants face, leaving further modeling of other aspects of the PCP for later work. We then assumed that the decisions participants are faced with in this game can be time-discretized into a series of "interact now vs. later" decisions, in which the option of interacting requires more immediate time investment than not interacting (or interacting less).

Next, we assumed the following approximations, which we take as a reasonable first approach to describing the PCP in the language of discrete Game Theory:

- the strategy of each player can be represented as a probability to interact or not interact (α), then later discretized to "low α" and "high α"
- interacting requires more time investment than not interacting, a relationship approximated with an integer factor (τ)

And we finally assumed the following relationships between the probabilities that can be defined given the approximations made thus far:

1. the probability of detecting a low α partner during an encounter with them, and whether interacting with them or not, is the same regardless of the time spent in the encounter ($\rho_{\mathcal{LL}} = \rho_{\mathcal{ML}}$)
2. when playing low α, the probability of detecting a high α partner is higher than the probability of detecting a low α partner ($\rho_{\mathcal{LM}} > \rho_{\mathcal{LL}}$)[13]
3. the probability of detecting a high α partner is higher when interacting than when not interacting [first because $\tilde{\rho}_{\mathcal{MM}} \geq \rho_{\mathcal{LM}}$, second because $P_{\mathcal{M}}(1) > \tilde{P}_m(1)$, and these combine to yield $w > \delta$]

These relationships are empirically supported. Previous work on the PCP has indeed shown that high perceptual awareness is associated with higher levels of turn-taking and behavior matching, is preceded by a period of passive stimulation, and is

[12]Recent versions of the paradigm set the trial duration to 1 min (Froese et al., 2014a).

[13]Note that results sometimes still hold in the case where this is not satisfied, but we opted for assuming this conservative hypothesis to make the analysis palatable.

also associated with longer interaction times (Kojima et al., 2017, in particular Figures 6–8).

Under these approximations and assumptions, it appears that the "interact now vs. later" decisions have the structure of an Assurance game, by which players maximize their likelihood of finding the other if they play the same strategy, and more so if they both choose to favor more interaction. While previous experimental work has extensively shown that mutual high α is without doubt the best team-level strategy to find each other, the Assurance game discovered here adds new light to the PCP. First, it shows that a mutual low α strategy is also a Nash Equilibrium, a fact that is only apparent when one takes into account the time cost of interaction and the small but non-zero probability of finding the other in mutual non-interaction. Second, it shows that investing in interaction alone bears a higher cost than the mutual low α equilibrium, which accounts for the difficulty of the task: choosing between investing time now, with a higher win-lose uncertainty, or leaving that uncertainty for a later moment in the trial.

This result provides us with an empirical point of contact between Embodied Rationality and the two approaches capable of accounting for team behavior in the rationality-as-consistency tradition: Team Rationality and Conditional Game Theory.

5. DISCUSSION

5.1. Standard Game Theory and Team Behavior

While the analysis process in Sections 3, 4 was couched in the language of Game Theory (consider the notions of choice, decision, strategy, or uncertainty), at this stage I am not suggesting that game-theoretic approaches are superior, or for that matter inferior, to other accounts of observed behaviors in PCP. Quite the contrary: given an empirical paradigm in which what seems like markedly human behaviors have been extensively documented, I aim to take the opportunity for conflicting accounts of human individual and group behavior to compete on a common ground. Two such approaches—Team Rationality and Conditional Game Theory—rely on the rationality-as-consistency framework and are therefore easy to assess in a game-theoretic framing. Besides, standard Game Theory itself *fails* to account for people's success in the Assurance game identified in PCP, such that the use of a game-theoretic description is really no more than a tool for rationality-as-consistency approaches to enter the debate. First then, let us see why standard Game Theory fails to account for PCP behavior, and set it aside.

The problem lies in the existence of several Nash Equilibriums in the Assurance game. Indeed, if the payoffs in a game are assumed to incorporate all the components of the preferences of players, and if that game then contains several Nash Equilibriums, standard Game Theory has no explanation for why an agent would prefer one equilibrium over another: by definition, all preferences have already been included in the derivation of the Nash Equilibriums. A good example of this problem appears in a recurrent yet misplaced criticism of the Prisoner's Dilemma game. As Ross (2021) describes it:

> Many people find it incredible when a game theorist tells them that players designated with the honorific "rational" must choose in this game in such a way as to produce the outcome [(defect, defect)]. The explanation [of the "rational" choice] seems to require appeal to very strong forms of both descriptive and normative individualism.

Ross (2021) continues, citing Binmore (1994):

> If players value the utility of a team they're part of over and above their more narrowly individualistic interests, then this should be represented in the payoffs associated with a game theoretic model of their choices.

In the case of the Prisoner's Dilemma, incorporating players' preferences for a more egalitarian outcome transforms the model into an Assurance game. Once this point is reached, standard Game Theory has no further tools to explain how agents choose one Nash Equilibrium over the other, even though the (Cooperate, Cooperate) equilibrium in the Assurance game is Pareto-optimal (Ross, 2021), and often chosen by people in practice. PCP is no exception here, and standard Game Theory is therefore ruled out as an account of well-documented behavior.

How, then, can convergence on the (Cooperate, Cooperate) equilibrium be accounted for? Both Team Rationality and Game Conditional Theory can answer this question. I will further propose an extension of ER, dubbed *Embodied Social Rationality*, which relies on the Linguistic Bodies approach to provide a third account of team behavior.

5.2. PCP Assurance Game Under Rationality-as-Consistency

Eschewing proposals that introduce components exogenous to rationality (such as heuristics or Schelling's notion of "focal points"), Sugden was the first to argue that accounting for team behavior should be done by extending the unit of agency. Thus, in cases where the existence of the team is already established, action are taken as part of a best-outcome plan for the team, subsuming the question of how to act depending on the action of one's partner (Sugden, 1993, p. 86):

> To act as a member of the team is to act as a component of the team. It is to act on a concerted plan, doing one's allotted part in that plan without asking whether, taking other members' actions as given, one's own action is contributing toward the team's objective. … It must be sufficient for each member of the team that the plan itself is designed to achieve the team's objective: the objective will be achieved if everyone follows the plan.

Team Rationality removes each player's concern for possibly detrimental moves from their partner: a team-member who does not follow their part of the plan is *team-irrational*. In this framework, rationality is not a matter of optimizing for individual preferences (which can therefore vary freely without this resulting in theoretical deadlocks), but a matter

of converging on mutually beneficial outcomes (Sugden, 2018, 2019). This form of reasoning can be illustrated in the PCP payoff landscape represented by **Figure 6**. If Alice is rational in the traditional, game-theoretic sense, she must consider Bob's strategy (α_{Bob}) as fixed, and her movements on the payoff landscape are restricted to horizontal lines. If Alice and Bob are team-rational, they are free to move *together* on the payoff landscape. In both cases, Alice and Bob know that the best mutually beneficial outcome would result from high α_{Alice} combined with high α_{Bob}. Yet in the first case, deciding under the assumption that the choice of the partner is fixed can prevent them from collectively reaching the best outcome, while in the second case they will each do their part in the concerted plan. The behavior of participants in the PCP Assurance game is thus understood using decision dynamics which span beyond individuals (Lecouteux, 2018). The role of a normative notion of individual preference, which has repeatedly been shown to conflict with empirical results (Infante et al., 2016), is also reduced.

Conditional Game Theory (Stirling, 2012, 2019) proposes a different account in which team agency is not needed, and for that matter need not exist. Instead, team behavior may emerge from the network of influences that agents' preferences exert on each other. Recall that a player's preferences are defined over the entire set of possible outcomes resulting from the actions of all players, such that conditioning on a player's preferences—instead of simply on their actions—substantially expands the dynamics possible in a Conditional Game Theory model. An analysis of the PCP Assurance game in this framework is beyond the scope of this discussion, yet the examples provided by Stirling and Tummolini (2018) and Hofmeyr and Ross (2019) for the Hi-Lo game suggest that the convergence of both participants on the "high α" behavior can be accounted for. This proposal has the additional benefit of applying to cases in which no team is established or payoffs are not as aligned as they are in the Hi-Lo and Assurance games. On the other hand, the way in which a player is influenced by another player's preferences may be a point of substantial variability across players. In particular, for an agent to obtain the actual (conditional) preferences of another agent influencing it, a fair amount of explicit communication or even computation may be required. This is in line with regular Game Theory's tradition of abstracting away from psychological details, but may render the conditional approach less applicable to the PCP case. By contrast, Team Rationality only requires players to be aware of the structure of the game, and consider themselves part of a team, both conditions which seem realized in the PCP.

5.3. The Evolution of Strategies

An important component of the enactive understanding of social agency in the PCP has so far not been addressed: the emergence and evolution of normative realms, that is, the horizon against which interactions are evaluated by participants. This notion encompasses both a participant's sensitivity to aspects of the interaction dynamics that take place and which they engage in, and their subjective valuation of such dynamics.

Contrary to the real Assurance game, strategy options in the PCP are open to change over time. Indeed, the existence of a strategy at a given point in time heavily depends on the history of interactions between the participants. After a small number of initial trials during which the framing from Sections 3, 4 is warranted, the PCP doesn't provide participants with fixed strategy options from which to choose. Instead, participants need to develop and stabilize their own set of dyadic interaction strategies. It is during this second phase of the experiment, once the initial strategies are being modified and tinkered with, that pairs of participants are able to develop social agency and genuinely perceive social presence. For instance, recent work shows that over successive trials the time spent with the other avatar increases (Hermans et al., 2020), along with an increase in the intensity of social awareness of the other (Froese et al., 2020), stronger levels of turn-taking and movement coordination, and longer interaction timescales (Kojima et al., 2017). The set of strategies to choose from at each encounter thus fluidly changes across trials as a result of the history of interactions in a pair.

The conceptual logic (if not the empirical unfolding) of this evolution is well explained by Participatory Sense-Making (De Jaegher and Di Paolo, 2007) and the Linguistic Bodies approach (Cuffari et al., 2015; Di Paolo et al., 2018). In a first step, two autonomous agents may maintain an initial contact without any pro-sociality, due to a stability related to each agent's sense-making process (i.e., each agent's regulation of self-environment interaction). In the PCP without click constraint, participants will come back on any object they sense, making the contact of two such agents stable over time. In a second step, a tension may emerge from the interference between the two agents' self-environment regulation processes. Indeed, at this point each agent's regulation process is active in an environment which includes the other agent, and therefore reacts differently to an environment from which the other agent is missing. In the PCP, this situation occurs when participants are constrained to a single click and the task is framed as cooperative: participants are more conservative with their click, and the cooperative framing may lead them to try and show themselves clearly to objects they encounter. This is the situation accounted for by the Assurance game.

Yet as agents actively explore different interaction dynamics, new co-regulation conventions emerge that solve the initial tension between the two agents' self-environment regulation processes. More elaborate stimulation of and reaction to the other participant's stimulation arises, marking the appearance of a co-regulation of the interaction. At this stage in the PCP, teams develop their own interaction conventions, associated with team-specific capacities for feeling each other, that is, a normative realm which sediments into a repertoire of shared acts: conventions which can be triggered by one participant (through a partial act) and call for an adequate response from the partner. Each shared act is a new form of meaningful interaction between partners, such that failing to respond adequately to a partial act can trigger new kinds of breakdown. Froese et al. (2014b) report the case of a participant feeling abandoned by their partner when an interaction was abruptly interrupted. On this view, such elaborate feelings result from the development and use of a repertoire of meaning-imbued shared acts, which constitute the new normative realm developed by the team.

The emergence of a repertoire of shared acts reconfigures the strategies that participants can use in each encounter: instead of remaining a fixed set, the strategies used by participants evolve over time, and are dependent on past interactions with the partner. This fundamental feature of the PCP, which underpins the emergence of a sense of social presence, cannot be explained by Conditional Game Theory or Team Rationality. Furthermore, abstracting the feature away for the purposes of models of collective behavior would negate the potential for evolution of choice sets as they emerge from agents' interactions themselves.

The time seems ripe, then, for a deeper comparison of Conditional Game Theory, Team Rationality, and the Linguistic Bodies approach. As the latter can account for fundamental features of the PCP which cannot be abstracted away by the former approaches, I believe that a second, deep comparison between the associated notions of rationality is also warranted: rationality-as-consistency, on one side, and Embodied Rationality, on the other. The analysis of the PCP presented in this paper shows that such comparisons are not only needed, but possible on empirical grounds.

6. CONCLUSION

In this article, I proposed a novel analysis of the PCP using the language of game theory. This analysis shows the existence of an Assurance game in the form of the "interact now-or-later" question that participants continuously need to solve. The existence of such a standard game in perceptual crossing sensorimotor interactions opened the door to comparing game theoretical approaches and the enactive theory of Participatory Sense-Making and Linguistic Bodies on two fronts. First, the capacity for participants to interactively solve the PCP Assurance game. Second, the evolution of choice landscapes resulting from the evolution of normative realms in the PCP. Finally, and most importantly, this work positions the PCP as an empirical meeting point between two radically different approaches to human interactions, namely, the economics tradition, interested in models of collective behavior such as markets, and the enactive approach.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. The Julia Pluto notebook used to generate these figures and explore the model is available at the following url: https://gitlab.com/wehlutyk/2021-10-shared-embodied-rationality-paper-public/-/blob/main/encounter-game.jl. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported by OIST Innovative Technology Research - Proof of Concept Program.

ACKNOWLEDGMENTS

The author would like to thank Tom Froese for the initial spur, and Soheil Keshmiri and Mark James for valuable discussions while writing the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.815691/full#supplementary-material

REFERENCES

Auvray, M., Lenay, C., and Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas Psychol.* 27, 32–47. doi: 10.1016/j.newideapsych.2007.12.002

Auvray, M., and Rohde, M. (2012). Perceptual crossing: the simplest online paradigm. *Front. Hum. Neurosci.* 6:181. doi: 10.3389/fnhum.2012.00181

Bacharach, M., Gold, N., and Sugden, R. (2006). *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton, NJ: Princeton University Press. doi: 10.1515/9780691186313

Becker, G. S. (1962). Irrational behavior and economic theory. *J. Polit. Econ.* 70, 1–13. doi: 10.1086/258584

Binmore, K. G. (1994). *Game Theory and the Social Contract, Vol. 1: Playing Fair*. Cambridge, MA; London: The MIT Press.

Cuffari, E. C., Di Paolo, E., and De Jaegher, H. (2015). From participatory sense-making to language: there and back again. *Phenomenol. Cogn. Sci.* 14, 1089–1125. doi: 10.1007/s11097-014-9404-9

De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making. *Phenomenol. Cogn. Sci.* 6, 485–507. doi: 10.1007/s11097-007-9076-9

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

Di Paolo, E. A., Cuffari, E. C., and De Jaegher, H. (2018). *Linguistic Bodies: The Continuity between Life and Language*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11244.001.0001

Di Paolo, E. A., Rohde, M., and Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas Psychol.* 26, 278–294. doi: 10.1016/j.newideapsych.2007.07.006

Engemann, D., Bzdok, D., Eickhoff, S., Vogeley, K., and Schilbach, L. (2012). Games people play-dash toward an enactive view of cooperation in social neuroscience. *Front. Hum. Neurosci.* 6:148. doi: 10.3389/fnhum.2012.00148

Froese, T., and Di Paolo, E. (2011). The enactive approach: theoretical sketches from cell to society. *Pragmat. Cogn.* 19, 1–36. doi: 10.1075/pc.19.1.01fro

Froese, T., Iizuka, H., and Ikegami, T. (2014a). Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. *Sci. Rep.* 4:3672. doi: 10.1038/srep03672

Froese, T., Iizuka, H., and Ikegami, T. (2014b). Using minimal human-computer interfaces for studying the interactive development of social awareness. *Front. Psychol.* 5:1061. doi: 10.3389/fpsyg.2014.01061

Froese, T., Lenay, C., and Ikegami, T. (2012). Imitation by social interaction? Analysis of a minimal agent-based model of the correspondence problem. *Front. Hum. Neurosci.* 6:202. doi: 10.3389/fnhum.2012.00202

Froese, T., Zapata-Fonseca, L., Leenen, I., and Fossion, R. (2020). The feeling is mutual: clarity of haptics-mediated social perception is not associated with

the recognition of the other, only with recognition of each other. *Front. Hum. Neurosci.* 14:560567. doi: 10.3389/fnhum.2020.560567

Gallagher, S. (2018). "Embodied rationality," in *The Mystery of Rationality: Mind, Beliefs and the Social Sciences*, eds G. Bronner and F. Di Iorio (Cham: Springer International Publishing), 83–94. doi: 10.1007/978-3-319-94028-1_7

Gallagher, S., Mastrogiorgio, A., and Petracca, E. (2019). Economic reasoning and interaction in socially extended market institutions. *Front. Psychol.* 10:1856. doi: 10.3389/fpsyg.2019.01856

Gigerenzer, G., and Selten, R. (eds.). (2002). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA; London: MIT Press. doi: 10.7551/mitpress/1654.001.0001

Hermans, K. S. F. M., Kasanova, Z., Zapata-Fonseca, L., Lafit, G., Fossion, R., Froese, T., et al. (2020). Investigating real-time social interaction in pairs of adolescents with the perceptual crossing experiment. *Behav. Res. Methods* 52, 1929–1938. doi: 10.3758/s13428-020-01378-4

Herschbach, M. (2012). On the role of social interaction in social cognition: a mechanistic alternative to enactivism. *Phenomenol. Cogn. Sci.* 11, 467–486. doi: 10.1007/s11097-011-9209-z

Hofmeyr, A., and Ross, D. (2019). "Team agency and conditional games," in *Contemporary Philosophy and Social Science: An Interdisciplinary Dialogue*, eds M. Nagatsu and A. Ruzzene (London: Bloomsbury), 1–25. doi: 10.5040/9781474248785.ch-003

Infante, G., Lecouteux, G., and Sugden, R. (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *J. Econ. Methodol.* 23, 1–25. doi: 10.1080/1350178X.2015.1070527

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475. doi: 10.1257/000282803322655392

Kelso, J. A. S. (2008). Haken-Kelso-Bunz model. *Scholarpedia* 3:1612. doi: 10.4249/scholarpedia.1612

Kojima, H., Froese, T., Oka, M., Iizuka, H., and Ikegami, T. (2017). A sensorimotor signature of the transition to conscious social perception: co-regulation of active and passive touch. *Front. Psychol.* 8:1778. doi: 10.3389/fpsyg.2017.01778

Lecouteux, G. (2018). What does "we" want? Team reasoning, game theory, and unselfish behaviours. *Rev. D'econ. Polit.* 128, 311–332. doi: 10.3917/redp.283.0311

Lenay, C. (2012). Minimalist approach to perceptual interactions. *Front. Hum. Neurosci.* 6:98. doi: 10.3389/fnhum.2012.00098

Lenay, C., Stewart, J., Rohde, M., and Amar, A. A. (2011). "You never fail to surprise me": the hallmark of the other: experimental study and simulations of perceptual crossing. *Interact. Stud.* 12, 373–396. doi: 10.1075/is.12.3.01len

Michael, J. (2011). Interactionism and mindreading. *Rev. Philos. Psychol.* 2:559. doi: 10.1007/s13164-011-0066-z

Michael, J., and Overgaard, S. (2012). Interaction and social cognition: a comment on Auvray et al.'s perceptual crossing paradigm. *New Ideas Psychol.* 30, 296–299. doi: 10.1016/j.newideapsych.2012.02.001

Murray, L., and Trevarthen, C. (1985). "Emotional regulations of interactions between two-month-olds and their mothers," in *Social Perception in Infants*, eds T. M. Field and N. A. Fox (Norwood, NJ: Ablex Pub), 177–197.

Murray, L., and Trevarthen, C. (1986). The infant's role in mother-infant communications. *J. Child Lang.* 13, 15–29. doi: 10.1017/S0305000900000271

Nadel, J., Carchon, I., Kervella, C., Marcelli, D., and Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Dev. Sci.* 2, 164–173. doi: 10.1111/1467-7687.00065

Nordham, C. A., Tognoli, E., Fuchs, A., and Kelso, J. A. S. (2018). How interpersonal coordination affects individual behavior (and vice versa): experimental analysis and adaptive HKB model of social memory. *Ecol. Psychol.* 30, 224–249. doi: 10.1080/10407413.2018.1438196

Overgaard, S., and Michael, J. (2015). The interactive turn in social cognition research: a critique. *Philos. Psychol.* 28, 160–183. doi: 10.1080/09515089.2013.827109

Petracca, E. (2021). Embodying bounded rationality: from embodied bounded rationality to embodied rationality. *Front. Psychol.* 12:710607. doi: 10.3389/fpsyg.2021.710607

Petracca, E., and Gallagher, S. (2020). Economic cognitive institutions. *J. Instit. Econ.* 1–19. doi: 10.1017/S1744137420000144

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

Reed, K., Peshkin, M., Hartmann, M. J., Grabowecky, M., Patton, J., and Vishton, P. M. (2006). Haptically linked dyads: are two motor-control systems better than one? *Psychol. Sci.* 17, 365–366. doi: 10.1111/j.1467-9280.2006.01712.x

Rizzello, S., and Turvani, M. (2000). Institutions meet mind: the way out of a deadlock. *Constit. Polit. Econ.* 11, 165–180. doi: 10.1023/A:1009085717188

Rolla, G. (2021). Reconceiving rationality: situating rationality into radically enactive cognition. *Synthese* 198, 571–590. doi: 10.1007/s11229-019-02362-y

Ross, D. (2014). Psychological versus economic models of bounded rationality. *J. Econ. Methodol.* 21, 411–427. doi: 10.1080/1350178X.2014.965910

Ross, D. (2021). "Game theory," in *The Stanford Encyclopedia of Philosophy*, ed E. N. Zalta (Metaphysics Research Lab; Stanford University). Available online at: https://plato.stanford.edu/archives/fall2021/entries/game-theory/

Rubinstein, A. (1998). *Modeling Bounded Rationality. Zeuthen Lectures*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/4702.001.0001

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

Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, MA; London: Harvard University Press.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychol. Rev.* 63, 129–138. doi: 10.1037/h0042769

Soussignan, R., Nadel, J., Canet, P., and Gerardin, P. (2006). Sensitivity to social contingency and positive emotion in 2-month-olds. *Infancy* 10, 123–144. doi: 10.1207/s15327078in1002_2

Stirling, W. C. (2012). *Theory of Conditional Games*. New York, NY: Cambridge University Press.

Stirling, W. C. (2019). Conditional coordination games on cyclic social influence networks. *IEEE Trans. Comput. Soc. Syst.* 6, 250–267. doi: 10.1109/TCSS.2019.2892025

Stirling, W. C., and Tummolini, L. (2018). Coordinated reasoning and augmented individualism. *Rev. D'econ. Polit.* 128, 469–492. doi: 10.3917/redp.283.0469

Sugden, R. (1993). Thinking as a team: towards an explanation of nonselfish behavior. *Soc. Philos. Policy* 10, 69–89. doi: 10.1017/S0265052500004027

Sugden, R. (2003). The logic of team reasoning. *Philos. Explor.* 6, 165–181. doi: 10.1080/10002003098538748

Sugden, R. (2018). *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford; New York, NY: Oxford University Press. doi: 10.1093/oso/9780198825142.001.0001

Sugden, R. (2019). The community of advantage. *Econ. Affairs* 39, 417–423. doi: 10.1111/ecaf.12374

Zapata-Fonseca, L., Dotov, D., Fossion, R., Froese, T., Schilbach, L., Vogeley, K., et al. (2019). Multi-scale coordination of distinctive movement patterns during embodied interaction between adults with high-functioning autism and neurotypicals. *Front. Psychol.* 9:2760. doi: 10.3389/fpsyg.2018.02760

Zapata-Fonseca, L., Froese, T., Schilbach, L., Vogeley, K., and Timmermans, B. (2018). Sensitivity to social contingency in adults with high-functioning autism during computer-mediated embodied interaction. *Behav. Sci.* 8:22. doi: 10.3390/bs8020022

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.

Copyright © 2022 Lerique. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

HYPOTHESIS AND THEORY
published: 25 April 2022
doi: 10.3389/fpsyg.2022.761808

What Can Deep Neural Networks Teach Us About Embodied Bounded Rationality

*Edward A. Lee**

Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, United States

"Rationality" in Simon's "bounded rationality" is the principle that humans make decisions on the basis of step-by-step (algorithmic) reasoning using systematic rules of logic to maximize utility. "Bounded rationality" is the observation that the ability of a human brain to handle algorithmic complexity and large quantities of data is limited. Bounded rationality, in other words, treats a decision maker as a machine carrying out computations with limited resources. Under the principle of embodied cognition, a cognitive mind is an *interactive* machine. Turing-Church computations are not interactive, and interactive machines can accomplish things that no Turing-Church computation can accomplish. Hence, if "rationality" is computation, and "bounded rationality" is computation with limited complexity, then "embodied bounded rationality" is both more limited than computation and more powerful. By embracing interaction, embodied bounded rationality can accomplish things that Turing-Church computation alone cannot. Deep neural networks, which have led to a revolution in artificial intelligence, are both interactive and not fundamentally algorithmic. Hence, their ability to mimic some cognitive capabilities far better than prior algorithmic techniques based on symbol manipulation provides empirical evidence for the principle of embodied bounded rationality.

Keywords: bounded rationality, embodied cognition, neural networks, artificial intelligence, computation

OPEN ACCESS

Edited by:
Riccardo Viale,
University of Milano-Bicocca, Italy

Reviewed by:
Gordana Dodig-Crnkovic,
Chalmers University of Technology, Sweden
Ahmed Rebai,
Centre of Biotechnology of Sfax, Tunisia
Adam Csapo,
Széchenyi István University, Hungary

***Correspondence:**
Edward A. Lee
eal@berkeley.edu

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *20 August 2021*
Accepted: *10 February 2022*
Published: *25 April 2022*

Citation:
Lee EA (2022) What Can Deep Neural Networks Teach Us About Embodied Bounded Rationality. Front. Psychol. 13:761808. doi: 10.3389/fpsyg.2022.761808

1. INTRODUCTION

From a computer science perspective, a rational process is step-by-step reasoning using clearly explicable rules of logic. Intractability arises when the number of steps or the amount of data that has to be stored gets too large. Simon's bounded rationality (Simon, 1972) can be interpreted as a recognition of the difficulty that the human mind has in carrying out such rational processes.

Computers, on the other hand, are superbly matched to this sort of rational process. When a decision problem can be formulated as an optimization problem with a clearly defined cost function, an algorithm can often be devised to make an optimal decision. These algorithms are rational in the same sense; they are step-by-step procedures where each step is justified using explicable rules of logic. Such algorithms have repeatedly proven tractable to computers even when hopelessly intractable to humans. When they prove intractable to computers as well, we can often refine them with heuristics and approximations that lead to close to optimal solutions, but even these heuristics are explicable and hence rational.

Among the successes of such computer-driven decision-making are those that lie in the field of optimal control, where a machine makes decisions in response to sensor inputs, and these decisions are used to drive actuators that change the physical world in such a way as to feed back into the

sensors. Self-driving cars, industrial robots, automated trains, and the electric power grid are all examples of such systems. The algorithmic decision-making in these systems interacts with the physical world in such a tight feedback loop that the behavior of the computer cannot be decoupled from the behavior of its physical environment. A whole branch of engineering called "cyber-physical systems" (CPS) has arisen to address the technological problems around such embodied robots (Lee, 2008).

Today, to the surprise and, in some cases, extreme frustration of many researchers, many optimization solutions in engineering are being routinely outperformed by deep neural networks (DNNs). While DNNs are often described as "algorithms," they do not rise to the level of "rational decision making" in the same sense. Although DNNs are typically realized algorithmically on computers, when a DNN produces a result, e.g., classifying an image as a Stop sign, there is no sequence of logical steps that you can point to that rationally leads to the classification. The classification is more like intuition than rationality.

The AlphaGo project (Silver et al., 2016) conclusively demonstrated this principle. The game of Go is notoriously intractable as an optimization problem, and heuristics lead to amateurish play. The best players do not arrive at their moves using a sequence of logical steps, and neither does AlphaGo. More precisely, although AlphaGo is realized as a computer program, that program does not describe any rational decision-making that has anything at all to with the game of Go. Instead, it describes how to build and use a very large data set, refining it by having computers play millions of games of Go with each other. The ability of the program to beat the best Go Masters lies much more in the learning process and the resulting data than in the sequence of rational steps that make and make use of the data. The data has been cultivated in much the same way that a Go Master builds expertise, by practicing.

The centrality of data here is easily and frequently misunderstood. The mantra that "data is the new oil" suggests that data is a resource lying all around us waiting to be exploited. Suppose, for example, that we had stored on disk drives somewhere a record of all the Go games ever played by Go Masters. This would certainly be valuable, but the AlphaGo team did not have such a data set, and their result would likely have been less spectacular had they trained their DNNs on that data set. If they had such a data set, they could have trained a DNN easily because each move in each board position is clearly labeled as a "winning" or "losing" move by the final outcome of the game. But that is not what they did. Instead, they programed their machines to play against each other. The first few million games were amateurish, but through the magic of backpropagation, each game refined the data driving the decisions such that each game got better. The data was not mined, it was created.

The process of training the AlphaGo machines is interactive, not observational. It is first person, not third person. By analogy, a human will never acquire the ability to outperform Go Masters by just watching masters play Go. The human has to interact with Go Masters to become a Go Master. Interaction is more powerful than observation. Not only do humans learn better by doing, so do machines.

The principle of embodied cognition puts interaction front-and-center. The mind is not a process in a brain observing the world through sensors. Instead, the mind is an interaction between processes in a brain and the world around it (Thelen, 2000). The kinds of problems that DNNs excel at are precisely those where interaction is front-and-center. And the decisions made by DNNs are frustratingly inexplicable, resisting any label as rational decisions.

In this article, I will show from several perspectives that interaction is more powerful than observation. There are things that can be accomplished through interaction that are impossible through observation. I will give technical and mathematical examples that are not possible without interaction.

I will also show that interaction can occur without algorithms. Although DNNs can be realized by computers, these realizations are brute-force simulations of processes that are not fundamentally algorithmic. The field of reservoir computing (Tanaka et al., 2019), for example, offers very different architectures that have little resemblance to Turing-Church computations and would be hard to describe as rational decision makers in the sense considered here. The field of feedback control, which is fundamentally about interaction, does not fundamentally need computers nor algorithms. Indeed, its earliest applications in the 1920s through the 1950s predate digital computers.

Proponents of embodied cognition often use the term "computation" much more broadly than I am using it here to mean any sort of information processing (Dodig-Crnkovic and Giovagnoli, 2013; Müller and Hoffmann, 2017; Dodig-Crnkovic, 2018). Any dynamic process that reacts to sensed information about its environment is capable of such "morphological computing" (Pfeifer and Bongard, 2007) or "natural computing" (Müller and Hoffmann, 2017). Such computation is performed by every living organism (Maturana and Varela, 1980; Stewart, 1995) and many non-living organisms (a thermostat, for example), and hence is much too broad to bear much if any relationship to bounded rationality in the sense of Simon (1972). In this article, "computation" will be limited to the meaning given by Turing and Church, as done for example by Piccinini (2007), and I will show in Section 4 that this meaning is not the same as information processing. I will argue that the Turing-Church meaning of "computation" does not even include many of the processes we accomplish today using digital computers. But it is this sense that matches the bounded rationality of Simon.

In the prevailing philosophy of science, observation trumps interaction. We are taught that the best science is objective, not subjective. Let the data speak for itself. Design your instruments to minimally disrupt what you are observing. But science also teaches us that observation without interaction is impossible. My claim is that it is also undesirable. We can accomplish much more if we embrace feedback and interaction.

The main contributions of this article are to point out that Turing-Church computations are objective, observational, and non-interactive processes; to clarify that an algorithm is the specifications of what a Turing-Church computation does; to show in several ways that first-person interaction, i.e., a

feedback system, can accomplish things that no Turing-Church computation can; to argue that deep neural networks are feedback systems and are not fundamentally algorithmic; and to argue that the efficacy of DNNs on certain cognitive tasks provides empirical support for the thesis of embodied bounded rationality.

2. BOUNDED RATIONALITY

In the 1970s, Herbert Simon challenged the prevailing dogma in economics, which assumed that agents act rationally. His key insight, for which he got the Nobel Prize in economics, was that those agents (individuals and organizations) do not have the capability to make the kinds of rational decisions that economists assumed they would. In his words:

> Theories that incorporate constraints on the information-processing capacities of the actor may be called theories of bounded rationality (Simon, 1972).

He identified three limitations: uncertainty about the consequences that would follow from alternative decisions, incomplete information about the set of alternatives, and complexity preventing the necessary computations from being carried out. He argued that "these three categories tend to merge," using the game of chess as an example and saying that the first and second, like the third, are fundamentally an inability to carry out computation with more than very limited complexity:

> What we refer to as "uncertainty" in chess or theorem proving, therefore, is uncertainty introduced into a perfectly certain environment by inability—computational inability—to ascertain the structure of that environment (Simon, 1972).

Three decades later, he reaffirmed this focus on the process of reasoning:

> When rationality is associated with reasoning processes, and not just with its products, limits on the abilities of Homo sapiens to reason cannot be ignored (Simon, 2000).

Reasoning and rationality as computation are central to his theory, and he argued that economists' assumptions that agents would maximize expected utility was unrealistic in part because that maximization is intractable to a human mind.

3. ALGORITHMS AND COMPUTATION

What is an algorithm? Merriam-Webster gives this definition: "a step-by-step procedure for solving a problem or accomplishing some end." Despite the simplicity of this definition, the term is widely used more broadly. Domingos (2015), for example, in his book *The Master Algorithm*, states that Newton's second law is an algorithm. Often expressed as $F = ma$, force equals mass times acceleration, Newton's second law is *not* an algorithm. There are no steps, there is no procedure, and there is no end. Instead, Newton's second law is a relation between two continuously varying quantities, force and acceleration, where the latter quantity expresses a rate of change of velocity, which in turn expresses a rate of change of position. Domingos seems to use the word "algorithm" to mean anything that is formally expressible. In this article, I will use the term "algorithm" in a narrower manner consistent with the Merriam-Webster definition.

Newton's second law is a differential equation. Acceleration is the second derivative of position. Not only is a differential equation not an algorithm, but many differential equations express behaviors for which *there is no algorithm*. Every algorithm that attempts to *simulate* a process described by such a differential equation is flawed. Newton's second law is a linear differential equation for which, for many input force functions, we can find a closed-form solution. Once we have such a solution, we can devise an algorithm that gives the position at any chosen point in time. However, for non-trivial force inputs, and for most non-linear differential equations, there is no such closed-form solution, and every algorithmic approximation exhibits arbitrarily large errors. Non-linear differential equations, in particular, often exhibit chaotic behavior, where arbitrarily small errors at any step become arbitrarily large errors in future steps. The discovery of such chaotic behavior is attributed to Lorenz (1963), who was frustrated by the inability of computer models to predict weather more than a few days in advance. The differential equations modeling the thermodynamics of weather are chaotic, and every algorithmic approximation develops arbitrarily large errors over time.

While time is central to differential equations, it is irrelevant to algorithms. The steps of an algorithm are discrete, entirely separable from one another, and the time it takes to complete a step is irrelevant to whether the algorithm is being correctly carried out. In contrast, in an interactive system or a feedback system where part of the interaction is a physical process, time plays a major role. Hence, under the principle of embodied cognition, time is central to cognition, a point forcefully made by Esther Thelen:

> It is precisely the continuity in time of the embedded and coupled dynamic systems essential for fluid, adaptive behavior that gives meaning to the notion of an embodied cognition (Thelen, 2000).

What is the relationship between algorithms and computation? Here again, I will stick to a rigorous use of this term, adopting the meaning established by Turing (1936) and Church (1932). In this meaning, a computation is a step-by-step procedure (i.e., a carrying-out of an algorithm) operating on digital information that terminates and gives an answer. What is now called the Church-Turing thesis states that every such computation can be computed by a Turing machine, a machine that realizes the algorithm. Turing showed that there is a particular Turing machine, or, equivalently, a particular algorithm, that can realize any other Turing machine. This machine is called a "universal Turing machine." Given enough time and memory, any modern computer can realize a universal Turing machine.

Unfortunately, many people misrepresent the universal Turing machine, calling it simply a "universal machine," and

stating that it can realize any other machine. For example, in his book *Tools for Thought*, Howard Rheingold states,

> The digital computer is based on a theoretical discovery known as "the universal machine," which is not actually a tangible device but a mathematical description of a machine capable of simulating the actions of any other machine (Rheingold, 2000, p. 15).

Rheingold misleads by speaking too broadly about machines. There is no universal machine, mathematical or otherwise. A universal Turing machine can only perform computations.[1]

With regard to computation, humans are much more limited than computers. Computers have no difficulty taking billions of steps in an algorithm to solve a problem, whereas humans struggle with a few dozen. Algorithmic reasoning may seem like the epitome of thought, but if it is, then humans fall far short of that epitome. So far short, in fact, that Simon may have not gotten it quite right. If human decisions are the result of a limited amount of computation, then it is an extremely limited amount. What if they are not the result of computation at all?

Kahneman, in his book, *Thinking Fast and Slow*, identifies two distinct human styles of thinking, a fast style (System 1) and a slow style (System 2). The slow style is capable of algorithmic reasoning, but the fast style, which is more intuitive, is responsible for many of the decisions humans make. It turns out that many of today's artificial intelligences (AIs) more closely resemble System 1 than System 2. Even though they are realized on computers, they do not reach decisions by algorithmic reasoning.

4. INFORMATION PROCESSING IS NOT (NECESSARILY) COMPUTATION

"Computation," in the sense that I am using the term in this article, is not the same as information processing, in the sense used in Dodig-Crnkovic and Giovagnoli (2013), Müller and Hoffmann (2017), and Dodig-Crnkovic (2018). In this article, computation is (a) algorithmic (consisting of a sequence of discrete steps, where each step is drawn from a finite set of possible operations); (b) terminating; (c) operating on discrete data (the inputs, outputs, and intermediate states are all drawn from countable sets); and (d) non-interactive (inputs are available at the start and outputs at termination). Turing-Church computation has all four of these properties. Under this definition, the set of all possible computations is countable. The core results in the theory of computation (e.g., undecidability, complexity measures, and the universality of Turing machines) all depend on this countability.

In Lee (2017) (Chapter 7), I define "information" as "resolution of alternatives." Using Shannon information theory, I point out that information need not be discrete. The alternatives may lie in a finite, countable, or uncountable set. I show that measurements of information (entropy) are incomparable when the alternatives lie in a finite or countable set vs. when they lie in an uncountable set. There is an infinite offset between these two measures of information. In particular, if the set of alternatives is countable, then entropy gives the expected number of bits needed to encode a selected alternative. This number of bits is a measure of the amount of information gained by observing a selected alternative. However, if the set of alternatives is uncountable, then entropy can still be finite, but it no longer represents a number of bits needed to encode a selected alternative. In fact, an infinite number of bits is required. Nevertheless, this entropy can still be interpreted as a measure of the amount of information in an observation of an alternative, and these amounts can be compared with each other, but these amounts are always infinitely larger than the amount of information in an observation drawn from a countable set of alternatives.

Many mistakes are made in the literature by ignoring this infinite offset. For example, Lloyd (2006) says about the second law of thermodynamics, "It states that each physical system contains a certain number of bits of information—both invisible information (or entropy) and visible information—and that the physical dynamics that process and transform that information never decrease that total number of bits." But the second law works absolutely unmodified if the underlying random processes are continuous, in which case the set of alternatives is uncountable, and the information is not representable in bits. The same mistake is made by Goyal (2012), who states "The fact that [the entropy of a black hole] is actually finite suggests that the degrees of freedom are not non-denumerably infinite." But the entropy of a black hole given by Bekenstein (1973) is based on a continuous probability density, so its finiteness does not imply countable degrees of freedom. Goyal (2012) continues, stating for example that in quantum physics, "the number of possible outcomes of a measurement *may* be finite or countably infinite" (emphasis added), and then implying that it is *always* finite or countably infinite. Goyal (2012) goes on to assert that "this stands in contrast with the classical assumption that all physical quantities (such as the position of a particle) can take a continuum of possible values." There are some physical measurements that have only a finite or countable number of outcomes, such as the spin of an electron, but position of a particle is not one of them. The Schrödinger equation operates in a time and space continuum and the wave function describing position is reasonably interpreted as a probability *density* function governing an uncountable number of possible alternatives. The discreteness of time and space is a later overlay on quantum theory that remains controversial and is not experimentally supported. Goyal takes a leap of faith, concluding "hence, discreteness challenges the classical idea that the continua of space and time are the fundamental bedrock of physical reality." In contrast, Dodig Crnkovic (2012) observes that "information is both discrete and continuous."

Information that lies in a continuum of alternatives can be operated on by processes that are neither algorithmic nor terminating. Ordinary differential equation models of the physical world can be interpreted as performing such operations. Such information processing is not, however, computation. Chaos theory shows that such information processing cannot

[1]Copeland (2017) has a nice section on common misunderstandings of the Turing-Church thesis.

even be approximated with bounded error by computation (Lee, 2017, Chapter 10).

Given these facts, we are forced to make one of two choices: either (A) information processing is richer than computation, or (B) the physical world does not have uncountable alternatives, Hypothesis (B) is sometimes called "digital physics" (Lee, 2017, Chapter 8). Some physicists and computer scientists go further and claim that everything in the material world *is actually* a Turing-Church computation.[2] I have previously shown, however, that hypothesis (B) is not testable by experiment unless it is *a priori* true (Lee, 2017, Chapter 8). Specifically, the Shannon channel capacity theorem tells us that every noisy measurement conveys only a finite number of bits of information, and therefore can only distinguish elements from a countable set of alternatives. Hence, hypothesis (B) is scientific, in the sense of Popper (1959), only if it is *a priori* true. Hence, hypothesis (B) is a matter of faith, not science.

If anything in the physical world forms a continuum (time or space, for example), then noise in measurements remains possible, no matter how good the measurement apparatus becomes. This follows from the incompleteness of determinism (Lee, 2016). A noiseless measurement of some physical system would have to be deterministic, in the sense that the same physical state should always yield the same measurement result. However, I have shown in Lee (2016) that any set of deterministic models of the physical world that includes both discrete and continuous alternatives and that is rich enough to include Newton's laws is incomplete. It does not contain its own limit points. Non-determinism, therefore, is inescapable unless digital physics is *a priori* true and there are no continuous alternatives. This means that at least some measurements will always be vulnerable to noise unless the hypothesis to be tested experimentally is already true. Hence, hypothesis (B) can only be defended by a circular argument.

Hypothesis (B) is not only a matter of faith, but it also a poor choice under the principle of Occam's razor. As I point out in Lee (2020) (Chapter 8), models based only on countable sets may be far more complex than models based on continuums. Diophantine equations, for example, which are widely used in physics, for example to describe the motions of bodies in gravitational fields, are chaotic and exhibit weird gaps when defined over countable sets. A more defensible position, therefore, is hypothesis (A), which allows for information processing as a reasonable model of the physical world without insisting that information processing have the form of computation. This position is also supported by Piccinini (2020) who states, "information processing may or may not be done by computing" (Chapter 6).[3]

[2]For a particularly bad exposition of this hypothesis, full of pseudo science and misinformation, see Virk (2019).

[3]Piccinini (2020) nevertheless defends a "computational theory of cognition" (CTC), though his use of computation is again broader than mine here, and even then, he admits that this theory may not provide a complete explanation.

5. DEEP NEURAL NETWORKS

Deep neural nets (DNNs), which have transformed technology by enabling image classification, speech recognition, and machine translation, to name a few examples, are inspired by the tangle of billions of neurons in the brain and rely on the aggregate effect of large numbers of simple operations. They are, today, mostly realized by computers, and hence are composed of "algorithms" and "computation." However, to view these realizations as Turing-Church computations is to ignore the role of feedback, a property absent in the Turing-Church model. This role is not incidental. Moreover, it also ignores the possibility that today's realizations of neural networks are brute force computational approximations of information processing that is not, at its root, computational.

A frustrating result of the recent successes in deep neural nets is that people have been unable to provide explanations for many of the decisions that these systems make (Lee, 2020, Chapter 6). In May 2018 a new European Union regulation called the General Data Protection Regulation (GDPR) went into effect with a controversial provision that provides a right "to obtain an explanation of the decision reached" when a decision is solely based on automated processing. Legal scholars, however, argue that this regulation is neither valid nor enforceable (Wachter et al., 2017). In fact, it may not even be desirable. I conjecture that sometime in the near future, someone will figure out how to train a DNN to provide a convincing explanation for *any* decision. This could start with a generative-adversarial network (GAN) that learns to provide explanations that appear to be generated by humans.

Humans are very good at providing explanations for our decisions. But our explanations are often wrong or at least incomplete. They are likely to be *post hoc* rationalizations, offering as explanations factors that do not or cannot account for the decisions we make. This fact about humans is well-explained by Kahneman, whose work on "prospect theory," like Simon's bounded rationality, challenged utility theory. In prospect theory, decisions are driven more by gains and losses rather than the value of the outcome. Humans, in other words, will make irrational decisions that deliver less value to them in the end. In *Thinking Fast and Slow*, Kahneman offers a wealth of evidence that our decisions are biased by all sorts of factors that have nothing to do with rationality and do not appear in any explanation of the decision.

Kahneman reports, for example, a study of the decisions of parole judges in Israel by Danziger et al. (2011). The study found that these judges, on average, granted about 65 percent of parole requests when they were reviewing the case right after a food break, and that their grant rate dropped steadily to near zero during the time until the next break. The grant rate would then abruptly rise to 65 percent again after the break. In Kahneman's words,

> The authors carefully checked many alternative explanations. The best possible account of the data provides bad news: tired and hungry judges tend to fall back on the easier default position of

> denying requests for parole. Both fatigue and hunger probably play a role (Kahneman, 2011).

And yet, I'm sure that every one of these judges would have no difficulty coming up with a plausible explanation for their decision for each case. That explanation would not include any reference to the time since the last break.

Taleb, in his book *The Black Swan*, cites the propensity that humans have, after some event has occurred, to "concoct explanations for its occurrence after the fact, making it explainable and predictable" (Taleb, 2010). For example, the news media always seems to have some explanation for movements in the stock market, sometimes using the same explanation for both a rise and a fall in prices.

Taleb reports on psychology experiments where subjects are asked to choose among twelve pairs of nylon stockings the one they like best. After they had made their choice, the researchers asked them for reasons for their choices. Typical reasons included color, texture, and feel, but in fact, all twelve pairs were identical.

Taleb also reports on some rather dramatic experiments performed with split-brain patients, those who have undergone surgery where the corpus callosum connecting the two hemispheres of the brain has been severed. Such surgery has been performed on a number of victims of severe epilepsy that have not responded to less aggressive treatments. These experiments support the hypothesis that the propensity for *post hoc* explanations has deep biological roots. An image presented to the left half of the visual field will go to the right side of the brain, and an image presented to the right half of the visual field will go to the left side of the brain. In most people, language is centered in the left half of the brain, so the patient will only be able to verbalize the right field experience. For example, a patient with a split brain is shown a picture of a chicken foot on the right side and a snowy field on the left side and asked to choose the best association with the pictures. The patient would correctly choose a chicken to associate with the chicken foot and a shovel to associate with the snow. When asked why the patient chose the shovel, the patient would reply that was "for cleaning out the chicken coop." Taleb concludes,

> Our minds are wonderful explanation machines, capable of making sense out of almost anything, capable of mounting explanations for all manner of phenomena, and generally incapable of accepting the idea of unpredictability (Taleb, 2010).

Demanding explanations from AIs could yield convincing explanations for anything, leading us to trust their decisions too much. Explanations for the inexplicable, no matter how plausible, are simply misleading.

Given that humans have written the computer programs that realize the AIs, and humans have designed the computers that execute these programs, why is it that the behavior of the programs proves inexplicable? The reason is that what the programs do is not well-described as algorithmic reasoning, in the same sense that an outbreak of war is not well-described by the interactions of protons and electrons. Explaining the implementation does not explain the decision.

Before the explosive renaissance of AI during the past two decades, AI was dominated by attempts to encode algorithmic reasoning directly through symbolic processing. What is now called "good old-fashioned AI" (GOFAI) encodes knowledge as production rules, if-then-else statements representing the logical steps in algorithmic reasoning (Haugeland, 1985). GOFAI led to the creation of so-called "expert systems," which were sharply criticized by Dreyfus and Dreyfus (1986) in their book, *Mind Over Machine*. They pointed out, quite simply, that following explicit rules is what novices do, not what experts do. Dreyfus and Dreyfus called the AI practitioners of the time,

> false prophets blinded by Socratic assumptions and personal ambition—while Euthyphro, the expert on piety, who kept giving Socrates examples instead of rules, turns out to have been a true prophet after all (Dreyfus and Dreyfus, 1984).

Here, Dreyfus and Dreyfus are reacting (rather strongly) to what really was excessive hyperbole about AI at the time. They were just the tip of a broad backlash against AI that came to be called the "AI winter," where funding for research and commercial AI vanished nearly overnight and did not recover until around 2010.

DNNs work primarily from examples, "training data," rather than rules. The explosion of data that became available as everything went online catalyzed the resurgence of statistical and optimization techniques that had been originally developed in the 1960s through 1980s but lay dormant through the AI winter before exploding onto the scene around 2010.

DNNs particularly excel at functions that, in humans, we call perception, for example the ability to classify objects in an image. Stewart (1995) attributes to the Chilean biologist and philosopher Maturana the perspective that, "perception should not be viewed as a grasping of an external reality, but rather as the specification of one." Indeed, the supervised training process that for a DNN such as Inception, which is distributed by Google as part of their open-source TensorFlow machine learning toolkit, results in a network that *specifies* a taxonomy rather than recognizing an objectively existing one. Because training images are labeled by humans, the resulting taxonomy is familiar to humans.

The techniques behind the AI renaissance are nothing like the production rules of GOFAI. A central one of these techniques, now called backpropagation, first showed up in automatic control problems quite some time ago. Kelley (1960) describes a controller that would carry a spacecraft from Earth's orbit to Mars's orbit around the sun using a solar sail. His controller, a feedback system, bears a striking resemblance to backpropagation, although his formulation is more continuous than the discrete form used in machine learning today. His formulation did not require a digital computer to realize it, and in fact, any computer realization would have been an approximation of his specification.

Based in part on Kelley's work, Bryson et al. (1961) describe a feedback system to control a spacecraft that is re-entering the earth's atmosphere to minimize heating due to friction. They adapted Kelley's method into a multistage technique that closely resembles the backpropagation technique used for DNNs today. The Kelley-Bryson technique was restated in a form closer to its usage today by Dreyfus (1962).

Backpropagation can be thought of as a technique for a system to continuously redesign itself by probing its environment (including its own embodiment) and adapting itself based on the reaction. DNNs realized in software are better thought of as programs that continuously rewrite themselves during their training phase. Today, it is common to freeze the program after the training phase, or to update it only rarely, but this practice is not likely to persist for many applications. Continuing to learn proves quite valuable.

There have been attempts to use machine learning techniques to learn *algorithmic* reasoning, where the result of the training phase is a set of explicable production rules, but these have proven to underperform neural networks. Wilson et al. (2018) created a program that could write programs to play old Atari video games credibly well. Their program generated random mutations of production rules, and then simulated natural selection. Their technique was based on earlier work that evolved programs to develop certain image processing functions (Miller and Thomson, 2000). The Atari game-playing programs that emerge, however, are far less effective than programs based on DNNs. Wilson et al. (2018) admit this, saying that the main advantage of their technique is that the resulting programs are more explainable. The learned production rules provide the explanations.

In contrast, once a DNN has been trained, even a deep understanding of the computer programs that make its decisions does not help in providing an explanation for those decisions. Exactly the same program, with slightly different training, would yield different decisions. So the explanation for the decisions must be in the data that results from the training. But those data take of the form of millions of numbers that have been iteratively refined by backpropagation, a feedback system. The numbers bear no resemblance to the training data and have no simple mapping onto symbols representing inputs and possible decisions. Even a deep understanding of backpropagation does little to explain how the particular set of numbers came about and why they lead to the decisions that they do. Fundamentally, the decisions are not a consequence of algorithmic reasoning.

Today, implementations of DNNs are rather brute force, using enormous amounts of energy and requiring large data centers with a great deal of hardware. The energy consumption of a human brain, in contrast, is quite modest. In an attempt to come closer, there is a great deal of innovation on hardware for machine learning. Some of this hardware bears little resemblance to modern computers and has no discernible roots in Turing-Church computation, using for example analog circuits. Reservoir computing (Tanaka et al., 2019) is a rather extreme example, where a fixed, non-linear system called a reservoir is used as a key part of a neural network. The reservoir can be a fixed physical system, such as a random bundle of carbon nanotubes and polymers. These innovations demonstrate that DNNs are not, fundamentally, Turing-Church computations, and they may eventually be realized by machines that do not resemble today's computers.

Simon developed his theory of bounded rationality well before DNNs, at a time when AI was all about symbolic processing. Newell and Simon (1976) say, "symbols lie at the root of intelligent action, which is, of course, the primary topic of artificial intelligence." They add, "a physical symbol system has the necessary and sufficient means for general intelligent action." They go further and commit to the universal machine hypothesis:

> A physical symbol system is an instance of a universal machine. Thus the symbol system hypothesis implies that intelligence will be realized by a universal computer (Newell and Simon, 1976).

We now know that this hypothesis is false. DNNs outperform symbolic processing on many problems, particularly on more cognitively difficult problems. Although their realizations in computers arguably use symbols (0 to represent "false" and 1 to represent "true," for example), those symbols have no relationship to the problem they are solving.

6. INTERACTION AND FEEDBACK

In the thesis of embodied cognition, the mind "simply does not exist as something decoupled from the body and the environment in which it resides" (Thelen, 2000). The mind is not a computation that accepts inputs from the environment and produces output, but rather the mind *is* an interaction of a brain with its body and environment. A cognitive being is not an observer, but rather a collection of feedback loops that include the body and its environment. Fundamentally, under this thesis, a cognitive mind is an interactive system.

If "rationality" is computation, and "bounded rationality" is computation with limited resources, then "embodied bounded rationality" is both more limited than computation and more powerful. By embracing interaction, embodied bounded rationality can accomplish things that bounded rationality or even unbounded rationality alone cannot.

Turing-Church computation is not interactive. There is no part of the theory that includes effects that outputs from the computation may have on inputs to the computation. Central to what a computation is, in this theory, is that the inputs are fully available at the start, and that the outputs are available when the computation terminates. If the computation does not terminate, there is no output and the process is not a computation. There is nothing in the formalism that enables the machine to produce intermediate outputs, allow the environment to react and provide new inputs, and then continue by reacting to those new inputs. The "universal" Turing machine proves to be far from universal because it does not include any such interactive machines.

To understand this point, it is critical to realize that the behavior that emerges from an interactive machine is not just a consequence of what the machine does, but also of what the machine's environment does. Hence, the only way to make Turing-Church computations truly "universal," including interactive machines, is to ensure that their environment is part of the Turing-Church computation. To do this in general, you have to assume digital physics, something you can only do on faith.

The biggest breakthroughs in AI replace the prior open-loop good old-fashioned AI (GOFAI) techniques with interaction

and feedback. Here, I use the term "feedback" for interaction where one part of the system provides a stimulus to another part, measures its response, and adjusts its actions to make future responses more closely resemble its goals. Deep neural networks are, fundamentally, feedback systems in this sense, and they yield results of such complexity as to be inexplicable (Lee, 2020, Chapter 6). The algorithms by which they are realized on computers are simply not good descriptions of what they do.

In this section, I will go through a series of illustrations of what can be accomplished with interaction and feedback that is not possible with Turing-Church computation alone. Some of these are quite technical and serve as proofs of the limitations of computation, while others are just better explanations of what is really going on.

6.1. Driving a Car

Wegner (1998) gives a simple example that illustrates the limitations of non-interactive machines, driving a car. Consider a cruise control system, which maintains the speed of a car close to a specified setpoint. In an interactive solution, the inputs to this system are measurements of the speed of the car, and the system simply accelerates (opens the throttle) if the speed is too low and decelerates if the speed is too high. The system continuously watches the effects of its actions and continuously corrects by adjusting the throttle. The system automatically compensates for changes in the environment, such as climbing a hill. This is a tight feedback loop, an embodied solution where the "smarts" of the cruise control is in its *interaction* with its body (the car) and its environment (the roadway). This solution is extremely simple, a feedback control system realizable with technology that was patented back in the 1930s (Black, 1934), well before digital computers.

Now, consider solving this problem as a Turing-Church computation without interaction. First, in order to terminate, the problem will only be able to be solved for a finite time horizon, and, to be algorithmic, time will need to be discretized. Assume a car is driven for no more than 2 h on each trip, and that we will get sufficient accuracy if we calculate the throttle level that needs to be applied each 100 ms. The output, therefore, will be a trace of 72,000 throttle levels to apply. What is the input? First, we need as input the elevation gains and losses along the trajectory to be taken by the car during all segments where the cruise control is to be active. We will also need a detailed model of the dynamics of the car, including its weight, the weight of each of the passengers and the contents in the trunk, and how the car responds to opening and closing the throttle. The computation will now need to solve complex differential equations governing the dynamics to calculate what throttle to apply to the car as it moves over the specified trajectory. The simple problem has become a nightmare of complexity requiring a great deal of prior knowledge and most likely yielding a lower quality result.

The reader may protest that what the cruise control actually does is rather simple computation. It takes as input a measurement of the current speed, subtracts it from the desired speed, multiplies by a constant, and adds the result to the current throttle position.[4] But is this a good description of what the machine does? By analogy, does a human mind take as input a grunt or squeal and produce as output a grunt or squeal? Or does it engage in conversation? Which is a better description? The latter is a description of what the brain, body, and environment accomplish *together*, whereas the former is a description only of what the brain does. The cruise control *system* includes the car, and what it accomplishes is not arithmetic but rather keeping a constant speed.

The cruise control system considered previously could be made more "intelligent" by endowing it with additional feedback. It could check, for example, that when it issues a command to further open the throttle that the car does indeed accelerate. This is a check for fault conditions that might prevent the cruise control system from operating properly. In a way, this check makes the system more "self aware," aware of its own body and the expected effects that its actions have on that body. Indeed, there is a fledgling subfield of engineering concerned with "self-aware systems," with a number of workshops worldwide addressing the question of how to design systems that gather and maintain information about their own current state and environment, reason about their behavior, and adapt themselves as necessary. Active interaction with the environment is an essential tool for such systems. The cruise control system has to open the throttle to perform the test to determine whether it is working correctly. Such interaction is a first-person activity, not a third-person observation, and it is a central principle behind embodied robots, which I consider next.

6.2. Embodied Robots

Clark and Chalmers (1998) used the term "cognitive extension" for the idea that the mind is not something trapped in the head but rather is spread out into the body and the world around it. Clark's work centers on the processes where the brain tries to predict what the senses will sense and then uses the differences between the predictions and what is sensed to improve the predictions. These feedback loops extend out into the world, encompassing the body and its physical environment so that they become an intrinsic part of thinking. In his words, "certain forms of human cognizing include inextricable tangles of feedback, feed-forward, and feed-around loops: loops that promiscuously crisscross the boundaries of brain, body, and world" (Clark, 2008). If Clark is right, then cognition in machines will not much resemble that in humans until they acquire ways to interact with the world like humans. Some computer programs are already starting to do this, particularly those that control robots.

Robots are, in a sense, embodied computers, but for the most part, they have not been designed in an embodied way. Clark (2008) compares Honda's Asimo robot to humans, observing that Asimo requires about sixteen times as much energy as humans to walk, despite being shorter and lighter. He attributes this to the style of control:

[4]What I have just described is the simplest form of a negative feedback controller, which is known as a proportional controller. A modern cruise control would more likely realize a PID controller (proportional, integral, derivative), but the computation would only be a slightly more complicated.

> Whereas robots like Asimo walk by means of very precise, and energy-intensive, joint-angle control systems, biological walking agents make maximal use of the mass properties and biomechanical couplings present in the overall musculoskeletal system and walking apparatus itself (Clark, 2008).

Clark points to experiments with so-called passive-dynamic walking (McGeer, 2001). Passive-dynamic robots are able to walk in certain circumstances with *no* energy source except gravity by exploiting the gravitational pull on their own limbs. You can think of these robots as performing controlled falling. McGeer's robots did not include any electronic control systems at all, but subsequent experiments have shown that robots that model their own dynamics in gravity can be much more efficient.

Conventional robotic controllers use a mechanism called a servo, a feedback system that drives a motor to a specified angle, position, or speed. For example, to control a robot arm or leg, first a path-planning algorithm determines the required angles for each joint, and then servos command the motors in each joint to move to the specified angle. The servos typically make little use of any prior knowledge of the physical properties of the arm or leg, their weight and moment of inertia, for example. Instead, they rely on the power of negative feedback to increase the drive current sufficiently to overcome gravity and inertia. It's no wonder these mechanisms are not energy efficient. They are burning energy to compensate for a lack of self-awareness.

Brooks (1992) articulates a vision of "embodied robots" that learn how to manipulate their own limbs rather than having hard-coded, preprogrammed control strategies. Gallese et al. (2020) observe,

> Newell and Simon's physical symbol system hypothesis was questioned when the "embodied robots" designed by Rodney Brooks proved able to simulate simple forms of intelligent behavior by externalizing most of cognition onto the physical properties of environments, thus dispensing with abstract symbolic processing" (Brooks, 1991, p. 377).

Brooks' vision was perhaps first demonstrated in real robots by Bongard et al. (2006). Their robot learns to pull itself forward using a gait that it develops by itself. The robot is not even programmed initially to know how many limbs it has nor what their sizes are. It makes random motions initially that are ineffective, much like an infant, but using feedback from its sensors it eventually puts together a model of itself and calibrates that model to the actual limbs that are present. This resembles the learning process in DNNs, where initial decisions are random and feedback is used to improve them. When a leg is damaged, the gait that had worked before will no longer be effective, but since it is continuously learning, it will adapt and develop a new gait suitable for its new configuration. If one of its legs "grows" (someone attaches an extension to it, for example), the robot will again adapt to the new configuration.

Pfeifer and Bongard (2007) assert that the very kinds of thoughts that we humans are capable of are both constrained and enabled by the material properties of our bodies. They argue that the kinds of thoughts we are capable of have their foundation in our embodiment, in our morphology and the interaction between the brain, the body, and its environment. Pfeifer and Bongard argue that fundamental changes in the field of artificial intelligence over the past two decades yield insights into cognition through "understanding by building." If we understand how to design and build intelligent embodied systems, they reason, we will better understand intelligence in general. Indeed, DNNs are teaching us that intelligence is not necessarily rational.

The classical servo-based robot control systems are simple feedback control loops like those developed by Black (1934). With a servo, the controller plans a path, and the mechanism forces the motion to match that path. Only recently have servos been realized using computers. In embodied robots, a second feedback loop is overlaid on this first one. In this second loop, the robot learns its own morphology and dynamics.

Higher-order cognitive feedback loops also enable humans to recognize flaws in our plans and attempt to improve them. Supporting Clark's argument, in *I Am a Strange Loop*, Hofstadter states:

> You make decisions, take actions, affect the world, receive feedback, incorporate it into your self, then the updated "you" makes more decisions, and so forth, round and round (Hofstadter, 2007).

Hofstadter emphasizes that feedback loops create many if not all of our essential cognitive functions. These feedback loops are entirely absent in Turing-Church computation.

6.3. Solving Undecidable Problems

Although it is complex, given good enough models, the cruise control and robotics problems are solvable, at least approximately, by a Turing-Church computation. Hence, these arguments do not, by themselves, speak to any *fundamental* limitations of Turing-Church computations.

Interactive systems, however, can sometimes solve problems that are provably unsolvable by Turing-Church computations. Back in the 1990s, a Ph.D. student of mine, Thomas Parks, surprised me by showing how to solve an undecidable problem (Parks, 1995). The problem was to determine whether a particular network of communicating processes built in a particular style due to Kahn and MacQueen (1977) can be executed for an unbounded amount of time using only a bounded amount of memory. Parks proved that the problem is undecidable, meaning that there is no Turing-Church computation that can yield an answer for all possible such networks. He then proceeded to solve the problem with an *interactive* solution. He provided a policy that provably uses bounded memory for any such network that can be executed in bounded memory.[5] Parks' solution is interactive, in that his scheduling policy makes decisions, watches how the program responds, and makes additional decisions accordingly. Such a strategy does not fit the Turing-Church model.

Strictly speaking, Parks doesn't really solve an undecidable problem, but rather solves a different but related problem. The question he answers is not *whether* a Kahn-MacQueen program

[5]This work was later generalized by Geilen and Basten (2003).

can execute in bounded memory, but rather *how* to execute a Kahn-MacQueen program in bounded memory. Computation plays a rather small role in the solution compared to interaction. By analogy, an automotive cruise control is not performing arithmetic; it is keeping the speed of the car constant.

Just as with the cruise control, computation forms a *part* of the interactive machine. Parks' solution performs a Turing-Church computation for each decision. Computers are used this way all the time. A user types something, the machine performs a computation and presents a resulting stimulus to the user, then the user types something more, and the machine performs *another* computation. Each computation is well-modeled by the Turing-Church formalism, but the complete closed-loop system is not.

6.4. Reasoning About Causation

Rational decision making frequently involves reasoning about causation. I do not smoke because smoking *causes* cancer. I click on Amazon's website because it *causes* goods to appear at my door. Pearl and Mackenzie (2018) show that it is impossible to draw conclusions about causation in a system by objectively observing the system. One must either *interact* with the system or rely on prior subjective assumptions about causation in the system.

A Turing-Church computation is an objective observer. It does not affect its inputs, as it would if it were an interactive system. To reason about causation, therefore, it can only encode the prior subjective assumptions of its designer. It cannot test those assumptions through interaction. Hence, it is unable to reason about causation.

To understand how interaction helps with reasoning about causation, suppose that we are interested in evaluating whether a particular drug can cause improvements in patients with some disease. In other words, we wish to measure the strength of a hypothesized causal relationship from treatment (whether a treatment is administered) to some measure of health. Suppose that there is risk of some factor that causes a patient to be more or less likely to take the treatment and also affects the patient's health. Such a factor is called a "confounder" in statistics. The confounder could be, for example, gender, age, or genetics. To be specific, suppose that the treatment for some disease is more appealing to women than men, and that women tend to recover more from the disease than men. In that case, gender is a confounder and failing to control for it will invalidate the results of a trial.

In many cases, however, we don't know what confounders might be lurking in the shadows, and there may be confounders that we cannot measure. There might be some unknown genetic effect, for example. We can't control for confounders that we can't measure or that we don't know exist. Is it hopeless, then, to evaluate whether a treatment is effective?

To guard against the risk of hidden confounders, Pearl and Mackenzie (2018) point out, active intervention is effective (when active intervention is possible), underscoring that interaction is more powerful than observation alone. We must somehow force the treatment on some patients and force the lack of treatment on others. Then controlling for the confounding factor is no longer necessary.

Randomized controlled trials (RCTs), are the gold standard for determining causation in medical treatments and many other problems. The way an RCT works is that a pool of patients is selected, and within that pool, a randomly chosen subset is given the drug and the rest are given an identical looking placebo. Ideally, both the patients and the medical personnel are unaware of who is getting the real drug, and the choice is truly random, unaffected in any way by any characteristic of the patients. The system is now interactive because we have forced the value of one of the variables, whether the drug is taken, for each of the patients, and then we observe the results.

RCTs are actually routinely used in software today. It is common at Facebook, for example, when considering a change to the user interface, to randomly select users to whom a variant of the user interface is presented. The reactions of the users, whether they click on an ad, for example, can be measured and compared to a control group, which sees the old user interface. In this way, Facebook software can determine whether some feature of a user interface causes more clicks on ads. This process can be automated, enabling the software to experiment and learn what causes users to click on ads. This is a much more powerful form of reasoning than mere correlation, and it can result in software designing and refining its own user interfaces. The software can even learn to customize the interface for individual users or groups of users. This software is not realizing a Turing-Church computation because it is interactive. The users are an intrinsic part of the system.

It is not always possible or ethical to conduct an RCT. Pearl and Mackenzie (2018) document the decades-long agonizing debate over the question of whether smoking causes cancer. Had it been possible or ethical to randomly select people and make them smoke or not smoke, the debate may have been over much earlier. Instead, we were stuck with tragic observation, watching millions die.

6.5. Act to Sense

Two of the three limitations in human rationality identified by Simon (1972), uncertainty about the consequences that would follow from alternative decisions and incomplete information about the set of alternatives, reflect limited information about the environment. Simon zeroed in on the limited ability humans have to *process* information from the environment, but there are also limitations in our ability to *sense*, to gather information from the environment.

Goddfrey-Smith (2016) tells us that sensing is greatly enhanced by feedback. He points out that you need not just *sense-to-act* connections, which even bacteria have, but also *act-to-sense*. To have cognitive function, you have to affect the physical world and sense the changes. Sense-to-act is open loop; you sense, you react. Combine this with act-to-sense, and you close the loop, creating a feedback system.

Goddfrey-Smith (2016) gives a rather nice example of act-to-sense in cephalopods, such as cuttlefish and octopuses, which can change the color of their skin for camouflage and communication. It turns out that cuttlefish are colorblind,

having only a single type of photoreceptor molecule. But these molecules are also found in the skin, and by modulating the chromatophores to change the color of the skin, the cephalopod creates a color filter for the light that enters the skin. Dynamically varying the filter reveals the color distribution of the incoming light. They "see" color through their skin *via* a sense-to-act, act-to-sense feedback loop.

Turing-Church computations can only sense-to-act. The formalism does not include any mechanism by which the computation can affect its own inputs.

6.6. Efference Copies

Sense-to-act and act-to-sense feedback loops are present in many higher level cognitive functions. Since at least the 1800s, psychologists have studied the phenomenon that the brain can internally synthesize stimulus that would result from sensing some action commanded by the brain. This internal feedback signal is called an "efference copy." In speech production, for example, while the body is producing sounds that the ears are picking up, at the same time, the brain generates an efference copy, according to this theory, which is fed back into a different part of the brain that calculates what the ears should be hearing, an "expected reafference." The brain then compensates, adjusting the motor efference to make the speech sound more closely resemble the expectation.

Many psychologists today believe that efference copies help distinguish self-induced from not self-induced sensory stimulus. All animals with sensors have evolved some form of efference copy mechanism because otherwise they would react to their own actions as if those actions were imposed by their environment.

The importance of the efference copy has been understood at some level since the nineteenth century. Grüsser (1995) gives a history, crediting a book by Johann Georg Steinbuch (1770–1818) that illustrated the essential concept with a simple experiment. He noticed that if you hold your hand still and roll an object around in it, say, a spoon, you will not be able to recognize the object from the sensations coming from your hand. But if you actively grasp and manipulate the object, you will quickly recognize it as a spoon. The motor efference, therefore, must play a role in recognition, which implies that the motor efference must be fed back to the sensory system.

Central to this thesis is that our knowledge of the world around us is not solely determined by stimulus that happens to arrive at our sensory organs, but rather is strongly affected by our actions. Without these feedback loops, we would not only suffer limited ability to perform the information processing on our inputs, but we would also have fewer and less meaningful inputs. Turing-Church computations have no efference copies and hence no mechanisms for gathering these more meaningful inputs.

6.7. Indiscernable Differences

Lest the reader assume that act-to-sense only makes sensing more efficient, I will now give two rather technical demonstrations that act-to-sense enables making distinctions that are not discernible without feedback. The first of these is the Brock-Ackerman anomaly, a well-known illustration in computer science that observation alone cannot tell the difference between two significantly different systems.

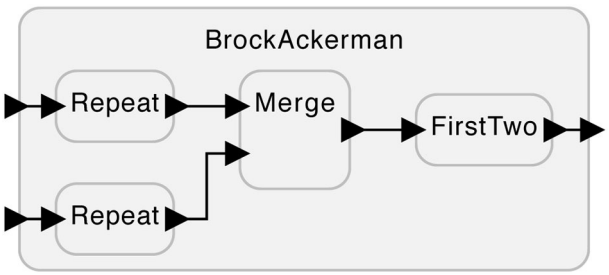


FIGURE 1 | A system with two variants that cannot be distinguished without feedback.

Consider the system shown **Figure 1**. This system has two inputs at the left that can accept sequences of numbers and one output at the right that produces a sequence of numbers. The subsystems labeled "Repeat" take each input number and repeat it twice on their outputs. For example, if the input to the top Repeat is the sequence (1, 2), the output will be the sequence (1, 1, 2, 2). The subsystem labeled "Merge" arbitrarily interleaves the input sequences it receives on its two inputs. For example, given the two input sequences (1, 2) and (3, 4), it can produce any of (1, 2, 3, 4), (1, 3, 2, 4), (1, 3, 4, 2), (3, 4, 1, 2), (3, 1, 4, 2), or (3, 1, 2, 4). If merge receives nothing on one of its inputs, then it will simply produce whatever it receives on the other input. The subsystem labeled "FirstTwo" simply outputs the first two inputs it receives. For example, given (1, 2, 3, 4), it will produce (1, 2).

Brock and Ackerman (1981) then gave two subtly different realizations of the FirstTwo subsystem:

1. The first realization produces outputs as it receives inputs. That is, as soon as it sees a 1 on its input, it will produce 1 on its output.
2. The second realization waits until there are two inputs available before producing any output. That is, it will not produce any output until both 1 and 2 are available, at which point it will produce the sequence (1, 2).

To an outside observer that can only passively watch the behavior of this system, these two realizations are indistinguishable. The possible output sequences are exactly the same for the same input sequences. For example, if the system is presented with inputs (5) and (6), i.e., two sequences of length one, the possible outputs for either realization are (5, 5), (5, 6), (6, 5), and (6, 6). The choice of realization has no effect on these possibilities.

Nevertheless, the two realizations yield different behaviors in some circumstances. Consider the system in **Figure 2**. The subsystem labeled "Increment" simply adds one to each input. For example, given the input sequence (1, 2), it will produce (2, 3). In this usage, it makes a difference which of the two realizations of FirstTwo is used. Suppose that the subsystem labeled "Source" provides on its output the length-one sequence (5). Under realization (1) of FirstTwo, there are two possible outputs from the BrockAckerman subsystem, (5, 5)

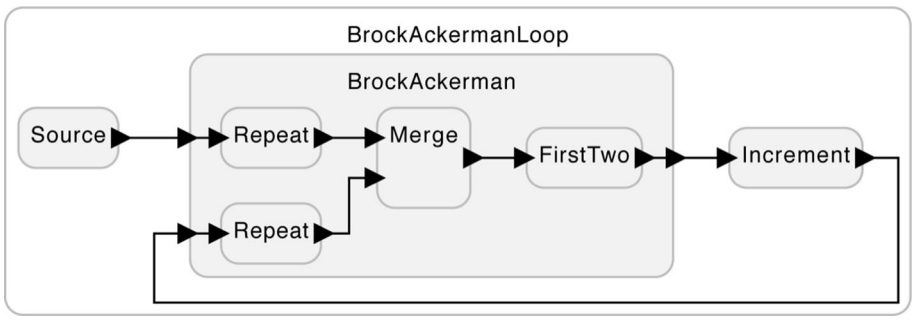


FIGURE 2 | A use of the system in **Figure 1** where the two variants of the FirstTwo subsystem yield different behaviors.

and (5, 6). But under realization (2), there is only one possible output, (5, 5).

With this example, Brock and Ackerman (1981) proved that two systems that are indistinguishable by a passive observer cannot be substituted one for the other without possibly changing the behavior. They are not equivalent. The feedback of **Figure 2** can be thought of as an "embodiment," where, by interacting with its environment, an otherwise indiscernible difference becomes evident.

Note that the Brock-Ackerman system is not a Turing-Church computation because of the non-deterministic Merge subsystem. The Turing-Church theory admits no such non-determinism. In Chapter 12 of Lee (2020), I show that a passive observer, one that can only see the inputs and outputs of a system, cannot tell the difference between such a non-deterministic system and a deterministic one (a deterministic one *would* be a Turing-Church computation). Only through interaction with the system is it possible to tell the difference. That argument depends on another celebrated result in computer science, Milner's concept of bisimulation.

6.8. Milner's Bisimulation

Milner (1980) developed a relation between systems that he called "simulation," where one system *A* "simulates" another *B* if, given the same inputs, *A* can match the outputs that *B* produces. Park (1980) noticed that there exist systems where *A* simulates *B* and *B* simulates *A*, but where the two systems are not identical. As with the Brock-Ackerman anomaly, the difference between the two systems is indiscernible to a passive observer, but discernible if you can interact with the system. This prompted (Milner, 1989) to revamp his system of logic and develop a stronger form of equivalence that he called "bisimulation." He then proved that any two systems that are "bisimilar" are indistinguishable not only to any observer, but also to any interactor. Sangiorgi (2009) gives an overview of the historical development of this idea, noting that essentially the same concept of bisimulation had also been developed in the fields of philosophical logic and set theory.

In Chapter 12 of Lee (2020), I give two possible models of tiny universes, the smallest imaginable universes where one entity in the universe is capable of modeling another entity in the same universe. I show two variants of entities in such a tiny universe, one where it is possible that the entity has free will, and one where the entity cannot possibly have free will. I then show that by passive observation alone, it is impossible to tell which entity you are modeling. But if interaction is allowed (using a bisimulation relation), the difference between the two entities can eventually become discernible to any desired degree of certainty. The two entities are not bisimilar. Without detailed knowledge of the inner structure of the entity being modeled, it is not possible to achieve 100% confidence in any conclusion about which entity is being modeled, but through repeated experiments, it is possible to get as close to 100% as you like.

Milner's simulation and bisimulation relations are relations between the possible *states* of two systems. Stretching a bit, one can imagine using these concepts to more deeply understand the relationship between mental states in a cognitive mind and the outside world that those states refer to. Philosophers use the term "intentionality" for such relationships, "the power of minds to be about, to represent, or to stand for, things, properties and states of affairs" outside the mind (Jacob, 2014). Searle (1983) argues that intentionality is central to cognition. Intentionality is about models of the universe that we construct in our brains. Dennett (2013) suggests the less formal term "aboutness" for intentionality. The relationship between mental states and the things that those states are about is essentially a modeling relationship. Milner shows us that such modeling works better when there is dialog, bidirectional interaction, or feedback. It may be that intentionality would likely not arise in a brain that can only observe the world. It must also be able to affect the world.

7. CONCLUSIONS

In Simon's "bounded rationality," rationality is the principle that humans make decisions on the basis of step-by-step (algorithmic) reasoning using systematic rules of logic to maximize utility. It becomes natural to equate rationality with Turing-Church computation. However, Turing-Church computation provides no mechanism for *interaction* or *feedback*, where the process provides outputs to its environment that then affect its inputs. The principle of embodied cognition suggests that human decision makers make use of feedback mechanisms for many of our cognitive functions, including rational decision making. Embodied bounded rationality, therefore, suggests that a rational

decision maker goes beyond Turing-Church computation, even if the ability to handle computational complexity is limited.

I have given a series of illustrations that show that interaction enables capabilities that are inaccessible to Turing-Church computation, including controlling a system in an uncertain environment, reasoning about causation, solving some undecidable problems, and discerning distinctions between certain kinds of systems. Bounded rationality, therefore, is not the same as a limited capacity to carry out Turing-Church computations because rational processes with feedback are capable of things that Turing-Church computations are not.

Interaction is the core idea in embodied cognition, which posits that a cognitive mind *is* an interaction of a brain with its body and environment. So, while it is true that the human brain has limited Turing-Church computational capability, it also transcends such computation by interacting with its body and environment. Key features of cognition, such as the ability to distinguish self from non-self and the ability to reason about causation, depend on such interaction. Since such interaction is missing from the Turing-Church theory of computation, the "universality" of such computation falls far short of true universality.

Deep neural networks, which have led to a revolution in artificial intelligence, are both interactive and not fundamentally algorithmic. Their ability to mimic some cognitive capabilities far better than prior algorithmic techniques based on symbol manipulation ("good old-fashioned AI") provides empirical evidence for the power of embodied bounded rationality.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The author thanks Rhonda Righter and reviewers for helpful suggestions on an earlier version of this article.

REFERENCES

Bekenstein, J. D. (1973). Black holes and entropy. *Phys. Rev. D* 7, 2333–2346. doi: 10.1103/PhysRevD.7.2333

Black, H. S. (1934). Stabilized feed-back amplifiers. *Electric. Eng.* 53, 114–120. doi: 10.1109/EE.1934.6540374

Bongard, J., Zykov, V., and Lipson, H. (2006). Resilient machines through continuous self-modeling. *Science* 314, 1118–1121. doi: 10.1126/science.1133687

Brock, J. D., and Ackerman, W. B. (1981). "Scenarios, a model of non-determinate computation," in *Conference on Formal Definition of Programming Concepts* (Peniscola: Springer-Verlag), 252–259. doi: 10.1007/3-540-10699-5_102

Brooks, R. A. (1992). "Artificial life and real robots," in *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, eds F. J. Varela and P. Bourgine (Cambridge, MA: MIT Press), 3–10.

Bryson, A. E., Denham, W. F., Carroll, F. J., and Mikami, K. (1961). A steepest-ascent method for solving optimum programming problems. *J. Appl. Mech.* 29, 247–257. doi: 10.1115/1.3640537

Church, A. (1932). A set of postulates for the foundation of logic. *Ann. Math.* 32, 346–366. doi: 10.2307/1968337

Clark, A. (2008). *Supersizing the Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195333213.001.0001

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19. doi: 10.1093/analys/58.1.7

Copeland, B. J. (2017). *The Church-Turing Thesis*. The Stanford Encyclopedia of Philosophy.

Danziger, S., Levav, J., and Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6889–6892. doi: 10.1073/pnas.1018033108

Dennett, D. C. (2013). *Intuition Pumps and Other Tools for Thinking*. New York, NY: W. W. Norton Co., Ltd.

Dodig Crnkovic, G. (2012). Information and energy/matter. *Information* 3, 751–755. doi: 10.3390/info3040751

Dodig-Crnkovic, G. (2018). "Cognition as embodied morphological computation," in *Philosophy and Theory of Artificial Intelligence*, ed V. C. Müller (Cham: Springer), 19–23. doi: 10.1007/978-3-319-96448-5_2

Dodig-Crnkovic, G., and Giovagnoli, R., (eds). (2013). *Computing Nature: Turing Centenary Perspective*. Berlin; Heidelberg: Springer.

Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York, NY: Basic Books.

Dreyfus, H. L., and Dreyfus, S. E. (1984). From Socrates to expert systems. *Technol. Soc.* 6, 217–233. doi: 10.1016/0160-791X(84)90034-4

Dreyfus, H. L., and Dreyfus, S. E. (1986). *Mind Over Machine*. New York, NY: Free Press.

Dreyfus, S. (1962). The numerical solution of variational problems. *J. Math. Anal. Appl.* 5, 30–45. doi: 10.1016/0022-247X(62)90004-5

Gallese, V., Mastrogiorgio, A., Petracca, E., and Viale, R. (2020). "Embodied bounded rationality," in *Routledge Handbook of Bounded Rationality*, ed R. Viale (London: Taylor & Francis Group), 14. doi: 10.4324/9781315658353-26

Geilen, M., and Basten, T. (2003). "Requirements on the execution of Kahn process networks," in *European Symposium on Programming Languages and Systems* (Warsaw: Springer), 319–334. doi: 10.1007/3-540-36575-3_22

Goddfrey-Smith, P. (2016). *Other Minds*. New York, NY: Farrar, Straus and Giroux.

Goyal, P. (2012). Information physics? Towards a new conception of physical reality. *Information* 3, 567–594. doi: 10.3390/info3040567

Grüsser, O.-J. (1995). "On the history of the ideas of efference copy and reafference," in *Essays in the History of Physiological Sciences: Proceedings of a Symposium Held at the University Louis Pasteur*, Strasbourg, ed C. Debru (The Wellcome Institute Series in the History of Medicine: Clio Medica) (Brill: London), 35–56. doi: 10.1163/9789004418424_006

Haugeland, J. (1985). *Artificial Intelligence*. Cambridge, MA: MIT Press.

Hofstadter, D. (2007). *I Am a Strange Loop*. New York, NY: Basic Books.

Jacob, P. (2014). *Intentionality*. Stanford, CA: Stanford Encyclopedia of Philosophy.

Kahn, G., and MacQueen, D. B. (1977). "Coroutines and networks of parallel processes," in *Information Processing*, ed B. Gilchrist (Amsterdam: North-Holland Publishing Co.), 993–998.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Kelley, H. J. (1960). Gradient theory of optimal flight paths. *ARS J.* 30, 947–954. doi: 10.2514/8.5282

Lee, E. A. (2008). "Cyber physical systems: design challenges," in *International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC)* (Orlando, FL: IEEE), 363–369. doi: 10.1109/ISORC.2008.25

Lee, E. A. (2016). Fundamental limits of cyber-physical systems modeling. *ACM Trans. Cyber Phys. Syst.* 1, 26. doi: 10.1145/2912149

Lee, E. A. (2017). *Plato and the Nerd*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11180.001.0001

Lee, E. A. (2020). *The Coevolution: The Entwined Futures of Humans and Machines*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/12307.001.0001

Lloyd, S. (2006). *Programming the Universe*. New York, NY: Alfred A. Knopf.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.* 20, 130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Maturana, H., and Varela, F. (1980). *Autopoiesis and Cognition*. Dortrecht; Boston, MA; London: D. Reidel Publishing Company. doi: 10.1007/978-94-009-8947-4

McGeer, T. (2001). Passive dynamic walking. *Int. J. Robot. Res.* 9, 62–82. doi: 10.1177/027836499000900206

Miller, J. F., and Thomson, P. (2000). "Cartesian genetic programming," in *European Conference on Genetic Programming* (Edinburgh: Springer), 121–132. doi: 10.1007/978-3-540-46239-2_9

Milner, R. (1989). *Communication and Concurrency*. Englewood Cliffs, NJ: Prentice Hall.

Milner, R. (1980). *A Calculus of Communicating Systems*. Berlin; Heidelburg: Springer. doi: 10.1007/3-540-10235-30

Müller, V. C., and Hoffmann, M. (2017). What is morphological computation? On how the body contributes to cognition and control. *Artif. Life* 23, 1–24. doi: 10.1162/ARTL_a_00219

Newell, A., and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. ACM* 19, 113–126. doi: 10.1145/360018.360022

Park, D. (1980). "Concurrency and automata on infinite sequences," in *Theoretical Computer Science*, Vol. 104 (Berlin; Heidelberg: Springer), 167–183.

Parks, T. M. (1995). *Bounded scheduling of process networks* (Ph.D. thesis). UC Berkeley, Berkeley, CA, United States.

Pearl, J., and Mackenzie, D. (2018). *The Book of Why*. New York, NY: Basic Books.

Pfeifer, R., and Bongard, J. (2007). *How the Body Shapes the Way We Think*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/3585.001.0001

Piccinini, G. (2007). Computational modelling vs. computational explanation: is everything a Turing machine, and does it matter to the philosophy of mind? *Austral. J. Philos.* 85, 93–115. doi: 10.1080/00048400601176494

Piccinini, G. (2020). *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford, UK: Oxford University Press. doi: 10.1093/oso/9780198866282.001.0001

Popper, K. (1959). *The Logic of Scientific Discovery*. New York, NY: Hutchinson & Co.; Taylor & Francis. doi: 10.1063/1.3060577

Rheingold, H. (2000). *Tools for Thought*. Cambridge, MA: MIT Press.

Sangiorgi, D. (2009). On the origins of bisimulation and coinduction. *ACM Trans. Program. Lang. Syst.* 31, 15:1–15:41. doi: 10.1145/1516507.1516510

Searle, J. R. (1983). *Intentionality*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9781139173452

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489. doi: 10.1038/nature16961

Simon, H. A. (1972). "Theories of bounded rationality," in *Decision and Organization*, eds C. B. McGuire and R. Radner (Amsterdam: North-Holland Publishing Company), 161–176.

Simon, H. A. (2000). Bounded rationality in social science: TODAY and tomorrow. *Mind Soc.* 1, 25–39. doi: 10.1007/BF02512227

Stewart, J. (1995). Cognition = life: implications for higher-level cognition. *Behav. Process.* 35, 311–326. doi: 10.1016/0376-6357(95)00046-1

Taleb, N. N. (2010). *The Black Swan*. New York, NY: Random House.

Tanaka, G., Yamane, T., Héroux, J. B., Nakane, R., Kanazawa, N., Takeda, S., et al. (2019). Recent advances in physical reservoir computing: a review. *Neural Netw.* 115, 100–123. doi: 10.1016/j.neunet.2019.03.005

Thelen, E. (2000). Grounded in the world. *Infancy* 1, 3–28. doi: 10.1207/S15327078IN0101_02

Turing, A. M. (1936). On computable numbers with an application to the entscheidungsproblem. *Proc. Lond. Math. Soc.* 42, 230–265. doi: 10.1112/plms/s2-42.1.230

Virk, R. (2019). *The Simulation Hypothesis: An MIT Computer Scientist Shows Why AI, Quantum Physics, and Eastern Mystics All Agree We Are in a Video Game*. Bayview Books.

Wachter, S., Mittelstadt, B., and Floridi, L. (2017). *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*. International Data Privacy Law. doi: 10.1093/idpl/ipx005

Wegner, P. (1998). Interactive foundations of computing. *Theoret. Comput. Sci.* 192, 315–351. doi: 10.1016/S0304-3975(97)00154-0

Wilson, D. G., Cussat-Blanc, S., Luga, H., and Miller, J. F. (2018). "Evolving simple programs for playing Atari games," in *The Genetic and Evolutionary Computation Conference (GECCO)* (Kyoto). doi: 10.1145/3205455.3205578

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.

Copyright © 2022 Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 06 May 2022
doi: 10.3389/fpsyg.2022.814844

Emergence and Embodiment in Economic Modeling

Shabnam Mousavi[1*†] *and Shyam Sunder*[2†]

[1] *Center for the History of Emotions, Max Planck Institute for Human Development, Berlin, Germany,* [2] *School of Management, Yale University, New Haven, CT, United States*

Exploratory ventures outside the established disciplinary boundaries can yield added insights and explanatory power. Imposing cognitive limitations on human logical reasoning ability (bounded rationality) is a well-known case in point. Extending cognition to parts of body outside the brain, and to environment outside the body is another. In contrast, the present article takes a constructive approach, also in an exploratory spirit. For the sake of exposition, we consider three tiered realms of scientific inquiry: physical or inanimate, biological or animate, and socio-psychological or sentient. In this three-tier framework, we explore the extent of gains in modeling human action within the confines of physical principles such as optimization. In this exercise, concepts of complexity and emergence account for the absence of analytically derivable mapping from micro or finer grain phenomena to macro or coarser grain phenomena. A general notion of embodiment captures the inclusion of a more expansive range of explanatory factors in modeling and understanding a given phenomenon. Emergence and embodiment play complementary roles in exploration of human behavior.

Keywords: embodiment, emergence, modeling behavior, three tiers, optimization

OPEN ACCESS

Edited by:
Riccardo Viale,
University of Milano-Bicocca, Italy

Reviewed by:
Antonio Mastrogiorgio,
IMT School for Advanced Studies Lucca, Italy
Matteo Convertino,
Tsinghua University, China

***Correspondence:**
Shabnam Mousavi
shabnam@jhu.edu

†ORCID:
Shabnam Mousavi
orcid.org/0000-0001-5664-7821
Shyam Sunder
orcid.org/0000-0001-8623-6409

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *14 November 2021*
Accepted: *21 March 2022*
Published: *06 May 2022*

Citation:
Mousavi S and Sunder S (2022) Emergence and Embodiment in Economic Modeling. Front. Psychol. 13:814844. doi: 10.3389/fpsyg.2022.814844

INTRODUCTION

Conceptual foundations of cognitive science of human (and animal) behavior rest on two assumptions to locate cognition in brain: objects in the environment being represented as symbols in the brain, and the brain functioning as a computer to process these symbols. During the past half-a-century it has been suggested that parts of the body outside the brain, as well as the environment outside the body, play a role in cognition. Furthermore, evidence points to the possibility of this dependence of cognition on extra-cranial parts of the body and external environment being structural, and not merely causal (Viale, 2012, 2014; Gallese and Cuccio, 2015; Varela et al., 2017; Gallagher, 2020; Vincini and Gallagher, 2021). The conceptual extensions beyond the traditional confines of cognitive science (representation and computation inside the brain) to other parts of the body and to the larger environment have taken several partially overlapping approaches (Wilson, 2002) under the labels of embodied, embedded, extended and enacted (collectively referred to as the 4Es in Newen et al., 2018), as well as distributed (Hutchins, 1995) and situated cognition (Gallagher and Varga, 2020). These developments either reject or reconfigure traditional cognitivism (Menary, 2010).

In this article, we ignore the distinctions among the diverse arguments and theories listed at the end of the paragraph above, and use "embodiment" as a common label for them in the meaning given therein. While obviously unsatisfactory for discussion of cognition, it would suffice for our objective of exploring the complementary role of embodiment and emergence in modeling human behavior. We use these two terms in the following intended meanings. Manifested within as well

as between tiers, emergence is the phenomenon of complex interplay among individual (or finer grain) elements giving rise to distinct coarser grain or aggregate level phenomena with properties absent in the parts. Embodiment implies that cognition is not limited to the brain, but includes parts of the body outside the brain, as well as elements of the environment outside the body and social interactions.

Emergence appears often in analyses of markets (Gode and Sunder, 1993; Sunder, 2006; Smith, 2008, 2009, 2010), and in complexity economics (a term coined by Doyne Farmer[1]). Viewing mental abilities (cognition) as emergent phenomena has precedent in cognitive science, development psychology and artificial intelligence research for many decades (Clark, 1997; McClelland, 2010). While institutions' role as location and enablers of emergence that extend agents' own minimal cognition is compatible with embodied cognition (Gilbert and Terna, 2000; Gallagher et al., 2019), on the whole, the idea of embodiment is relatively newer in economics than emergence. Moreover, with the exception of cognitive economy (Rosch, 1978), which is instantiated through embodiment, the mainstream cognitive economics, like cognitive psychology, is primarily focused on what is in the mind (Kimball, 2015). Overall, work that combines embodiment with emergence remains scarce (for philosophical instances see Garrison, 2022; Ryan, 2022).

It may be useful to start with a thumbnail sketch of developments in modeling human behavior in economics. Axiomatization of choice by von Neumann and Morgenstern (1944) was expanded to include subjective expected utility by Savage (1954). To date, expected utility theory (EUT) remains the corner stone of economic analysis of human behavior and the economic theory of choice (on its empirical failure, see Friedman et al., 2014). Given the ubiquity of methodological individualism and the concomitant psychological foundations of microeconomics, rise of cognitive science in the middle of the twentieth century led to a behavioral critique of economic theory. It was rooted in the discrepancies between the psychological assumptions about human decision-making on one hand and observed human cognitive abilities on the other. By incorporating known limitations of human cognitive abilities, bounded rationality was introduced as a revised framework for economics (and related aspects of other social sciences) to reshape the classical microeconomic approach, which had remained rooted in unbounded cognitive abilities (Simon, 1957). In other words, bounded rationality sought to improve the explanatory power of economic models using the accepted cognitive science framework, referred to as cognitivism, that keeps cognition firmly located in the brain (Mousavi and Garrison, 2003). Section Economic Models Keep Cognition in the Brain expands on this and provides the larger context of this development. We believe that the inclusion of the roles of extra-cranial and environmental phenomena in the expanded conceptual scheme of embodied cognition call for revisiting its implications for the use of economic theory to organize observed phenomena. Extend the current economic theory in this manner promises to produce better explanatory power and newer insights. However, that ambitious task is beyond the scope of this article.

Instead of expanding outside the traditional boundaries, this article takes a confining approach to the study of human behavior. More specifically, we take a few steps back, away from higher faculties such as intention and cognition, and even from evolutionary and other biological attributes. Limiting our exploration within the boundaries of inanimate existence, we examine just how much can be understood and what can be gained from modeling human action by framing it only in physical terms. This is not a reductionist approach; we remain fully cognizant of the aspects of behavior that cannot be understood without biology and socio-psychology. What we want to emphasize is to keep the interpretability of principles of every discipline within its confines, while also allowing their use as structuring tools on the outside. An example of a powerful organizing tool is in the domain of physical sciences (Sunder, 2006). To illustrate, we use the principle of least action from physics to reconfigure some extant models, and to compare that to a fully physical representation of the same phenomena. This reconfiguration does not enhance the explanatory power; instead, it helps address the well-known criticism of using optimization to model human behavior in economics on grounds of limited cognition. Once the cognitive or biological element is not a part of the model, questions about the applicability of optimization to modeling human behavior lose their relevance. Distinguishing reality from models, our physics-first approach is implemented in a three-tier framework. It is introduced in the next section, where the prevalence of cognitivism in modeling behavior is discussed. Section Where We Start Modeling Matters focuses on the shared physical existence of animate and inanimate worlds, where optimization is a powerful explanatory principle. The takeaway is not that all phenomena can be reduced to their physical existence. Section Causality Is Also in Tandem with Modeling Direction discusses causality. Section Methodological Individualism: Trouble No More argues that starting the modeling of human behavior from physical domain recasts longstanding concerns about methodological individualism in a new light by adopting an approach that keeps principles of each discipline within its boundaries. Section Embodiment Is an Outward Approach portrays embodiment as an outward perspective, and Section Concluding Remarks concludes.

ECONOMIC MODELS CONFINE COGNITION IN THE BRAIN

Human mind is thought to operate in a neurobiological brain in the physical world and with culture in the social world. Its existence comprises multiple discrete interacting layers (Anderson, 2007; Frégnac and Laurent, 2014). Cognitive capabilities are not only traced in the activities of neurons in the brain (e.g., fMRI), but also in muscle memory that produces embodied automatic behavior (e.g., athletic training). Embodied cognition offers an alternative framework that admits cognition to extend outside the brain. Such extension is largely missing in economics. More than a century of scientizing of

[1]See http://www.doynefarmer.com/book.

economics in the image of mathematics and physics has focused on refining reverse-engineered models of human behavior from observations. By and large, neuroscientists are actively engaged in similar activities (on neuroeconomics, see McCabe, 2008). Along the way, behavioral economics rose by focusing on imperfections of such models, and attempted to increase explanatory power by drawing on evidence from cognitive psychology and other disciplines. These efforts have been criticized for concentrating largely on blind alleys of unusual or contrived experiences covered neither by human evolution (Aumann, 2019) nor life experiences of most individuals. In the terminology of behavioral economics, anomaly is a generic label for observations that deviate from the predictions of the expected utility theory about individual behavior. Complications, messier mathematics, unobservability of explanatory constructs and the consequent decline in intuitive appeal are the main challenges to increasing the explanatory power sought from accounting for an ever-expanding list of anomalies.

At the aggregate level, either average behavior or emergence from complex interactions among micro-behavior can become salient, making it more observable and easier to theorize about. But for making economic policy and supporting recommendations, ease of analysis is not without its own disadvantages. Applying equilibrium and optimization concepts from physics beyond its traditional inanimate realm—say, to sentient phenomena—has important consequences that merit scrutiny. Do we need to completely abandon optimization (or in general, tools of analysis in physics) to produce sensible results? We suggest a technical modification to the use of methods, instead.

In an attempt to revisit methods of modeling human behavior, as an instrument, we developed a simple three-tier framework (Mousavi and Sunder, 2019, 2020). Consider organizing observed phenomena in three tiers using metaphor of crust, mantle and core in planet earth: human actions are manifested in the crust, biology in the mantle, and physics in the core (see **Figure 1**). While subject matter of physics concerns the universe of inanimate matter and energy, including the smallest of particles, human behavior encompasses sentient phenomena, with biological perspective situated in between the two.

Note that the extant method, by and large, takes an inward approach to modeling human behavior: from crust to mantle to the core in the earth metaphor in **Figure 1**. For example, efforts to model altruism start with social-psychological attributes such as utility, reciprocity, empathy, and identity. Appeal to principles of biological evolution may contribute some additional explanatory power through survival of the species. Only then might the modeler resort to abstract mathematical apparatus from physics.

The applicability of optimization to human choice behavior has long been debated. The main defense lies in good performance and lack of better alternative (Stigler and Becker, 1977). In Grether and Plott's words:

> The fact that preference theory and related theories of optimization are subject to exception does not mean that they should be discarded. No alternative theory currently available appears to be capable of covering the same extremely broad range of phenomena. In a sense the exception [preference reversal] is an important discovery, as it stands as an answer to those who would charge that preference theory is circular and/or without empirical content. It also stands as a challenge to theorists who may attempt to modify the theory to account for this exception without simultaneously making the theory vacuous (Grether and Plott 1979, p. 629).

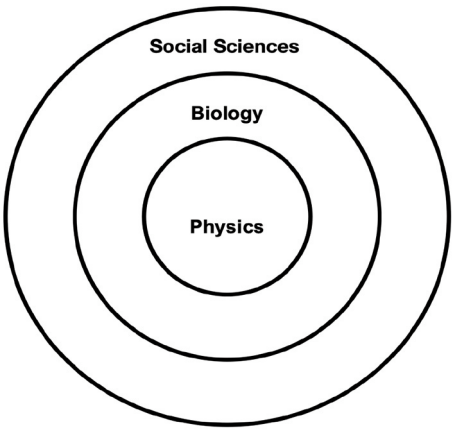


FIGURE 1 | Earth metaphor for the three-tier framework for modeling human action. The familiar direction is from the crust inwards and we explore the outward direction (Art by Anoush Kheirandish).

Notably, acknowledgment of cognitive limits in models of bounded rationality has not led to discarding optimization as a powerful tool for analysis. Models under bounded rationality paradigm also construct paths of action of satisficing agents that are optimal subject to their cognitive and procedural attributes. Moreover, cognition remains firmly located in the brain, in both bounded as well as unbounded paradigms.

Is optimization principle a legitimate tool for analyzing human behavior? We believe it can be, as long as it is confined to analysis in the physical core shared by animate and inanimate existence. Therefore, optimization presents a meaningful frame for modeling human action as long as its implications remain within that core. Doing so only requires the modeler to focus on the shared properties of matter and energy in the universe first, before attending to the biological and social-psychological characteristics of the animate and sentient phenomena. In context of the earth metaphor in **Figure 1**, we suggest an outward approach to deployment of tools of analysis and to confine the interpretability of each set of principles to its respective tier. This implies using optimization for analyzing human behavior, but only to capture the components that might be shared with inanimate phenomena as elaborated in the next section.

WHERE WE START MODELING MATTERS

Consider four examples from different domains, each presenting an action or end point of an action: (a) a jar filled with small smooth marbles, (b) the network of nerves among

TABLE 1 | Using the principle of least action to model catching a fly ball and the nematode nervous system (Source: Mousavi and Sunder, 2021).

	Method of modeling	WHAT: given variables	HOW: action element	Path of action
To catch a fly ball	*Current Method: Inward approach with three-tiers*	Time a fly ball takes to reach ∼1.5 m above ground	Use the evolutionary capacity of holding gaze on a moving object	A curved path, depending on when the angle of gaze is first fixed
	Proposed method: In the first physics tier only	Same as above	Keep a *fixed angle* of gaze (change = 0)	Same as above
Arrange nervous system network	*Current method: Inward approach in the second tier*	Location of ganglia in a combinatorial space	Economize the use of biological resources for connecting (ganglia)	A path of fiber connections with minimal length of connections
	Proposed method: In the first physics tier only	*Number of ganglia*	Minimize distance among ganglia and position them concurrently	Same as above

ganglia (nodes) of a nematode worm, (c) a baseball player running to catch a fly ball, and (d) iron filings on a plate in the force field surrounding a magnet. Now, let us explore how far can optimization takes us in organizing these four observed phenomena.

a) When a large jar full of small smooth marbles is shaken for a few seconds in a gravitational field, packing of its contents approaches a local optimum arrangement.
b) Connections among the ganglia in a nematode's nervous system are optimized to save wire:
   At multiple hierarchical levels—brain, ganglion, and individual cell—physical placement of neural components appears consistent with a single, simple goal: minimize cost of connections among the components. The most dramatic instance of this "save wire" organizing principle is reported for adjacencies among ganglia in the nematode nervous system; among about 40,000,000 alternative layout orderings, the actual ganglion placement in fact requires the least total connection length. In addition, evidence supports a component placement optimization hypothesis for positioning of individual neurons in the nematode, and also for positioning of mammalian cortical areas (Cherniak 1986, p. 1).
c) Cognitive scientists have used data gathered from the field to model how animals and humans catch fly balls and other moving objects. They keep a constant angle of gaze on the ball above the horizon while moving toward the ball until catching it. If we take an outward approach to model this phenomenon (to catch a fly ball) a mere minimization of changes in the angle of gaze fixed at the ball would suffice. However, the inward approach consists of the following elements: (1) cognitive attribution of catching the object by deploying the cognitive ability to hold the gaze on a moving object against a noisy background; (2) biological attribution: evolution of capabilities for preys to evade predators and for predators to catch their preys; and (3) physics scheme: solving an optimization problem with the objective of minimizing the change in the angle of gaze—ideally, keeping the angle fixed (for a comprehensive overview of the phenomena, see Hamlin, 2017).
d) Orientation of the iron filings aligned with the direction of the magnetic field represent an optimal outcome albeit subject to approximation depending on the size of the filings, strength of the field, and friction with the supporting surface.

All four seemingly disparate phenomena discussed above exhibit presence of optimization at work. Our proposed outward approach offers alternative ways of organizing a given phenomenon at various levels. In what follows, first we organize items b (the nervous system of a simple worm), and c (catching a fly ball) by using the physical principle of least action (PLA) and remaining confined to physical attributes. We also use PLA to organize the inward modeling approach for the same two phenomena. By juxtaposing the resulting structures, we show how this exercise can produce a method for comparing modeling elements among different tiers (see **Table 1**). Second, we generalize our three-tier framework to organize scientific inquiry across fields of study (see **Table 2**).

The first exercise demonstrates that using a physics principle for organization does not imply that all elements of the observed phenomenon need to be only physical. Let us examine how one physics principle can be used to organize and compare a physics-only model with a biology-based description. The principle of least action structures an observed phenomenon with specifying three elements: (1) an action element that is the argument of optimization, (2) given or fixed element(s) that are not affected by the action but constrain it, and (3) a path of movement on which the action is realized. **Table 1** lists the physical forms of these elements for catching of the fly ball and the nematode nervous system when the modeler remains confined in the physical core, as well as the biological (evolutionary) forms of the same element that are used in the familiar inward methods of modeling. Organized as such, comparison and connections among elements of modeling is straightforward. This can facilitate interdisciplinary communication and collaboration. Indeed, we consider our framework as a productive and generalizable method for detecting cross sections of scientific pursuits and initiating cross-disciplinary exchanges for virtually all fields of study.

The second exercise features thinking in terms of three tiers as a powerful tool that can be used for structuring not only modeling

TABLE 2 | Subject matter and principles in three domains of scientific inquiry (Source: Mousavi and Sunder, 2021).

Domain	Animate	Animate-Inanimate	Inanimate
Discipline	Social Sciences	Biology/Molecular Chem.	Physics/Chemistry
Subject matter	Person/group/institution	Large molecules/Cells/Organism/group	Matter and energy (detectable and dark)
Principles, concepts and terms	Theories of mind Perception and cognition Nature vs. nurture Demand and supply Behavior, labor, capital, trade, contract, judgment, personality, development State and society…	Evolution by natural selection (Matching) Longevity vs. reproduction Function of organs Anatomy and physiology DNA, RNA, cells, protein, life…	Least action Force fields Chemical binding Inertia and Symmetry Relativity Effort, flow, motion, time …
Shared features	Physical existence in all domains is subject to physical laws.		

practices concerned with human behavior, but also a general view of scientific inquiries into the inanimate, animate and sentient existence. Our attempt to organize fields of study in this manner is summarized in **Table 2**.

Where we start modeling matters. Economists traditionally, and psychologists increasingly produce policy recommendations and intervention designs. We argued that social scientists in general take an inward approach to modeling human behavior. This means that they can easily ignore the eventual effect of physical structures at work. This inattention can be consequential in a large scale, especially when the modeling and observation methods are assumed to be neutral with respect to the outcome, or independent of each other. Scientific observation at large has a history and evolving structure:

> The scientific observation of the organic world (including humans) went through three stages: first, intensive observation of very small samples (still pursued in primatology); second, statistical observation of large samples to extract averages (still used in much of social science); and third, observation of larger samples that focused on variability rather than erasing it with averages (striven by Darwin's insight that it is individual variability that drives evolution). All three modes of observation are still very much in use and often complement one another: for example, a puzzling statistical effect may need a more granular ethnographic study to discover the causal mechanisms at work[2].

Physicists are wary of the observer's role, and statisticians' motto is: if you beat the data long enough, it will eventually confess. Social scientists regularly talk about scientific facts derived from data. Both the sequence and the limits are of particular consequence in using results from physics models to draw societal implications. We therefore propose a careful observation of the sequence in which scientific insights are gained, cognizant of the similarities and differences between social, biological, and physical phenomena as summarized in **Table 3**.

Choice of starting point of modeling matters. What is and is not carried across the tiers of scientific inquiry also matters. Critiques of a physics-based approach to socio-economic phenomena can be recast by switching the direction of sequence of investigation from inward to outward. At a time when behavioral policymaking has spread far and wide in public (and private) sectors, it is refreshing to recount astrophysicist John Stewart's insights:

TABLE 3 | Properties of different phenomena.

Subject matter	Physical phenomenon	Biological and social phenomenon
Scientific inquiry		
Observation effect	Yes	Yes
Principle universality	Yes	No
Method neutrality	No	No
Explanatory equivalence	Yes	No

> There is no longer [an] excuse for anyone to ignore the fact that human beings, on the average and at least in certain circumstances, obey mathematical rules resembling in a general way some of the primitive "laws" of physics. "Social physics" lies within the grasp of scholarship that is unprejudiced and truly modern. When we have found it, people will wonder at the blind opposition its first proponents encountered.
> Meanwhile, let "social planners" beware! Water must be pumped to flow uphill, and natural tendencies in human relations cannot be combated and controlled by singing to them. The architect must accept and understand the law of gravity and the limitations of materials. The city or national planner likewise must adapt his studies to natural principles (Stewart, 1947, p. 485, emphasis added).

CAUSALITY IS ALSO IN TANDEM WITH MODELING DIRECTION

Just as it is customary in scientific practice to start analysis of action with attribution to most salient, immediate and proximate variables (an inward approach), it is not unusual to assume the arrow of causation and dynamics pointing in the opposite direction from such variables to observations, especially if the

[2]Personal correspondence with historian of science Lorraine Daston.

former carries an earlier time stamp. Consider this example from a textbook on biology for engineers on effort (cause) and flow (effect):

> There are two basic kinds of variables that describe the action of a physical system. Effort variables are those things that cause an action to occur. Flow variables are the responses to effort variables, usually involving movement but not always. For the simple case of a running animal, the effort variable is the force required to propel the animal; the flow variable is the velocity of movement. Heat loss from that same animal, which is the flow variable, occurs in response to a temperature difference, an effort variable. Sexual attraction to an animal of the opposite sex (effort variable) can result in a wide range of activities, including copulating (a flow variable). Hunger (an effort variable) can result in feeding (a flow variable). Thus, there are a wide variety of causes and effects related to biological activity, and these can be thought about in terms of effort and flow variables, which tend to simplify the concepts of biological activities. For any activity of a biological organism or system, searching for the effort variable, the flow variable, and relationships between these two can make it easier to comprehend not only how and why the activity occurs, but also the intensity of the activity (Johnson, 1941, p. 32–33).

This effort-flow frame captures a wide range of phenomena across domains from force and acceleration in Newtonian mechanics to motivation and work in social sciences. Extending this form of framing generates amusing views. For example, framed in economic terms, the outcome of sustainability can be achieved by optimizing on the flow variables of consumption and reproduction: "Consumption and reproduction have been and remain the basic values of human societies. These two lie at the root of our moral codes. Even virtue is promoted with the promise of entitlement to more consumption in the future. Development, prosperity and welfare are euphemisms for higher consumption" (Sunder 2012, p. 1).

Economics is the most physical of the social sciences, and has directly adopted physical terminology such as equilibrium, friction, efficiency. However, economics is not alone in this. It does not require much effort to trace conceptual links also between physical laws and other social sciences and with humanities.

The "path of least resistance" as the underlying principle of inductive sociology was introduced more than a century ago (Giddings, 1906). The linguist Zipf (1949) built the "biosocial physics" theory of human behavior whose principle of minimum effort yields the eponymous law of frequency distribution anywhere from of words in a language to city populations. Zipf considered mind as a system of "mentation", by analogy extended the philology of semantics in spoken language and cultural preconceptions to the structure of every human action. In the context of embodiment approach, psychologist Rosch proposed cognitive economy as the first of her two principles that govern how human being categorize their world of language, people, animals, vegetation, and just about everything else in order "to provide maximum information with least cognitive effort …" (Rosch 1978, p. 28). Similar analogical exercises have been undertaken with the concept of inertia that links effort-flow and capacity. Economist Bewley (1987, 2002) used inertia to formulate economist Knight's (1921/2006) notion of uncertainty. In general, for framing cognition inertia has long been considered as a fundamental law. In the words of Schiller (1846–1937):

> Our curious result of this inertia, which deserves to rank among the fundamental laws of nature, is that when a discovery has finally won tardy recognition it is usually found to have been anticipated, often with cogent reasons and in great detail (Johnson 1941, p. 35).

This very phenomenon is dubbed as the knew-it all-along effect in contemporary literature. Finally, the Lagrange principle for probability with constraint that views physics in terms of energy and entropy, is used in socio-physics to frame a wide range of phenomena across social sciences: planned vs. spontaneous, collective vs. individual, law as right vs. wrong, or order vs. disorder; society as bondage vs. freedom; and economics as rational vs. chances (Chakrabarti et al., 2006).

Overall, the inward deployment of principles—from social-psychological to biological to physical properties—has generated a body of coherent models, partially generalizable theories, as well as numerous ongoing disagreements. In the next section, we argue that our outward-confined approach is not an effort to answer a major critique to this practice, namely methodological individualism, instead it is a new way of thinking and organizing the matter that addresses the problem at hand in a different light.

METHODOLOGICAL INDIVIDUALISM: TROUBLE NO MORE

Philosopher of science, Longino, takes issue with the general thrust of modeling behavior:

> [T]he question [of behavioral sciences] is why people fall into one or the other of these categories, or fall into a particular range of a multiple-valued quantitative (more or less) trait. Behavioral sciences seek to answer this question. Even when the research methodology permits only correlations among behaviors and studied factors, it is intended ultimately to contribute to an understanding of the causes of behaviors. To ask about the causal influences on the expression of a trait in a population is already to be committed to an individualistic point of view.…factors maybe genetic, hormonal, neurological, or environmental. The question for researchers is how these factors influence an individual's disposition to respond to situations in one way or another (Longino, 2019, p. 4).

Methodological individualism is a cornerstone of much economic thought of the recent century. Starting with individual (human being) as the primary unit of axiomatization, actions and analysis, economic theory derives and predicts outcomes and economic behavior of organizations, markets, and societies.

In economics, as in other social sciences, engineering approach to designing the parts to serve the functions of the whole and building the whole from the parts takes the form of methodological individualism. Even when the macro or

coarser grain phenomena phenomenon is of primary interest, modeling starts with specifying attributes of the individual at micro level, where the "representative agent" manifests shared attributes. Macro outcomes of the model result from a constructivist process that derives properties of the collective from behavior of sophisticated individuals who demonstrate rationality in anticipation, learning, and goal-seeking. Reflecting on this common practice, economist Arrow highlights the social nature of all economic phenomena:

> In the usual versions of economic theory… seems commonly to be assumed methodological individualism, that it is necessary to base all accounts of economic interaction on individual behavior… A specific version of this has invaded other social sciences, under the name of rational-actor models. … [There exists] explicit advocacy of methodological individualism among the Austrian school…[and] useful implications of methodological individualism for positive economics. It is usually thought that mainstream economics is the purest exemplar of methodological individualism [but].… In fact, every economic model one can think of includes irreducibly social principles and concepts… social variables, not attached to particular individuals, [which] are essential in studying economy or any other social system… (Arrow 1994, p. 8).

Methodological individualism lies at the heart of choice theory and rests on two key psychological elements: preferences (whether static or adaptive) and choice of preferred alternative(s) from the individual's opportunity set specified by the environment. Preferences of an individual are a mapping from objects of choice to the real line, so each object is either more, or less or equally preferred to every other object under consideration. Preferences may be static, dependent on the state of the world, and may change over time according to a knowable law. Remarkably, if the law by which the preferences change is not knowable ex ante, anything goes, and they could not serve as a basis of a theory (for an engaging study, see Pastor-Bernier et al., 2017).

The rational-actor model encounters two hurdles in scaling up to social phenomena. As the number of agents increases, so do their opportunity sets, strategies, actions and interactions among them, rapidly rendering analysis of interactions intractable. Representative agents and other simplifying assumptions made to facilitate analysis add the risk of excluding important social dimensions. Second, when macro-level phenomena emerge from non-linear complex interactions among many parts, the properties of such outcomes may not be derivable and cannot be constructed from the micro-level properties.

Deployment of emergence and embodiment in tandem as modeling apparatus can be portrayed as follows. Emergence can be a tool for explaining social phenomenon that cannot be adequately captured by economic modeling of rational agents. Similarly, embodiment provides a perspective for cognitive psychology, and for behavioral sciences in general, to account for the context that may include the body of actors and their social interactions.

EMBODIMENT IS AN OUTWARD APPROACH

Embodiment literature, replete with irrelevance of optimization to understanding cognitive phenomena, is in conformity with our proposal to keep principles of each tier confined within. For example, when moving to the animate domain, evolutionary capacities can be understood through their viability not optimization:

> One of the more interesting consequences of this shift from optimal selection to viability is that the precision and specificity of morphological or physiological traits, or of cognitive capacities, are entirely compatible with their apparent irrelevance to survival. To state this point in more positive terms, much of what an organism looks like and is "about" is completely *underdetermined* by the constraints of survival and reproduction. Thus, adaptation (in its classical sense), problem solving, simplicity in design, assimilation, external "steering" and many other explanatory notions based on considerations of parsimony not only fade into the background but must in fact be completely reassimilated into new kinds of explanatory concepts and conceptual metaphors (Varela et al. 2017, p. 196).

It is commonplace to think of technology—fire, hammer, knife, eyeglasses, car, and telephone as external devices developed by humans over the ages and put to use to make their life easier and better. In this everyday perspective, the evolution of humans is confined to what is covered by their skin. However, this perspective can be questioned at three levels. First, to the extent tools and technologies enhance the ability of our own bodies to perform various functions, the former can be viewed as evolutionary extensions of the later. For example, hammer could be seen as evolution of hand, and bicycle as evolution of legs, where evolution extends outside the body. This perspective includes inanimate objects created by humans as extensions of humans themselves, fusing them across the matter-energy vs. biology boundary line.

Second, without crossing the inanimate/animate boundary, humans like other organisms have large microbiomes added to their "human" genetic endowment. The two parts of the total genome have evolved together to a state where their independent separate existence as life forms may not be viable. No humans are known to survive the destruction of gut bacteria. This relationship makes it difficult to decide if the genome of billions of microorganisms that reside in the gut of every human being are or are not to be regarded as a "part" of the human body.

Third, is the case of human body and the structure of societies in which humans live—such as family groups including two, three or more generations with specialization of work by age and gender. These social structures themselves could be seen as extensions of human evolution in a non-trivial sense.

All three kinds of evolution—within biology, between biological and social, and between biological and physical worlds—have a long history. Persuasive arguments have been made about their co-evolution. Early stone tools and electronic computers today are not only results of human brain but also helped shape the brain that created them. Same could be said

of co-evolving species in the animate world. Human body and mind and tools surely co-evolved with each other, as also with the structure of families and human societies. Human child, for example remains dependent for longer than any other animal, and development of tools and fire may have been related to the length of gestation of child rearing.

In sum, our examination of embodiment in the three-tier framework reveals the outward direction of deployment of scientific principles. It is in this manner that embodiment provides better understanding of human behavior as well as more powerful tools for describing and explaining observed phenomena.

CONCLUDING REMARKS

Familiar approach to modeling economic behavior starts with specifying social-psychological preferences and goals to construct an objective function, specifying the opportunity set by constraints, and then seeking optimal choice of action from the set. For example, an effort to understand the price and availability of coffee may start with attributing preferences to consumers, production technology to producers, and opportunity sets to them both, before deriving price and allocations from a model that attributes maximization of their respective goals—utility of consumers and profit of producers arising from their sentient and conscious nature. We introduced a three-tiered framework with physics in the core, biology in the middle, and socio-psychology on the top tier. Our framework characterizes this familiar method of modeling as an inward approach that originates in the crust with the possibility of proceeding to the biological middle, and to formalize uses the tools and principles from the physical core.

We discussed that optimization is a superb organizing principle for modelers. It manifests in the physical phenomena most vividly, for example in mechanics, sound, light, electricity, magnetism, and elementary particles. We set out to explore an outward approach to deployment of scientific principles. This attempt was motivated by the idea that if photons do not need cognitive equipment to optimize their paths from a candle to a book, there is no reason to presume in modeling—as is in the inward approach—that all human behavior must necessarily arise from animate adaptive and cognitive faculties at its physical level. Moreover, emergence of social phenomena and their properties, when optimal, can be decoupled from what is derivable from individual parts of the system. We argued that through an outward-confined approach, our three-tier framework organizes physical, biological and social science principles, proposing a new and broader perspective on human behavior, sans reductionism. Viewing embodiment and emergence as exploration methods that deploy scientific principles in an outward direction highlights their tandem role in scientific inquiry, each enriching insights into human behavior.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

SM acknowledges the Original-isn't it? grant from the Volkswagen Foundation.

ACKNOWLEDGMENTS

This article draws heavily on our discussion article, Framing Human Action in Physics: Valid Reconstruction, Invalid Reduction (https://cowles.yale.edu/sites/default/files/files/pub/d23/d2326.pdf). We acknowledge helpful comments and suggestions from two referees that helped shaping and sharpening our arguments. We are indebted to Jim Garrison and Frank Ryan for pointing us to related literature and for in-depth reflections on an earlier version of this article. The usual disclaimer applies.

REFERENCES

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York, NY: Oxford University Press.

Arrow, K. J. (1994). Methodological individualism and social knowledge. *Am. Econ. Rev.* 84, 1–9.

Aumann, R. J. (2019). A synthesis of behavioural and main stream economics. *Nat. Hum. Behav.* 3, 666–670. doi: 10.1038/s41562-019-0617-3

Bewley, T. F. (1987). *Knightian Decision Theory. Part II: Intertemporal Problems.* Cowles Foundation Discussion Paper No. 835.

Bewley, T. F. (2002). Knightian Decision Theory. Part I. *Decis. Econ. Fin.* 25, 79–110. doi: 10.1007/s102030200006

Chakrabarti, B. K., Chakraborti, A., and Chatterjee, A. (eds.). (2006). *Econophysics and Sociophysics.* Weinheim: Wiley.

Cherniak, C. (1986). *Minimal Rationality.* Cambridge, MA: MIT Press.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again.* Cambridge, MA: MIT Press.

Frégnac, Y., and Laurent, G. (2014). Neuroscience: where is the brain in the human brain project? *Nat. News* 513, 27. doi: 10.1038/513027a

Friedman, D., Isaac, M., James, D., and Sunder, S. (2014). *Risky Curves: On the Empirical Failure of Expected Utility.* Hoboken, NJ: Cambridge University Press.

Gallagher, S. (2020). "Phenomenology of agency and the cognitive sciences," in *The Routledge Handbook of the Phenomenology of Agency,* eds C. Erhard and T. Keiling (London; New York, NY: Routledge), 334–348.

Gallagher, S., Mastrogiorgio, A., and Petracca, E. (2019). Economic reasoning and interaction in socially extended market institutions. *Front. Psychol.* 10, 1856. doi: 10.3389/fpsyg.2019.01856

Gallagher, S., and Varga, S. (2020). The meshed architecture of performance as a model of situated cognition. *Front. Psychol.* 11, 2140. doi: 10.3389/fpsyg.2020.02140

Gallese, V., and Cuccio, V. (2015). "The paradigmatic body. embodied simulation, intersubjectivity and the bodily self," in *Open MIND*, eds T. Metzinger and J. M. Windt (Frankfurt: MIND Group), 1–23.

Garrison, J. (2022). "Transactional perspectivalism: the emergence of langauage, minds, selves and temporal sequence," in *Deweyan Transactionalism in Education: Beyond Self-action and Inter-action*, eds J. Garriosn, J. Öhman, and L. Östman (London: Bloomsbury Academic), 39–50. doi: 10.5040/9781350233348.ch-3

Giddings, F. H. (1906). *Readings in Descriptive and Historical Sociology*. New York, NY: Macmillan.
Gilbert, N., and Terna, P. (2000). How to build and use agent-based models in social science. *Mind Soc.* 1, 57–72. doi: 10.1007/BF02512229
Gode, D., and Sunder, S. (1993). Allocative efficiency of markets with zero intelligence traders: market as a partial substitute for individual rationality. *J. Polit. Econ.* 101, 119–137.
Grether, D. M., and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *Am. Econ. Rev.* 69, 623–638.
Hamlin, R. P. (2017). The gaze heuristic: biography of an adaptively rational decision process. *Top. Cogn. Sci.* 9, 264–288. doi: 10.1111/tops.12253
Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
Johnson, A. T. (1941). *Biology for Engineers.* Cleveland, OH: CRC Press.
Kimball, M. S. (2015). *Cognitive Economics*. NBER Working Paper No. 20834. Available online at: https://www.nber.org/system/files/working_papers/w20834/w20834.pdf (acessed October 12, 2021).
Knight, F. H. (2006). *Risk, Uncertainty and Profit*. New York, NY: Dover (1921).
Longino, H. E. (2019). Scaling up; scaling down: what's missing? *Synthese* 196, 1–15. doi: 10.1007/s11229-019-02249-y
McCabe, K. A. (2008). Neuroeconomics and the economics sciences. *Econ. Philos.* 24, 345–368. doi: 10.1017/S0266267108002010
McClelland, J. L. (2010). Emergence in cognitive science. *Topics* 2, 751–770. doi: 10.1111/j.1756-8765.2010.01116.x
Menary, R. (2010). *The Extended Mind*. Cambridge, MA: MIT Press.
Mousavi, S., and Garrison, J. (2003). Towards a transactional theory of decision making: creative rationality as functional coordination in context. *J. Econ. Methodol.* 10, 131–156. doi: 10.1080/1350178032000071039
Mousavi, S., and Sunder, S. (2019). *Physical Laws and Human Behavior: A Three-Tier Framework*. Cowles Foundation Discussion Paper No. (2173).
Mousavi, S., and Sunder, S. (2020). Physics and decisions: an inverted perspective. *Mind Soc.* 19, 293–298. doi: 10.1007/s11299-020-00244-2
Mousavi, S., and Sunder, S. (2021). *Framing Human Action in Physics: Valid Reconstruction, Invalid Reduction*. Cowles Foundation Discussion Paper No. (2326).
Newen, A., De Bruin, L., and Gallagher, S. (eds.). (2018). *The Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press.
Pastor-Bernier, A., Plott, C. R., and Schultz, W. (2017). Monkeys choose as if maximizing utility compatible with basic principles of revealed preference theory. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1766–E1775. doi: 10.1073/pnas.1612010114
Rosch, E. (1978). "Principles of categorization," in *Cognition and Categorization,* eds E. Rosch and B. Lloyd (Hinsdale, NJ: Lawrence Erlebaum Associations), 27–48.
Ryan, F. X. (2022). "Philosopher's problems: transaction in philosophy and life," in *Deweyan Transactionalism in Education: Beyond Self-action and Inter-action*, eds J. Garriosn, J. Öhman, and L. Östman (London: Bloomsbury Academic), 17–39. doi: 10.5040/9781350233348.ch-2
Savage, L. J. (1954). *The Foundations of Statistics*. New York, NJ: Dover.
Simon, H. A. (1957). *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York, NY: John Wiley.
Smith, V. L. (2008). *Rationality in Economics: Constructivist and Ecological Forms*. Cambridge: Cambridge University Press.
Smith, V. L. (2009). *Rationality in Economics: Constructivist and Ecological Forms*. Cambridge: Cambridge University Press.
Smith, V. L. (2010). *Rationality in Economics: Constructivist and Ecological Forms*. Cambridge: Cambridge University Press.
Stewart, J. Q. (1947). Empirical mathematical rules concerning the distribution and equilibrium of population. *Geograph. Rev.* 37, 461–485.
Stigler, G. J., and Becker, G. S. (1977). De Gustibus non est disputandum. *Am. Econ. Rev.* 67, 76–90.
Sunder, S. (2006). Economic theory: structural abstraction or behavioral reduction? *History Polit. Econ.* 38, 322–342. doi: 10.2139/ssrn.800786
Sunder, S. (2012). *Consumption, Numbers and Time: Arithmetic of Sustenance*. Remarks for Sarah Smith Memorial Conference, Yale Divinity School.
Varela, F. J., Thompson, E., and Rosch, E. (2017). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
Viale, R. (2012). *Methodological Cognitivism: Vol. 1: Mind, Rationality, and Society*. Berlin: Springer-Verlag.
Viale, R. (2014). *Methodological Cognitivism: Vol. 2: Cognition, Science, and Innovation*. Berlin: Springer-Verlag.
Vincini, S., and Gallagher, S. (2021). Developmental phenomenology: examples from social cognition. *Continent. Philos. Rev*. 54, 183–199. doi: 10.1007/s11007-020-09510-z
von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/bf03196322
Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. New York, NY: Hafner Publishing.

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.

Copyright © 2022 Mousavi and Sunder. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

TYPE Conceptual Analysis
PUBLISHED 17 November 2022
DOI 10.3389/fpsyg.2022.841972

Check for updates

OPEN ACCESS

EDITED BY
Riccardo Viale,
University of Milano-Bicocca, Italy

REVIEWED BY
Patrick Simen,
Oberlin College, United States
Pierluigi Cordellieri,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Samuel A. Nordli
snordli@indiana.edu

SPECIALTY SECTION
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

RECEIVED 23 December 2021
ACCEPTED 24 October 2022
PUBLISHED 17 November 2022

CITATION
Nordli SA and Todd PM (2022) Embodied and embedded ecological rationality: A common vertebrate mechanism for action selection underlies cognition and heuristic decision-making in humans. *Front. Psychol.* 13:841972. doi: 10.3389/fpsyg.2022.841972

COPYRIGHT
© 2022 Nordli and Todd. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Embodied and embedded ecological rationality: A common vertebrate mechanism for action selection underlies cognition and heuristic decision-making in humans

Samuel A. Nordli[1,2]* and Peter M. Todd[1,2]

[1]Cognitive Science Program, Indiana University, Bloomington, IN, United States, [2]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, United States

The last common ancestor shared by humans and other vertebrates lived over half a billion years ago. In the time since that ancestral line diverged, evolution by natural selection has produced an impressive diversity—from fish to birds to elephants—of vertebrate morphology; yet despite the great species-level differences that otherwise exist across the brains of many animals, the neural circuitry that underlies motor control features a functional architecture that is virtually unchanged in every living species of vertebrate. In this article, we review how that circuitry facilitates motor control, trial-and-error-based procedural learning, and habit formation; we then develop a model that describes how this circuitry (embodied in an agent) works to build and refine sequences of goal-directed actions that are molded to fit the structure of the environment (in which the agent is embedded). We subsequently review evidence suggesting that this same functional circuitry became further adapted to regulate cognitive control in humans as well as motor control; then, using examples of heuristic decision-making from the ecological rationality tradition, we show how the model can be used to understand how that circuitry operates analogously in both cognitive and motor domains. We conclude with a discussion of how the model encourages a shift in perspective regarding ecological rationality's "adaptive toolbox"—namely, to one that views heuristic processes and other forms of goal-directed cognition as likely being implemented by the same neural circuitry (and in the same fashion) as goal-directed action in the motor domain—and how this change of perspective can be useful.

KEYWORDS

ecological rationality, vertebrate motor control, cortico-basal ganglia-thalamo-cortical loop, habit formation, exaptation

Introduction

The field of ecological rationality (e.g., Todd et al., 2012) is predicated on the assumption that any answer to questions regarding the "rationality" of a given animal's behavior must necessarily include a proper accounting of (1) the evolved structure of the animal that exhibited the behavior, (2) the structure of the environment in which that behavior occurred, and (3) the structure of the environments in which the animal's ancestral species evolved (if structural differences between present and past environments are plausible). Although researchers in ecological rationality have often restricted their analyses to the structure of a decision-maker's mind (setting aside the mind's implicit dependence on the structure of the brain/body), work in this tradition ideally seeks to understand behavior and cognition from the broadest relevant vantage point—which naturally includes the embodied perspectives and embedded contexts of thinking and acting agents. Ecology is the study of interactions between organisms and their environments; indeed, *ecological* rationality is so named to specifically call out those interactions, and hence already implies an embodied and embedded perspective. Moreover, von Uexküll's (1957) inherently embodied and embedded concept of the Umwelt has been used explicitly within the ecological rationality community for years (e.g., Todd and Gigerenzer, 2012), and so the phrase "embodied and embedded ecological rationality" may admittedly seem redundant to some readers; however, we use it to draw attention to this connection, because others have criticized work in ecological rationality for overemphasizing environmental structure while underemphasizing the species-specific (and specificity-dependent) nature of decision-making environments (for further discussion/debate, see Felin et al., 2017; Chater et al., 2018; Felin and Koenderink, 2022).

This paper seeks to emphasize that the full extent and import of an agent's embodiment and embeddedness may be obscured by the lenses through which that agent's behavior and cognition are understood and described. For instance, although the heuristics and biases program (e.g., Tversky and Kahneman, 1974) takes inspiration from Simon's (1955) concept of bounded rationality, one of the fundamental insufficiencies of that program (from an ecological rationality perspective) is the failure to fully account for natural selection, leading to an impoverished understanding and description of behavioral data. In the heuristics and biases view, the predictable use and failure of specific heuristics in certain contexts is seen as evidence of human irrationality and presented as the conclusion of a cautionary tale. From an ecological rationality perspective, the same findings are instead a starting point for further inquiry—indicative of how underlying cognitive and behavioral mechanisms typically function, as well as providing insight into why those mechanisms evolved to operate that way.

The ecological rationality tradition has also not been immune to such perspectival limitations. Applying the same general criticism through the lens of a Marrian perspective (Marr, 1982), the ecological rationality literature has historically tended to restrict itself to computational and algorithmic levels of analysis (Gallese et al., 2020), focusing on the structure of environmental problems and the algorithmic tools used to solve them, while tending to eschew consideration of how those tools are constructed and implemented in terms of their underlying neuroanatomy and physiology. In fairness, many psychologists and cognitive scientists will openly admit that they ignore the brain in their thinking and research (at least as often as they can). Behavior and cognition can fruitfully be both studied and modeled, irrespective of whatever might be going on at the level of neurons, brain regions, and circuits, so why bother with the substrate? Given that this substrate happens to be the most complex object in the known universe, it may seem altogether appropriate to investigate higher-level cognitive and behavioral phenomena as a line of scientific inquiry that remains largely independent—if not completely divorced—from neuroscience. The intention here is not to accuse, but rather to acknowledge (1) that *all* viewpoints are limited, and (2) that neuro-agnostic cognitive scientists and psychologists may specifically benefit from a broader vantage that includes some degree of implementation-level understanding. To the point, researchers who are sympathetic to the ecological rationality approach should accept that a proper accounting of an animal's structural organization and limitations requires an appreciation of its embodied particulars, including the evolved neural architecture and perceptual apparatus that underlie behavior and cognition in the animal and its conspecifics (at least to the extent that it may be practically relevant).

As a specific example of the potential usefulness of this embodied/embedded ecological rationality perspective, this paper argues in favor of a greater implementation-level awareness of cortico-basal ganglia-thalamo-cortical—or CBGTC—circuitry, sometimes referred to as the CBGTC loop (e.g., Parent and Hazrati, 1995). Critical to the regulation of goal-directed action selection in vertebrate motor control, the architecture of CBGTC circuitry (or its functional equivalent in species that lack a neocortex) has been conserved throughout the evolution of every vertebrate species (Reiner, 2010; Stephenson-Jones et al., 2011); equivalently fundamental to motor control and motor learning in each of those species, this circuitry allows individuals within the vertebrate lineage to both learn basic sequences of goal-directed actions and to successfully achieve their goals by recalling and executing those sequences in situationally appropriate contexts (e.g., Grillner and Robertson, 2016). Furthermore, this same neural circuitry is implicated in cognition (e.g., Graybiel, 1997; Middleton and Strick, 2000), including the production and comprehension of human language (e.g., Lieberman, 2002; Reimers-Kipping et al., 2011). The evidence suggests that this evolutionarily

ancient circuitry evolved as an effective and efficient means for learning and regulating sequences of goal-directed motor behaviors, and that this functionality was extended over time *via* exaptation[1]—at least in human evolution—to serve analogously in cognitive control, providing us with the means to learn and regulate sequences of goal-directed cognitive operations. The extent of this functional overlap between motor and cognitive control makes these circuits an attractive starting point for an expanded implementation-level awareness among neuro-agnostic students of behavior and cognition.

We begin the rest of this paper with a summary overview of the recurrent structure of CBGTC circuitry, as well as its relevance to the regulation of action selection and the coordinated sequencing of goal-directed action (e.g., Park et al., 2020; Dhawale et al., 2021); we then review the role of this circuitry in procedural learning and the development of action sequence protocols and (in some repeating contexts) the transition away from voluntary execution of those protocols and toward their automatic expression in response to contextual triggers—i.e., habit formation (e.g., Graybiel, 1995). Following that basic overview, we outline a symbolic model of these implementation-level processes, which provides a general framework for understanding and describing behavioral phenomena in terms of an embodied agent, its goals, and the ecological contexts that emerge between goal-directed action (*via* perception and motor control) and environmental structure; this model also provides a common language that helps illustrate the relevance of CBGTC circuitry for cognition by highlighting the functional overlap between motor control and cognitive control. We then apply this framework to heuristic decision-making and the "adaptive toolbox" (e.g., Gigerenzer and Todd, 1999b) and discuss how our model may benefit current thinking and future work in ecological rationality and other areas of cognitive science.

A rough sketch of voluntary motor control and sequential goal-directed behavior

The neural circuitry of the CBGTC loop is complex, but it is not difficult to convey a simplified understanding of what the brain is doing during (and immediately prior to) voluntary action in the case of motor control. From the endpoint of the literal muscular activations that resulted in one of the authors typing on a keyboard, we can roughly trace the sequence of neural activation backward through the relevant circuitry to the initial intention to write the words you've just read (because all voluntary motor control invariably begins with a goal; for an accessible and less-physiologically-oriented overview of this process in greater detail, see Wong et al., 2015).

The motor cortex is ultimately responsible for the literal execution of voluntary movement *via* coordinated muscular activation; when a sentence is typed on a keyboard, it is because the appropriate somatotopic regions of the motor homunculus (M1) have been activated in order to move the muscles controlling the fingers just so, such that the goal of typing this or that word is ultimately achieved. This sequential, temporally coordinated pattern of activation is processed in premotor areas of the cortex (such as Broca's area), but the finalized sequence is ultimately forwarded to M1 only after the relevant cortical regions have been stimulated by excitatory subcortical projections from the thalamus; before the relevant thalamic neurons may excite those proper cortical pathways through to M1 in that way, the thalamus must first be selectively disinhibited by the subcortical nuclei of the basal ganglia[2]. One role of the basal ganglia is to serve as gatekeepers of behavioral expression—generally inhibiting thalamic activation while selectively opening the "gates" (*via* targeted selective cessation of that inhibition) to permit specific thalamic excitation to occur—coordinating which behaviors are ultimately expressed (and when). Prior to the basal ganglia releasing their inhibitory grasp on the particular thalamic neurons that will go on to excite motor areas of the cortex, the basal ganglia receive input from the prefrontal cortex (and elsewhere) regarding the motor goal, a motor plan that is predicted to achieve that goal, and sensory input associated with perception of the current context (i.e., sitting/staring at the computer, working to complete a draft of this document).

To summarize this progression in its proper order, (1) prior to typing, an intention in the cortex—e.g., to type the word *cortex*—forms the basis of a motor goal that leads to the selection of a planned motor sequence—e.g., to move particular fingers in series over the keyboard—which is predicted to achieve that goal; (2) this information is then projected subcortically to the basal ganglia, which (3) sequentially disinhibit select regions of the thalamus that (4) will correspondingly excite the cortex, leading to the behavioral execution of the motor plan. Of course, how fluidly this progression unfolds depends on one's prior experience/facility at typing. To type the same word, a student first learning to type may initially need to form distinct

1 Exaptation is the co-opting or repurposing of existing structure/functionality over the course of evolution by natural selection—a process by which pre-evolved structure/functionality is subsequently further adapted or co-opted, extending its use or operation to fit new modes or contexts for which it was not originally adapted (e.g., Gould, 1991; for relevant discussion on exaptation in the context of rationality, see Mastrogiorgio et al., 2022).

2 Technically, the relevant areas of the thalamus are always "attempting" to excite the cortex, but normally they are reined in by tonic inhibitory input from the basal ganglia, which persists until selective disinhibition allows targeted thalamic excitation to stimulate the cortex in a controlled fashion—this is why pathology of the basal ganglia can either lead to a chaotic excess of unintended movement (as in Huntington's chorea) when generalized inhibition falters, or deficits in voluntary motor control (as in Parkinson's disease) when selective disinhibition is impaired.

intentions, goals, and motor plans in order to press particular letter keys individually with specific fingers (and not others); however—over the natural course of procedural learning—the actions that achieve the lower-order goals of individually pressing the C-O-R-T-E-X keys may come to be sequenced together automatically when pursuing the single higher-order goal of typing the word *cortex*.

Procedural learning and habit formation in vertebrates

In general, if the execution of a motor plan in some context successfully achieves the motor goal that inspired that plan's initial selection, dopaminergic neurons provide reinforcement signals to the relevant sections of the CBGTC loop; this process of reinforcement forms associations that result in an increased likelihood of re-selecting that same motor plan in any future instance in which that same goal recurs within that same context (or similar contexts). When trial-and-error exploration is added, this combination of goal-directed motor control and reinforcement amounts to a basic description of procedural learning: Simpler behavioral elements that achieve lower-order goals are strung together (serially and/or in parallel) to form a more complex action sequence, which is executed in pursuit of a more complex higher-order goal (that the sequence is predicted to achieve); when a sequence of behavior achieves its goal, it is contextually reinforced in association with that goal and its concurrent/immediately preceding ecological features; the more frequently a given sequence achieves its goal and is reinforced in a consistent context, the deeper the association becomes between that goal, the sequence of behavior that achieved it, and other contextual features that consistently coincided with/preceded them—and the more consistently and efficiently that sequence is then selected and executed in the future when that constellation of reinforced associations subsequently realigns.

If the process of contextual and procedural reinforcement recurs consistently and frequently enough, the selection and coordinated execution of an action sequence may crystallize into a habit. In behavioral neuroscience, a habit describes a stereotyped sequence of goal-directed behavior that has become automatic[3] through "overlearning" (i.e., through consistent repetition within a stable context): over the course of many trials, individual behaviors of a sequence gradually fuse together into a singular "chunk" of behavior that becomes associated with—and triggered by—its context (e.g., Graybiel, 1998, 2005; Smith and Graybiel, 2014, 2016). In other words, the associations between goal, behavior, and coincidental contextual cues eventually become so strong (under the right conditions), that perceiving the associated cues will trigger the entire sequence of behavior through to its completion at the achievement of the goal. A study by Barnes et al. (2005) provides a window into the neurological development of a habit within the CBGTC loop. For this experiment, rats were repeatedly placed in a simple T-shaped maze; as a rat approached the T junction, a tone from the left or right reliably signaled which arm of the maze the rat could follow in order to find a chocolate pellet reward (which it was allowed to eat, if it chose the correct arm). Initially, single-unit recording within the rats' basal ganglia revealed a constant and chaotic pattern of activation that corresponded with the halting exploratory motion with which the rats first examined the maze; however, as the rats became accustomed to the structure of this task environment (over the course of many trials), the pattern of striatal activation changed as their behavior became more efficient and consistently successful: task-irrelevant neural activity dropped off drastically, and task-relevant firing clustered around the beginning and end of the task. After this period of overlearning, the rats entered an "extinction" phase of trials—in which the source of the tone no longer reliably indicated which arm of the T-maze contained chocolate—followed by a "reacquisition" phase that re-established the consistency between tone and reward; the rats' neural activity reverted to initial levels of chaotic activation during extinction trials, but rapidly resumed pre-extinction firing patterns after the onset of reacquisition trials (Barnes et al., 2005).

The pre-extinction shift in activation reflects the general nature of procedural learning and (later) habit formation: What was once a series of distinctly-exploratory actions, executed individually in pursuit of multiple disjointed goals (e.g., *check over there*; *try forward and to the left*; *now right*; *ooh, eat this chocolate!*), becomes consolidated into a unified "chunk" of behavior, executed collectively in response to a set of contextual triggers that has become associated with that behavioral chunk and its achievement of a single, overarching goal. What was once an unfamiliar context—in which exploration *occasionally* resulted in a chocolate reward—has become a recognized context in which adherence to a strict behavioral protocol *always* results in a reward. A habit naturally starts to form as any vertebrate animal (e.g., a rat) experientially discovers that a recurrent goal (receiving the chocolate pellet upon solving a maze) is repeatedly achieved *via* the execution of a stereotypical sequence of behavior (following a direct route to the maze's end, given a tone on one side) whenever it perceives that it has reencountered that context[4]. After a habit has become

3 Roughly in the dichotomous sense of automatic vs. controlled processing (Schneider and Shiffrin, 1977).

4 Evidence supports a kind of retrograde contextual expansion in the development of a habit. A habit's endpoint is naturally tied to the achievement of its associated goal, but the neurological markers of a habit's onset apparently may shift backwards in time (relative to achieving the goal) in a way that reflects an updating of when/where that habit's context begins (effectively enlarging the "chunk"). Barnes et al. (2005) report that these neurological-onset markers for their rat's habits were

established, the perception of its associated contextual cues automatically triggers the onset of that habit, which runs through to its completion (whereupon the goal is achieved).

The ecological context model: A formal account of embodied/embedded motor control

Generally, in the context of a desired goal in a particular environment, the process of procedural learning *via* trial-and-error exploration and reinforcement can be summarized as the construction (*via* motor control) of a novel action sequence that is discovered to be successful at achieving the desired goal (in that particular environment). In *recurrent* contexts—where a desired goal is repeatedly pursued in a particular type of environment that is stable enough to support the reuse of stereotyped behavior over the course of repeated encounters—the processes of procedural reinforcement (and habit formation) can be summarized as streamlining the selection of a sequence of actions that consistently achieves its goal in the associated environments (and the consolidation of that sequence into a singular behavioral chunk). Given this basic understanding, we can roughly characterize how vertebrates physically navigate their environments and pursue their goals, flexibly stringing simpler behaviors together into more complex sequences in an exploratory fashion, using trial-and-error learning to discern which sequences achieve their goals, and—in recurrent contexts—refining behavioral protocols and developing habits to efficiently and effectively exploit stable (i.e., predictable) environmental structure.

From here, we establish a symbolic description of what occurs in these phenomena, which might be considered a generalized extension of Lewin's (1936) field theory equation in which behavior B is expressed as a function of the interactions between a person, which we will generalize to an agent A and its environment E as such:

$$B = f(A, E).$$

While Lewin's equation importantly entails that an agent's behavior necessarily depends on the ecological interactions between that agent and its environment, the nature of the function and the particulars of A and E are unspecified (Todd and Gigerenzer, 2020).

Rather than starting with behavior, we begin with an ecological context C, which refers to the unique situational configuration that arises when an individual agent A is oriented toward a specific goal g within a particular local environment E, as follows:

$$C = \{A(g), E\},$$

where—for a given context—the environment E consists of a set of structural features, where

$$E = \{f_1, f_2, f_3, \ldots, f_n\},$$

and E contains an agent-dependent subset— $E(A)$—consisting of structural features that are hypothetically perceptible by the agent, depending on its perceptual apparatus[5]. The agent A possesses a given repertoire of possible behaviors Br—whether learned or innate—where

$$Br = \{b_1, b_2, b_3, \ldots, b_n\}.$$

The agent A also possesses a set of recent (including current[6]) perceptions P, following from phenomenal awareness/experience of some perceived subset of $E(A)$, where

$$P = \{p_1, p_2, p_3, \ldots, p_n\},$$

and A similarly possesses a set of related behavioral associations Ba—given g and P—where

$$Ba(g, P) = \{b_1p_1g,\ b_1p_2g,\ b_2p_1g, \ldots\}.$$

Based on Ba, the agent plans and executes a behavior or sequence of behaviors B, drawn from Br, where (for example)

$$B = [b_i,\ b_j, b_k, \ldots],$$

and B is predicted to achieve g; if that prediction is successful and g is achieved—g^+—following B, then B is reinforced, and

originally recorded around when experimental trials began as the maze door opened and the rats entered the maze; however, over the course of further trials, these markers began to occur earlier and earlier in time, with recorded activation eventually settling around when experimenters first placed the rats into the pre-trial antechamber (where they waited for a few moments before the maze door opened and trials "officially" began). This suggests that habits are constructed in reverse for cases in which the structural stability (i.e., invariability) of a recurrent context supports the use of a stereotyped behavioral protocol to achieve a goal, and that the protocol expands to match the temporal/structural invariability of its context.

5 For example, the presence and reflectance of ultraviolet light would always be considered structural features in a given environment, but they would not normally be perceptible features for humans (absent special tools) in the way that they would be for most birds.

6 This ecological context model is organized around a single goal and its pursuit, to keep it simpler, but there may be multiple competing goals that are simultaneously "vying" to be pursued in any given moment (or which might be pursued in tandem), and the particular goal that takes precedence may change from moment to moment. The model could be adjusted to better capture continuous-time dynamics by changing the singular goal to a set of goals with associated motivation levels that fluctuate in response to real-time perceptions and/or changes to internal and external environmental factors (e.g., a function of rising hunger over time would increase the motivation to seek food, or the sudden appearance/perception of a dangerous predator would cause a spike in the motivation to fight and/or flee, eventually or suddenly leading to a shift in goal orientation that would entail the formation of a new ecological context in the model), which would govern transitions between contexts.

its related associations in $Ba\left(g, P\right)$ are updated, such that B is then more likely to be executed in a similar context— C'—in the future, where g recurs in C' and there is overlap between the agent's perceived features in $P(C)$ and $P(C')$.

In keeping with the themes of embodiment and embeddedness, the agent is only nominally separate from the environment in this model out of convenience: E and $E(A)$ should be understood to contain features that are internal to the agent as well as external—e.g., E includes the agent's cognitive architecture, behavioral repertoire, circadian rhythm, etc., and the agent-specific perceptible subset of E, $E(A)$, includes the agent's memories, emotions, interoceptions, and any other internal characteristics or processes that might enter its phenomenal awareness.

If any element in the configuration of a given ecological context is altered, that new configuration necessarily entails a different context. Because the set of an agent's perceptions, P, includes perceptions in the present moment, this might occur (for example) if the agent is unexpectedly interrupted and reorients toward a new goal (e.g., upon the arrival of a potential mate); or this may occur if the agent returns to a familiar and unchanged environment with an expanded or reduced behavioral repertoire (e.g., subsequent to learning or a restricting injury, respectively); or if the environment has changed (even imperceptibly—e.g., a trap has been set and thoroughly hidden); etcetera. And although it may seem unwieldy to differentiate between ecological contexts, given even the slightest changes to their constituent elements, this practice highlights the primacy of goals and their associations: the cyclical recurrence of goals (e.g., the periodic importance of the goals to eat and drink, inspired by oscillations in hunger and thirst) connects contexts over time, allowing agents to discover and exploit structural invariance across those contexts (e.g., by repeatedly returning to the location of a reliable watering hole to drink); an extended description of an agent/environment system over time can be characterized as a succession of contexts, depending on the prevailing goal of the agent within a given moment.

To organize and summarize, an ecological context can be expressed as comprising an agent's orientation toward a specific goal in its present environment,

$$C \quad = \left\{A(g), E\right\}, \tag{1}$$

where an agent's behavior in a given context can be expressed as a function of its behavioral repertoire and its behavioral associations (given its present goal and recent/current phenomenal awareness/perception within that context),

$$B\left(C\right) \quad = f\left(Br, Ba\left(g, P\right)\right), \tag{2}$$

and where the achievement of a goal in a given context can be expressed as a function of an agent's behavior and the structure of its environment,

$$g^{+}\left(C\right) \quad = f\left(B, E\right). \tag{3}$$

This degree of formalism allows us to systematically characterize a wide range of observed behavioral phenomena in terms of their associated contexts and contextualized interactions.

Procedural learning and habit formation in the ecological context model

Within the framework of this model, we can describe the characteristic progression from trial-and-error-based exploration through procedural learning and habit formation as a transition through a series of ecological contexts—following (1)—in which g and the external structure of E are held constant. In early contexts, the agent's expressed behavior—following (2)—is exploratory and unpredictable, but as $Ba\left(g, P\right)$ is updated (*via* reinforcement), later contexts in the series become more and more autocorrelated as behavior under (2) converges upon a stereotyped protocol that consistently achieves g under (3), given the fixed external structure of E—i.e., after a point, the outcome of (3) becomes predictable for all subsequent contexts in which the relevant structural features[7] of E and the perceived features P are effectively stable. When $B\left(C\right)$ becomes "chunked" into a single behavior (as occurs in habit formation), it is considered to have been added to the agent's behavioral repertoire, such that

$$B\left(C\right) \quad = b_{n+1},$$

and where b_{n+1} has been appended to the set Br, and may subsequently be recruited (à la transfer of learning; e.g., Day and Goldstone, 2012) in new behavioral sequences—following (2)—potentially in pursuit of unrelated goals in different contexts (e.g., during future trial-and-error exploration).

This framework may similarly be used to illuminate how and why habits occasionally break down and result in error. When an individual who typically drives their own car habitually attempts to shift from PARK into DRIVE in an unfamiliar rental, this will often result in the individual grasping a fistful of air (instead of the shifter) if it happens to be located behind the steering wheel rather than its accustomed spot in

7 Relevant in the sense of being integral to the successful execution of the behavior—e.g., the color of two otherwise identical cars is irrelevant to driving behaviors, but relevant to trying to locate one in a parking lot.

the center console of the familiar car (or vice versa); this can readily be understood in terms of preparing to drive in the familiar context of the known car C, and preparing to drive in the different, but structurally similar context of the unfamiliar rental C'. In both contexts, the goal g (to shift from PARK into DRIVE) is the same, and overlap in perceived features across $P(C)$ and $P(C')$ is sufficient to trigger the habitually-chunked sequence of behavior B in both contexts, following (2); however, structural differences between $E(C)$ and $E(C')$ are such that the habit fails to achieve g^+ in C' where it is consistently successful in C [following (3)]. After the failure, g persists unachieved in C', which typically motivates visual exploration to update P'—i.e., to perceptually locate the shifter—followed by the formation and execution of an adjusted motor plan B' that is predicted to achieve g and which might (given consistent repetition and reinforcement) become a new habit in C' if the unfamiliar car is driven frequently enough over a sufficient period of time.

This descriptive model was designed primarily to provide a conceptual pivot point—to facilitate a shift in discussion from motor control in vertebrates, generally, to cognitive control in humans, specifically. Given its ubiquity in vertebrates, the functional architecture of CBGTC circuitry is extremely well-studied (e.g., Foster et al., 2021). To reiterate, the same functional circuitry in humans is found even in the relatively simple lamprey, which diverged from the rest of our ancestral vertebrate line ∼560 million years ago: The basal ganglia in lamprey brains perform the same role in action selection as they do in modern humans, inhibiting most behavior but selectively disinhibiting actions in sequence to achieve specific motor goals (Grillner and Robertson, 2016). This suggests that the basal ganglia evolved (at least in part) to facilitate action selection in a pre-vertebrate species, and they were so effective that they remain virtually unchanged among all vertebrate species over half a billion years later (Reiner, 2010). This evidence strongly suggests that all vertebrates use the same CBGTC circuitry (or its functional equivalent) to orchestrate the timing and sequencing of motor actions—selected from among a general repertoire of possible actions—in the pursuit of various motor goals (Stephenson-Jones et al., 2011). Moreover, evidence *also* suggests that humans use CBGTC circuitry to orchestrate the specific timing and sequencing of *cognitive* actions—also selected from among a general repertoire of possible operations—in the pursuit of various cognitive goals (e.g., Lieberman, 2007; Graybiel, 2008). The next section formulates some decision-making research in terms of the ecological context model framework, to highlight the significance of this circuitry for cognition, and to demonstrate the potential benefits of viewing decision making and other cognitive phenomena through the lens of this particular embodied and embedded perspective.

Heuristics and the adaptive toolbox in the ecological context model

As traditionally conceived in the ecological rationality literature, a heuristic is an algorithmic process that uses limited environmental information in order to make effective and efficient decisions, assuming that the structure of the task environment appropriately matches the heuristic (Gigerenzer and Todd, 1999a). For instance, the *elimination-by-aspects* heuristic is a choice-making algorithm that first compares the available options on the basis of a single cue: If one option outscores the rest on that criterion, that is the choice—but if no option outscores any others on the basis of that cue, the sample of options is potentially reduced, the next cue is selected and checked to see if it determines a unique choice, and the process repeats down the line of possible cues until a choice is made (Tversky, 1972). If you are in a new town and need to decide on a restaurant to visit for dinner, you could be an unagi fan and pick sushi as your first cue, but then find that 3 out of 14 nearby restaurants serve sushi, so you limit the field to those 3, use price as your next cue, check out their menus, and then make your choice based on which spot offers unagi for the best price. This heuristic works well across various contexts, making an efficient choice in environments where options differ on a range of attributes.

Within the formal system outlined previously, we can reframe such heuristic decision-making algorithms like this in such a way that they are rendered indistinguishable from the context-sensitive execution of refined motor sequences and habits as described above. Just as habits, a heuristic can be represented in terms of this framework as a sequence of goal-directed cognitive operations—a decision mechanism selected (from among others in an existing repertoire) because it is expected to achieve a specific goal in a given ecological context. In the case of *elimination-by-aspects*, we would predict that this heuristic would be likely to be selected for use in any context in which the goal g is to make a choice in a decision-making environment E and in which three assumptions are met: First, that its perceptible features $E(A)$ include multiple choice options with discernibly (or conceivably) differentiating attributes—following (1); second, that the perceived features of that context, P, overlap with perceived features in prior contexts in which expressed behavior—following (2)—was the *elimination-by-aspects* heuristic; and third—following (3)—that the use of this heuristic resulted often enough in the achievement of g when it was deployed in similar contexts in the past.

When reframed in this manner, heuristic-based errors may also be rendered formally indistinguishable from habit-based errors, such as in the above example of the habitual shifter-grasping error that sometimes occurs when driving an unfamiliar car. As indicated above, grabbing at the air above a rental car's center console can be formulated as an instance in which habitual behavior is erroneously triggered in an unfamiliar context because it shares a goal and has overlapping features with a familiar context (in which the habitual behavior has previously been effective); in this case, behavior that would have been successful in one context leads to failure in the other, because the environmental structure of the second context is incompatible with goal achievement, given that behavior, as per equation (3). This can be seen to parallel successes and failures in the case of heuristic decision making.

For example, consider use of the *recognition heuristic*: Roughly, when facing a decision between two options wherein one is recognized and the other is not, choose the recognized option. Goldstein and Gigerenzer (2002) showed that students tend to use the recognition heuristic when asked which of two cities is more populous; the recognition heuristic is often successful in the context of questions like this, because cities that are larger tend to be more famous—hence more often talked about and consequently more recognizable—than smaller cities (Todd, 2007). But consider how American students would likely fare (on average) if asked which city is more populous in two different contexts: (a) comparing Japan's two largest cities, Tokyo and Yokohama, and (b) comparing Yokohama and Nagasaki, which is toward the bottom of the top 50 most populous Japanese cities. In both contexts, the goal is the same, as are the perceptible features, so we would predict (absent explicit individual knowledge) the use of the recognition heuristic in each instance, following (2). Consequently, because Tokyo and Nagasaki are likely both highly recognized (relative to Yokohama), Tokyo would likely be (correctly) chosen in the first comparison, and Nagasaki would likely be (incorrectly) chosen in the second. From the perspective of the ecological context model framework laid out above, this occurs because the underlying structure of the first context supports the recognition heuristic's successful use, following (3), but the structure of the second context is incompatible with that heuristic (while being similar enough to compatible contexts in order to elicit it).

The points of similarity between habits and heuristics suggest the possibility that there is little difference between them (at least in terms of their formation and implementation, according to our model). If this is correct, it would imply that at least some—if not many or most—heuristics in the adaptive toolbox have likely been formed in the same way that procedural memories and habits form: individually, *via* trial-and-error-based procedural learning, by combining available operations in pursuit of a specific goal in the context of a particular environment. In the last section of this paper, we argue that this is likely the case—namely, that the evolutionarily conserved circuitry that underlies vertebrate motor control has been coopted to facilitate the use of cognitive control to pursue and achieve cognitive goals analogously to how motor goals are pursued and achieved *via* motor control.

Goal pursuit in motor and cognitive domains: Evidence for generalized implementation

In general, if a description of a cognitive phenomenon—like a heuristic—can be expressed in terms of a sequence of actions or operations that are executed in pursuit of an identifiable goal in some context(s), we suspect it is a reasonable first assumption that—at an implementation level—the phenomenon in question relies on CBGTC circuitry (or its functional equivalent, as is the case for any overt sequence of goal-directed motor behavior in every species of vertebrate). Some of the advantages that this approach may bring to the study of cognition can be appreciated in terms of its analogous success in previous research comparing internal and external search (e.g., Hills et al., 2008, 2015a; Todd and Hills, 2020). This work suggests that search behavior in both physical and cognitive domains likely relies on a shared set of underlying neural mechanisms, and that these mechanisms almost certainly first evolved to facilitate exploration through external space and subsequently (much later) were further adapted *via* exaptation to similarly regulate exploration throughout internal space as well. Whereas cognitive search is possibly unique to humans (and its observation is relatively obscured by our skulls), physical search is practically ubiquitous in the animal kingdom (and is relatively straightforward to observe); given their shared implementation and common evolutionary provenance, a well-developed understanding of the nature of physical search—which is relatively easy to attain—can serve as a natural source of valuable insight into the nature of cognitive search (Hills et al., 2015b; for general discussion, see Todd and Miller, 2018).

Akin to the ubiquity of search, much of what we and other animals *do*, both in terms of our behavior and our cognition—including exploration, communication, and decision-making—amounts to the pursuit of various types of goals within various types of environments. Ultimately, it appears that CBGTC circuitry allows for specific behaviors and/or cognitive operations to be pieced together (serially and in parallel) into goal-directed sequences, to recognize when specific sequences are rewarding in particular contexts (because they achieve their associated goals), and to consolidate or "chunk" sequences of rewarded actions into singular protocols (which themselves may then be recruited in other

contexts to create even larger sequences in pursuit of more complex goals, eventually contributing to the formation of even larger chunked protocols). Given this perspective of CBGTC circuitry as a kind of recursive[8] sequencing engine that constructs, executes, and evaluates the efficacy of goal-directed action patterns, it appears possible (if not likely) that its functional architecture is so highly conserved among vertebrates precisely (or at least in part) because of how successfully it regulates the embodied pursuit of goals and the learning of embedded goal-pursuit protocols that are custom-molded to fit and exploit structural regularity in the environment wherever possible.

To adopt this neurologically-grounded perspective of heuristics as goal-directed behavior/cognition is to explicitly connect ecological rationality to the common neural architecture that is responsible for orchestrating goal-oriented motor behavior in vertebrate brains. This has the immediate benefit of simplifying the *strategy selection problem* (e.g., Marewski and Link, 2014)—in which the mechanism for choosing a given heuristic or strategy in any given context is unspecified and difficult to implement artificially—as the answer to this problem reduces to the analog of the combined processes of trial-and-error-based procedural learning, context recognition, and goal-directed action selection in vertebrate motor control (which are relatively well-studied in non-human vertebrates). Additionally, this view emphasizes the primacy of specific goals in behavior as well as in cognition. There is often an implicit generalization and abstraction of goals when "rationality" is defined traditionally in terms of "human reasoning" whereby "to act optimally" or "to make an optimal decision" could effectively characterize the presumed goal in any given situation. In contrast, the ecological context model that we propose encourages a perspective of rationality that is relative to an embodied agent's pursuit of its own *specific* goal (or set of goals) within the environment in which it is embedded; in this view, rationality—with respect to some agent's behavior—must be conceptualized and evaluated in terms of the agent's goal(s) that gave rise to the behavior in an environment, and whether the behavior in question was successful—with respect to the agent's goal(s)—in that context.

Ultimately, the ecological context model is a conceptual framework that may inform a range of approaches to interdisciplinary scientific inquiry. Tying goal-directed cognition to the neurophysiology of goal-directed motor control constrains the possible ways in which goal-directed cognition may have emerged during the course of evolution. This suggests that evolutionary theorists may gain insight by developing a greater understanding of the normal functioning of CBGTC circuitry in non-human vertebrates (e.g., Desrochers et al., 2010), the cognitive and behavioral consequences of its malfunctions in related pathology (e.g., in cases of Huntington's, Parkinson's, and FOXP2 mutations), as well as the abnormal behavior and anatomy/physiology of CBGTC circuitry in non-human animals that have been reared with humanized[9] genes that affect the development of that circuitry (e.g., Schreiweis et al., 2014). Further, neuroanatomists and behavioral neuroscientists may uncover new insights by investigating structural and functional differences in CBGTC circuitry across humans, non-human primates, and other mammals. Moreover, as CBGTC circuitry is so functionally conserved in vertebrate motor control and motor learning, psychologists and cognitive scientists may themselves derive new insights from existing work in neuroanatomy and behavioral neuroscience relevant to CBGTC circuitry: Some questions about hypothetical mechanisms of human cognition may become simpler when plausibly grounded by comparisons to potentially-corollary mechanisms in vertebrate motor control/learning that are already relatively well-studied in non-human animals.

Author contributions

SN wrote the first draft of the manuscript. Both authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research was supported in part by the John Templeton Foundation grant, "What drives human cognitive evolution".

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.

8 Recursive in the sense used by Hauser et al. (2002), who characterize recursion as "the capacity to generate an infinite range of expressions from a finite set of elements" (p. 1,569).

9 For example, mice may be reared with the two amino acid substitutions that differ between the human FOXP2 gene and the mouse Foxp2 gene (Enard et al., 2009).

References

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., and Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437, 1158–1161. doi: 10.1038/nature04053

Chater, N., Felin, T., Funder, D. C., Gigerenzer, G., Koenderink, J. J., Krueger, J. I., et al. (2018). Mind, rationality, and cognition: An interdisciplinary debate. *Psychonom. Bull. Rev.* 25, 793–826. doi: 10.3758/s13423-017-1333-5

Day, S. B., and Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educ. Psychol.* 47, 153–176. doi: 10.1080/00461520.2012.696438

Desrochers, T. M., Jin, D. Z., Goodman, N. D., and Graybiel, A. M. (2010). Optimal habits can develop spontaneously through sensitivity to local cost. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20512–20517. doi: 10.1073/pnas.1013470107

Dhawale, A. K., Wolff, S. B., Ko, R., and Ölveczky, B. P. (2021). The basal ganglia control the detailed kinematics of learned motor skills. *Nat. Neurosci.* 24, 1256–1269. doi: 10.1038/s41593-021-00889-3

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., et al. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137, 961–971.

Felin, T., and Koenderink, J. (2022). A Generative view of rationality and growing awareness. *Front. Psychol.* 13:807261. doi: 10.3389/fpsyg.2022.807261

Felin, T., Koenderink, J., and Krueger, J. I. (2017). Rationality, perception, and the all-seeing eye. *Psychonom. Bull. Rev.* 24, 1040–1059. doi: 10.3758/s13423-016-1198-z

Foster, N. N., Barry, J., Korobkova, L., Garcia, L., Gao, L., Becerra, M., et al. (2021). The mouse cortico–basal ganglia–thalamic network. *Nature* 598, 188–194. doi: 10.1038/s41586-021-03993-3

Gallese, V., Mastrogiorgio, A., Petracca, E., and Viale, R. (2020). "Embodied bounded rationality," in *Routledge handbook of bounded rationality*, ed. R. Viale (Abingdon: Routledge), 377–390.

Gigerenzer, G., and Todd, P. M. (1999a). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.

Gigerenzer, G., and Todd, P. M. (1999b). "Fast and frugal heuristics: The adaptive toolbox," in *Simple heuristics that make us smart*, eds G. Gigerenzer, P. M. Todd, and The ABC Research Group (Oxford: Oxford University Press), 3–36.

Goldstein, D. G., and Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychol. Rev.* 109:75. doi: 10.1037/0033-295X.109.1.75

Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. *J. Soc. Issues* 47, 43–65. doi: 10.1111/j.1540-4560.1991.tb01822.x

Graybiel, A. M. (1995). Building action repertoires: Memory and learning functions of the basal ganglia. *Curr. Opin. Neurobiol.* 5, 733–741. doi: 10.1016/0959-4388(95)80100-6

Graybiel, A. M. (1997). The basal ganglia and cognitive pattern generators. *Schizophrenia Bull.* 23, 459–469. doi: 10.1093/schbul/23.3.459

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* 70, 119–136. doi: 10.1006/nlme.1998.3843

Graybiel, A. M. (2005). The basal ganglia: Learning new tricks and loving it. *Curr. Opin. Neurobiol.* 15, 638–644. doi: 10.1016/j.conb.2005.10.006

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31, 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Grillner, S., and Robertson, B. (2016). The basal ganglia over 500 million years. *Curr. Biol.* 26, R1088–R1100. doi: 10.1016/j.cub.2016.06.041

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Hills, T. T., Todd, P. M., and Goldstone, R. L. (2008). Search in external and internal spaces: Evidence for generalized cognitive search processes. *Psychol. Sci.* 19, 802–808. doi: 10.1111/j.1467-9280.2008.02160.x

Hills, T. T., Todd, P. M., and Jones, M. N. (2015a). Foraging in semantic fields: How we search through memory. *Topics Cogn. Sci.* 7, 513–534. doi: 10.1111/tops.12151

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., Cognitive Search, et al. (2015b). Exploration versus exploitation in space, mind, and society. *Trends Cogn. Sci.* 19, 46–54. doi: 10.1016/j.tics.2014.10.004

Lewin, K. (1936). *Principles of topological psychology*. New York, NY: McGraw-Hill. doi: 10.1037/10019-000

Lieberman, P. (2002). On the nature and evolution of the neural bases of human language. *Yearb. Phys. Anthropol.* 45, 36–62. doi: 10.1002/ajpa.10171

Lieberman, P. (2007). The evolution of human speech: Its anatomical and neural bases. *Curr. Anthropol.* 48, 39–66. doi: 10.1086/509092

Marewski, J. N., and Link, D. (2014). Strategy selection: An introduction to the modeling challenge. *Wiley Interdiscip. Rev. Cogn. Sci.* 5, 39–59. doi: 10.1002/wcs.1265

Marr, D. (1982). *Vision*. Missouri: Freeman.

Mastrogiorgio, A., Felin, T., Kauffman, S., and Mastrogiorgio, M. (2022). More thumbs than rules: Is rationality an exaptation? *Front. Psychol.* 13:805743. doi: 10.3389/fpsyg.2022.805743

Middleton, F. A., and Strick, P. L. (2000). Basal ganglia output and cognition: Evidence from anatomical, behavioral, and clinical studies. *Brain Cogn.* 42, 183–200. doi: 10.1006/brcg.1999.1099

Parent, A., and Hazrati, L. N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res. Rev.* 20, 91–127. doi: 10.1016/0165-0173(94)00007-C

Park, J., Coddington, L. T., and Dudman, J. T. (2020). Basal ganglia circuits for action specification. *Annu. Rev. Neurosci.* 43, 485–507. doi: 10.1146/annurev-neuro-070918-050452

Reimers-Kipping, S., Hevers, W., Pääbo, S., and Enard, W. (2011). Humanized Foxp2 specifically affects cortico-basal ganglia circuits. *Neuroscience* 175, 75–84. doi: 10.1016/j.neuroscience.2010.11.042

Reiner, A. (2010). "The conservative evolution of the vertebrate basal ganglia," in *Handbook of behavioral neuroscience*, Vol. 20, (Amsterdam: Elsevier), 29–62. doi: 10.1016/B978-0-12-374767-9.00002-0

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 84:1. doi: 10.1037/0033-295X.84.1.1

Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., et al. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14253–14258. doi: 10.1073/pnas.1414542111

Simon, H. A. (1955). A behavioral model of rational choice. *Q. J. Econ.* 69, 99–118. doi: 10.2307/1884852

Smith, K. S., and Graybiel, A. M. (2014). Investigating habits: Strategies, technologies and models. *Front. Behav. Neurosci.* 8:39. doi: 10.3389/fnbeh.2014.00039

Smith, K. S., and Graybiel, A. M. (2016). Habit formation. *Dialogues Clin. Neurosci.* 18:33. doi: 10.31887/DCNS.2016.18.1/ksmith

Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B., and Grillner, S. (2011). Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Curr. Biol.* 21, 1081–1091. doi: 10.1016/j.cub.2011.05.001

Todd, P. M. (2007). How much information do we need? *Eur. J. Operational Res.* 177, 1317–1332. doi: 10.1016/j.ejor.2005.04.005

Todd, P. M., and Gigerenzer, G. (2012). "What is ecological rationality?," in *Ecological rationality: Intelligence in the world*, eds P. M. Todd, G. Gigerenzer, and The ABC Research Group (Oxford: Oxford University Press), 3–30. doi: 10.1093/acprof:oso/9780195315448.003.0011

Todd, P. M., and Gigerenzer, G. (2020). "The ecological rationality of situations: Behavior = f(Adaptive Toolbox, Environment)," in *The Oxford handbook of psychological situations*, eds J. F. Rauthmann, R. A. Sherman, and D. C. Funder (Oxford: Oxford University Press), 143–158. doi: 10.1093/oxfordhb/9780190263348.013.29

Todd, P. M., and Hills, T. T. (2020). Foraging in mind. *Curr. Direct. Psychol. Sci.* 29, 309–315. doi: 10.1177/0963721420915861

Todd, P. M., and Miller, G. F. (2018). The evolutionary psychology of extraterrestrial intelligence: Are there universal adaptations in search, aversion, and signaling? *Biol. Theory* 13, 131–141. doi: 10.1007/s13752-017-0290-6

Todd, P. M., Gigerenzer, G., and The Abc Research Group. (2012). *Ecological rationality: Intelligence in the world*. Oxford: Oxford University Press.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychol. Rev.* 79:281. doi: 10.1037/h0032955

Tversky, A., and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science* 185, 1124–1131.

von Uexküll, J. (1957). "A stroll through the worlds of animals and men: A picture book of invisible worlds," in *Instinctive behavior: The development of a modern concept*, ed. C. H. Schiller (New York, NY: International Universities Press), 5–79.

Wong, A. L., Haith, A. M., and Krakauer, J. W. (2015). Motor planning. *Neuroscientist* 21, 385–398. doi: 10.1177/1073858414541484

frontiers | Frontiers in Psychology

TYPE Conceptual Analysis
PUBLISHED 18 May 2023
DOI 10.3389/fpsyg.2023.1152866

Check for updates

OPEN ACCESS

EDITED BY
Massimiliano Palmiero,
University of Teramo, Italy

REVIEWED BY
Nathalie Gontier,
University of Lisbon, Portugal
Firat Soylu,
University of Alabama, United States

*CORRESPONDENCE
Riccardo Viale
✉ viale.riccardo2@gmail.com

RECEIVED 28 January 2023
ACCEPTED 19 April 2023
PUBLISHED 18 May 2023

CITATION
Viale R, Gallagher S and Gallese V (2023) Bounded rationality, enactive problem solving, and the neuroscience of social interaction. *Front. Psychol.* 14:1152866. doi: 10.3389/fpsyg.2023.1152866

COPYRIGHT
© 2023 Viale, Gallagher and Gallese. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Bounded rationality, enactive problem solving, and the neuroscience of social interaction

Riccardo Viale[1,2]*, Shaun Gallagher[3,4] and Vittorio Gallese[5,6]

[1]Department of Economics and BIB-Ciseps, University of Milano-Bicocca, Milan, Italy, [2]Cognitive Insights Team, Herbert Simon Society, Turin, Italy, [3]Department of Philosophy, University of Memphis, Memphis, TN, United States, [4]SOLA, University of Wollongong, Wollongong, NSW, Australia, [5]Department of Medicine and Surgery, Unit of Neuroscience, University of Parma, Parma, Italy, [6]Italian Academy for Advanced Studies, Columbia University, New York, NY, United States

This article aims to show that there is an alternative way to explain human action with respect to the bottlenecks of the psychology of decision making. The empirical study of human behaviour from mid-20th century to date has mainly developed by looking at a normative model of decision making. In particular Subjective Expected Utility (SEU) decision making, which stems from the subjective expected utility theory of Savage (1954) that itself extended the analysis by Von Neumann and Morgenstern (1944). On this view, the cognitive psychology of decision making precisely reflects the conceptual structure of formal decision theory. This article shows that there is an alternative way to understand decision making by recovering Newell and Simon's account of problem solving, developed in the framework of bounded rationality, and inserting it into the more recent research program of embodied cognition. Herbert Simon emphasized the importance of problem solving and differentiated it from decision making, which he considered a phase downstream of the former. Moreover according to Simon the centre of gravity of the rationality of the action lies in the ability to adapt. And the centre of gravity of adaptation is not so much in the internal environment of the actor as in the pragmatic external environment. The behaviour adapts to external purposes and reveals those characteristics of the system that limit its adaptation. According to Simon (1981), in fact, environmental feedback is the most effective factor in modelling human actions in solving a problem. In addition, his notion of *problem space* signifies the possible situations to be searched in order to find that situation which corresponds to the solution. Using the language of embodied cognition, the notion of problem space is about the possible solutions that are enacted in relation to environmental affordances. The correspondence between action and the solution of a problem conceptually bypasses the analytic phase of the decision and limits the role of symbolic representation. In solving any problem, the search for the solution corresponds to acting in ways that involve recursive feedback processes leading up to the final action. From this point of view, the new term *enactive problem solving* summarizes this fusion between bounded and embodied cognition. That problem solving involves bounded cognition means that it is through the problem solver's enactive interaction with environmental affordances, and especially social affordances that it is possible to construct the processes required for arriving at a solution. Lastly the concept of *enactive problem solving* is also able to explain the mechanisms underlying the adaptive heuristics of rational ecology. Its adaptive function is effective both in practical and motor tasks as well as in abstract and symbolic ones.

KEYWORDS

bounded rationality, embodied cognition, problem solving, decision making, enaction

1. Introduction

We begin with a brief background history of Subjective Expected Utility (SEU) decision making. On this view, the cognitive psychology of decision making precisely reflects the conceptual structure of formal decision theory. In relation to this structure and the normative component derived from it, empirical research in the cognitive psychology of decision making has been developing since the 1950s. This article shows that there is an alternative to this view that recovers Newell and Simon's bounded rationality account of problem solving and integrates it into the recently developed research program of embodied cognition. The role of embodied cognition is fundamental in the pragmatic activity of problem solving. It is through the problem solver's enactive interaction with environmental affordances, and especially social affordances that it is possible to construct the processes required for arriving at a solution. In this respect, the concept of bounded rationality is reframed in terms of embodied cognition.

2. Bounded rationality is bounded by the decision making programme

The empirical study of human behaviour from the mid-20th century to date has mainly developed by looking at a normative model of decision making. In particular Subjective Expected Utility (SEU) decision making, which stems from the subjective expected utility theory of Savage (1954) that itself extended the analysis of Von Neumann and Morgenstern (1944).[1]

In decision theory, the von Neumann–Morgenstern utility theorem[2] shows that under certain axioms of rational behaviour, such as completeness and transitivity, a decision maker faced with risky (probabilistic) outcomes of different choices will behave as if he or she is maximizing the expected value of some function defined over the potential outcomes at some specified point in the future. The theory recommends which option rational individuals should choose in a complex situation, based on their risk appetite and preferences. The theory of subjective expected utility combines two concepts: first, a personal utility function, and second a personal probability distribution (usually based on Bayesian probability theory).[3]

The concepts used to define the decision are therefore information about the world; the risk related to outcomes and consequences; preferences over alternatives; the relative utilities on the consequences; and, finally, the computation to maximize the subjective expected utility. Even if in formal decision theory no explicit reference is made to the actual mental and psychological characteristics of the decision maker, in fact the concepts that define decision can be mapped onto psychological processes, such as the processing of external perceptual incoming inputs or internal mnemonic inputs, mental representations of the states of the world on the basis of information, hedonic evaluations[4] of the states of the world, and deductive and probabilistic computation on the possible decisions to be implemented on the basis of hedonic evaluations (Viale, 2023a).

On this view, the cognitive psychology of decision making precisely reflects the conceptual structure of formal decision theory. In relation to this structure and the normative component derived from it, empirical research in the cognitive psychology of decision making has been developing since the 1950s. Weiss and Shateau (2021), highlight that in the 1950s Edwards (1992), the founder of the psychology of decision making, began to carry out laboratory experiments to unravel the way in which people actually decide. His experiments, which became the reference of subsequent generations and in particular of Daniel Kahneman and Amos Tversky's Heuristics and Biases program, have two fundamental characteristics: firstly, the provisions of the SEU are set as a normative reference, and the experimental work has the aim of evaluating when and how the human decision maker deviates from the requirements of the SEU. Ultimately, the aim is to discover the irrational components in the decision which constitutes its bounded rationality.[5] Secondly, the experiments are not carried out in the real decision-making contexts of everyday life, but in abstract situations of games, gamblings, bets and lotteries. In these abstract experimental situations, characterized by risk, the informative characteristics typical of the real environment - such as uncertainty, complexity, poor definition of data, instability of phenomena, dynamic and interactive change with the decision maker, and so on - are entirely absent (Viale, 2023a,b).

This situation is highlighted by Lejarraga and Hertwig (2021). Psychological experimentation on decision making,[6] particularly within the Heuristics and Biases program, uses experiments that represent descriptions of statistical events on which a probabilistic judgment is asked. These are generally descriptions of games, bets and lotteries and other situations that do not correspond to the decision-making reality and the natural habitat of the individual and which, above all, exclude learning. The experiments in the Heuristics and

1 The way in which this escalation developed is discussed in detail in Mousavi and Tideman (2021).

2 Von Neumann and Morgenstern never intended axiomatic rationality to describe what humans and other animals do or what they should do. Rather, their intention was to prove that if an individual satisfies the set of axioms, then their choice can be represented by a utility function.

3 This theoretical model has been known for its clear and elegant structure and it is considered by some researchers to be one of "the most brilliant axiomatic theory of utility ever developed." In contrast, assuming the probability of an event, Savage defines it in terms of preferences over acts. Savage used the states (something that is not in your control) to calculate the probability of an event. On the other hand, he used utility and intrinsic preferences to predict the outcome of the event. Savage assumed that each act and state are enough to uniquely determine an outcome. However, this assumption breaks down in the cases where the individual does not have enough information about the event. In reality Savage explicitly limited the theory to small worlds, that is, situations in which the exhaustive and mutually exclusive set of future states S and their consequences C are known.

4 The hedonic approach to economic assessment can be used for evaluating the economic value of goods.The hedonic approach is based on the assumption that goods can be considered aggregates of different attributes, some of which, as they cannot be sold separately, do not have an individual price.

5 Bounded Rationality was introduced by Herbert Simon (1982) to characterize the constraints of human action. As it is represented in the scissor's metaphor there are two set of constraints: one is about the computational limitations of the mind and the other is about the complexity and uncertainty of the environment (task). The psychology of decision making and behavioural economics focussed mainly on the first cognitive set of constraints forgetting the second set.

6 The lack of ecological soundeness applies to many areas of cognitive psychology.

Biases program do not fulfill the Brunswik (1943, 1952, 1955, 1956) requirements for psychological experiments. Since the psychological processes are adapted in a Darwinian sense to the environments in which they function, then the stimuli should be sampled from the organism's natural ecology to be representative of the population of the stimuli to which the organism has adapted and to which the experimenter wishes to generalize. Therefore, an experiment should correspond to an experience and not to a description; it should be continuous and not discrete; and it should be ecological, normal and representative, and not abstract and unreal.

Furthermore, the highly artificial experimental protocols of the Heuristics and Biases program are frequently based on one-shot situations.[7] They do not correspond to how people learn and decide in a step-by-step manner, thus adapting to the demands of the environment. There is no room for people to observe, correct and craft their responses as experience accumulates. There is no space for feedback, repetition or opportunities to change. Consequently, conclusions about the irrationality of the human mind have been based on artificial experimental protocols (Viale, 2023a).

In summary, the psychology of decision making reflects the conceptual *a priori* structure of SEU theory. The formal concepts used to define decision making are mapped onto psychological processes involving perception, memory, mental representations of the states of the world, hedonic evaluations, and deductive and probabilistic computation on the possible decisions to be implemented on the basis of hedonic evaluations. The limits of this research tradition are evident in relation to bounded rationality (Viale, 2023a,b):

a) The provisions of the SEU are set as a normative reference, and the experimental work has the aim of evaluating when and how the human decision maker deviates from the requirements of the SEU. Ultimately, the aim is to discover the irrational performances in the decision.
b) Secondly, the experiments are not carried out in the real decision-making contexts of everyday life, but in an abstract one of games, bets and lotteries. In these abstract experimental situations, characterized by risk, the informative characteristics typical of the real environment - such as uncertainty, complexity, poor definition of data, instability of phenomena, dynamic and interactive change with the decision maker, and so on - are entirely absent. Accordingly, such experiments do not fulfil the Brunswik ecological requirements.

3. Problem solving as an alternative programme

When Herbert Simon introduced the arguments about the limits of rationality (Simon, 1947), he did so by referring to behaviour in public administration and industrial organizations. Unlike consumer behaviour whose rationality is evaluated in relation to the SEU theory, behaviour in organizations is evaluated above all at a routine or problem-solving level. The routines of the different hierarchical levels are the main way in which problems related to the processing of information complexity and uncertainty of the external environment are solved. But it is above all in solving new problems that Simon characterizes non-routine behaviour. Depending on successful problem solving in areas such as Research & Development, marketing, distribution, human resources, finance, etc. an organization may or may not survive. The problem-solving behaviours, that can subsequently become routines, express the adaptive capacity of an organization in a more or less competitive environment. The decision-making model linked to the SEU theory does not seem relevant to the organizational context and does not seem to be at the origin of the concept of Bounded Rationality (Viale, 2023a,b).

Simon (1978) emphasizes the importance of problem solving and differentiates it from decision making, which he considers a phase downstream of the former. In fact, Simon's research in AI, economic and organizational theory is almost entirely dedicated to problem solving that seems to absorb the evaluation and judgment phase (Viale, 2023c). In dealing with a task, humans have to frame problems, set goals and develop alternatives. Evaluations and judgments about the future effects of the choice are the optional final stages of the cognitive activity.[8] This is particularly true when the task is an ill-structured problem. When a problem is complex, it has ambiguous goals and shifting problem formulations; here cognitive success is characterized mainly by setting goals and designing actions. Simon offers the example of design-related problems:

> [T]he work of architects offers a good example of what is involved in solving ill-structured problems. An architect begins with some very general specifications of what is wanted by a client. The initial goals are modified and substantially elaborated as the architect proceeds with the task. Initial design ideas, recorded in drawings and diagrams, themselves suggest new criteria, new possibilities, and new requirements. Throughout the whole process of design, the emerging conception provides continual feedback that reminds the architect of additional considerations that need to be taken into account (Simon, 1986, p. 15).

Most of the problems in corporate strategy or governmental policy are as ill-structured as problems of architectural and engineering design or scientific activity. Reducing cognitive success to predictive ability (Schurz and Hertwig, 2019) seems to branch from the decision-making tradition and in particular from SEU theory. The latter deals solely with analytic judgements and choices, and it is not interested in how to frame problems, set goals and develop a suitable course of action (Viale, 2021,

7 This is not a characteristic merely of Heuristic & Biases experiments, but of the majority of lab experiments in psychology and economics with some exceptions in repeated games experiments as in ultimatum games with multiple players. Nevertheless the perseverance to use artificial experiments protocol relies on some methodological advantages as easy control of the crucial variables, random sampling and clear task conditions.

8 On the traditional models, problem solving includes the steps of judgement and evaluation, but does not include the stage of action. Problem solving and action, however, are both part of the phenomenon that we dub "enactive problem solving." It is a dynamic process based on pragmatic recursive attempts and related positive or negative feedback from the environment. Constructing the meaning of one's attempts at a solution and ultimately selecting the final solution are only possible through the problem solver's enacting interaction with environmental affordances (Viale, 2023a).

2023a,b). In the SEU approach empirical phenomena lose their epistemic and material identity and are symbolically deconstructed and manipulated as cues with only statistical meaning (tallied, weighted, sequenced and ordered) (Felin and Koenderink, 2022).

In contrast, cognitive success in most human activities is based precisely on the successful completion of the phases of problem-solving described by Simon. Problem-solving is not the computation of a decision based on an analytical prediction activity performed on data coming from deconstructed empirical phenomena, but rather a pragmatic recursive process made up of many attempts and related positive or negative feedback from the environment.

Simon's approach to problem solving highlights the influence of American pragmatism, and in particular of Dewey (1910), Peirce (1931), and James (1890), on his work. For the pragmatists, the centre of gravity of the rationality of action lies in the ability to adapt. And the centre of gravity of adaptation is not so much in the internal environment of the actor, that is, in his or her cognitive characteristics, as in the pragmatic external environment. Simon and Newell write: "For a system to be adaptive means that it is capable of grappling with whatever task environment confronts it. Hence, to the extent that a system is adaptive, its behaviour is determined by the demands of the task environment rather than by its own internal characteristics. Only when the environment stresses [the system's] capacities along some dimension - presses its performance to the limit - do we discover what those capabilities and limits are, and are we able to measure some of their parameters" (Newell and Simon, 1971, p. 149).

4. Enactive problem solving and 4E cognition

In this section we argue that the role of embodied cognition is fundamental in this pragmatic activity. We take embodied cognition in a broad sense to include what has been termed 4E (embodied, embedded, extended and enactive) cognition (Newen et al., 2018). On this view, the body's neural and extra-neural processes, as well its mode of coupling with the environment, and the environmental feedback that results, play important roles in cognition. Similar to Simon's approach, 4E cognition has philosophical roots in pragmatism (see especially Gallagher, 2017; Crippen and Schulkin, 2020), but also incorporates insights from phenomenology, analytic philosophy of mind, developmental and experimental psychology and the neurosciences.

Wilson (2002) outlined a set of principles embraced by most proponents of embodied or 4E cognition.

1. cognition is situated
2. cognition is time-pressured
3. we off-load cognitive work onto the environment
4. the environment is part of the cognitive system
5. cognition is for action
6. cognition (in both basic and higher-order forms) is based on embodied processes

Proponents of 4E approaches, however, vary in what they emphasize as explanatory for cognition. The body can play different roles in shaping cognition. Non-neural bodily processes are sometimes thought to shape sensory input prior to, and motor output subsequent to central or neural manipulations (e.g., Chiel and Beer, 1997). According to proponents of extended cognition minimal, action-oriented representations add further complexity (Clark, 1997a; Wheeler, 2005). Enactive approaches emphasize the idea that the body is dynamically coupled to the environment is important ways (Di Paolo, 2005; Thompson, 2007); they point not only to sensorimotor contingencies (where specific kinds of movement change perceptual input) (O'Regan and Noë, 2001), but also to bodily affectivity and emotion (Gallese, 2003; Stapleton, 2013; Colombetti, 2014) as playing a nonrepresentational role in cognition. Embedded and enactive approaches emphasize action affordances that are body- and skill-relative (Chemero, 2009). More generally, most theorists of embodied cognition hold that these ideas help to shift the ground away from orthodox, purely computational cognitive science, which clearly informs the cognitive psychology of decision making. In this respect, it's not just the internal processes of the mind or brain, but the brain–body-environment system that is the unit of explanation.

Relevant to the idea of problem solving, there is general agreement that the environment scaffolds our cognitive processes, and that our engagement with the environmental structure, and environmental features, including external props and devices, can shift cognitive load. Already, within the scope of Simon's own work it's clear that only through the enactive interaction between problem solver and environmental affordances is it possible to construct a solution. The metaphor of the ant on the beach (Simon, 1981) is illuminating: imagine an ant walking on a beach. Now let us say you wanted to understand why the ant is walking in the particular path that it is. In Simon's parable, you cannot understand the ant's behaviour just by looking at the ant: "Viewed as a geometric figure, the ant's path is irregular, complex, hard to describe. But its complexity is really a complexity in the surface of the beach, not a complexity in the ant" (Simon, 1981, p. 80). In other words, to predict the path of the ant, we have to consider the effects of the beach – the context that the ant is operating in. The message is clear: we cannot study what individuals want, need or value detached from the context of the environment that they are in. That environment shapes and influences their behaviour. In this example, the *procedural rationality* of the ant (finding a suitable behaviour on the beach) requires its *substantial rationality* (the adaptivity to the irregularity of the beach).

From this metaphor Simon derives a philosophical principle very much in tune with the broad sense of 4E cognition[9]: "A man

9 We note that although the concept of bounded rationality acknowledges the role of the environment in problem solving, it does this from an information processing perspective. In this respect bounded rationality is historically tied to a computational/cognitivist approach, rather than an embodied approach that emphasizes action-perception loops, affordances, and dynamic brain–body-environment assemblies. Some embedded and extended versions of embodied cognition can be viewed as consistent with the information processing/computational framework (e.g., Clark, 2008). Others, like the radical enactive approaches tend to reject this framework (e.g., Hutto and Myin, 2017). Our aim in this paper is not to resolve such debates in the embodied cognition literature. On our view, it remains an open question whether one can reframe bounded rationality in strict non-computational enactivist terms. In any case, Simon's pragmatist epistemology and his account of the importance of environmental feedback in solving problems draws him closer to the enactive aspects of embodied cognition. For a contrast between extended and enactive approaches in the context of institutional economics, see Clark (1997b) and Gallagher et al. (2019).

considered as a system capable of having a behaviour is very simple. The apparent complexity of his behaviour over time is largely a reflection of the complexity of the environment in which he finds himself" (Simon, 1981, p. 81). The behaviour adapts to external purposes and reveals those characteristics of the system that limit its adaptation.

> When agents coordinate their activity with environmental resources such as external artifacts, cognitive processes may be productively constrained or enabled by objective features, or enhanced by the affordances on offer. Examples include using written notes to reduce demands on working memory, setting a timer as a reminder to do something, using a map, or the surrounding landscape to assist in navigation, or, since the environment is not just physical, but also social, asking another person for directions (Gallagher, in press).

For the idea of enactive problem solving, however, it is important to emphasize two things. First, the relational nature of affordances. It is not just the environment that constrains behaviour; it is also the body's morphology and motor possibilities, and the agent's past experience and skill level that will define what counts as an affordance. The way in which the body couples (or can couple) to the environment, will delineate the set of possibilities or solutions available to the agent. Likewise, affordances can also be limited by an agent's affective processes, emotional states, and moods. It is sometimes not just what "I can" do (given my skill level and what the environment affords), but what "I feel like (or do not feel like)" doing (given my emotional state).

Second, as the pragmatists pointed out, the environment is not just the physical surroundings; it's also social and cultural and characterized by normative structures. As Gibson (1979) indicated, affordances can be social. Enactive problem solving also highlights the important role of social and intersubjective interactions (De Jaegher, 2018). Again, it's not only what "I can" do, but also what "I cannot" (or "I ought not") do given normative or institutional constraints, as well as cultural factors that have to do with, for example, gender and race. These are larger issues that range from understanding how dyadic interactions shape our developing skills, to how institutional factors can either enable or constrain our social interactions.

It is also the case that cultural practices can determine the way in which the environment is represented, thereby changing our ability to interact with it. Think of how much arithmetic was simplified by transitioning from Roman to Arabic numerals and to positional notation. The success of the Arabic number system was dictated by the positive pragmatic aspects it delivered in our ability to efficiently represent the world in quantitative terms.[10] In other words, it was the retroactive adaptation that allowed the Arabic number system to prevail. Embodied processes are primitive and original in the cultural development of mathematical calculus and geometry. In a set of well-known experiments, Goldin-Meadow et al. (2001) showed that hand gesture may add to or supplement mathematical thinking. Specifically, children perform better on math problems when they are allowed to use gestures. In addition, Lakoff and Nunez (2000, p. 28) argue that mathematical reasoning builds on innate abilities for "subitizing," i.e., discriminating, at a glance, between there being one, or two, or three objects in one's visual field, and on basic embodied processes involving "spatial relations, groupings, small quantities, motions, distribution of things in space, changes, bodily orientations, basic manipulations of objects (e.g., rotating and stretching), iterated actions, and so on." Thus, the concept of a set is derived from perception of a collection of objects in a spatial area; recursion builds upon repeated action; derivatives (in calculus) make use of concepts of motion, boundary, etc. (Lakoff and Nunez, 2000, pp. 28–29).[11] Likewise, Saunders Mac Lane (1981) provides "examples of advances in mathematics inspired by bodily and socially embedded practices: counting leading to arithmetic and number theory; measuring to calculus; shaping to geometry; architectural formation to symmetry; estimating to probability; moving to mechanics and dynamics; grouping to set theory and combinatorics" (Gallagher, 2017, p. 209). All such practices involve environmental feedback as an essential part of the process.

According to Simon (1981), in fact, environmental feedback is the most effective resource for modelling human actions in solving a problem. Design activity is shaped by the logic of complex feedback. A purpose is followed in the design, which is to solve a given problem (e.g., design a smooth urban plan for the regulation of road traffic), and when you think you have reached it, feedback is generated (e.g., from the political, social and geographical environment) that introduces a new, unforeseen purpose (e.g., energy saving constraints). This leads to reworking the design and generating new retroactive effects. The same selectivity in the solution of a problem is based on feedback from the environment (Simon, 1981, p. 218).

Newell and Simon (1971) propose the notion of the problem space. They write (p.150): a "problem space is about the possible situations to be searched in order to find that situation which corresponds to the solution." The concept of problem space can easily be characterized in terms of enactive interaction and coupling with environmental affordances. A problem space is equivalent to the possible solutions that can be enacted given the landscape of affordances (Rietveld and Kiverstein, 2014). Some of the resources that define a solution will come from past experience and one's skill set; some others from the consequences of the actions that have been attempted in pursuit of the solution. The actions leading to the solution manipulate the world in a recursive feedback process, whereas processes of forecasting, which often lead the problem solver into a dead end, have limited importance. In fact, for Simon (1981, p. 231) the distinction between "state description" that describes the world as it is and "process description" that characterizes the steps in manipulating the world to achieve the desired end is important. To use another Simonian figure: given a certain dish, the aim is to find

10 See, e.g., Overmann (2016, 2018). It is important to consider the role of materiality in defining physical affordances (found in paper and pencil, and the formation of doodles, images, and script), as well as physical practices with our hands that can lead to abstract modes of thought (Gallagher, 2017, p. 196n3; Overmann, 2017). Malafouris (2013, 2021) highlights how the fact that making straight lines was easier than making curved ones led to the development of more and more abstract forms in pictographs/ideographs. This promoted greater simplicity and speed of language production.

11 Lakoff and Nuñez frame their analysis in terms of metaphor. For views closer to enactive approaches, see Abrahamson (2021) and Gallagher and Lindgren (2015).

the corresponding recipe (Simon, 1981, p. 232). This research takes place through successive actions with phenomenological/sensory-motor feedback (taste, smell, texture) selectively directing us towards the final result. And, we may add, this happens not only when the problem is not well structured, as in the case in which we do not have the recipe data, but also when we know the necessary ingredients.

The correspondence between action and the solution of a problem conceptually bypasses the analytic phase of the decision and limits the role of symbolic representation. The decision-making model based on SEU theory does not correspond to the empirical reality of individual action. In solving any problem, whether opening a door, running to catch a falling ball, replacing a car tyre, calculating for a financial investment, solving tests and puzzles or negotiating with a competitor, the search for the solution corresponds to acting in the sense of *wide* and *strong* embodied cognition, including the idea of a recursive feedback process leading up to the final action. From this point of view, the concept of 'enactive problem solving' summarizes the integration of multiple factors and could well represent the complexity of the phenomenon (Viale, 2023a).

The importance of the embodied aspects of human cognition that emerge from the concept of enactive problem solving can also be demonstrated in the actions generated by the simple heuristics studied within the ecological rationality program (Gigerenzer, Todd, and ABC Group, 1999; Gallese et al., 2021). Ecological rationality represents the direct development of bounded rationality. Most ecological rationality heuristics have to do nominally with decision making, but in actuality are often enactive problem solving mechanisms, and they can be analysed in terms of embodied cognition. In support of this thesis, consider the main mental abilities that heuristics use in their activation. The core mental capacities exploited by the building blocks of simple heuristics include *recognition memory*, *frequency monitoring* and additionally, three typical embodied cognition capacities: visual *object tracking*, *emotion* and *imitation* (Hertwig and Herzog, 2009; Gigerenzer and Gassamaier, 2011; Hertwig and Hoffrage, 2011).

Gigerenzer (2022) writes that he "reserves the term 'embodied heuristics' for rules that require specific sensory and/or motor abilities to be executed, not for rules that merely simplify calculations" (Gigerenzer, 2022). In reality, the very capacity of frequency monitoring seems to reflect a dimension of embodiment. A confirmation of this comes from the considerations of Lejarraga and Hertwig (2021) on the importance of experimental protocols that include learning and experience. Why are the heuristics and biases experimental protocols in behavioural decision research that rely on described scenarios rather than learning and experience able to cause so many biases? Which qualities of experience make it different from description and thus potentially foster statistical intuitions? Lejarraga and Hertwig write: "A learner experiencing a sequence of events may, for instance, simultaneously receive sensory and motor feedback (potentially triggering affective or motivational processes); obtain temporal, structural, and sample size information" (Lejarraga and Hertwig, 2021, p. 557). In other words, the ability to respond correctly in repeated and experience-based statistical tests is derived from the adaptive role of the sensorimotor and affective feedback-loop associated with the task. Thus, enactive problem solving is also able to explain the mechanisms underlying the adaptive heuristics of rational ecology. Its adaptive function seems effective both in practical and motor tasks as well as in abstract and symbolic ones.

5. The inside story

In 4E approaches much of the emphasis falls on embodied and environmental processes. Perhaps this is a reaction to the overemphasis in classic computational cognitive science that emphasizes processes internal to the individual agent. 4E cognition, however, does not deny the important role of brain processes. Neural processes are dynamically coupled to non-neural bodily processes. Indeed, the explanatory model is brain–body-environment. So how should we characterize what is happening in the brain in this model, especially as it relates to affordance-related processes and social cognition and interaction?

In regard to the latter, we note that primates learn from others' behaviour and base their decisions also on the prediction of others' choices. The discovery of 'mirror neurons' in macaque monkeys (Gallese et al., 1996; Rizzolatti et al., 1996), and then of similar mechanisms in humans (see Gallese et al., 2004), revealed the cognitive role of the motor system in social cognition, enabling the start of social neuroscience. The solipsistic stance of classic cognitivism, addressing the 'problem of other minds' by means of a disembodied computational architecture applied to a social arena populated by other cognitive monads was finally challenged, giving way to an embodied account of intersubjectivity, grounded on what the phenomenologist, Merleau-Ponty (2012), called intercorporeity. Indeed, mirror neurons reveal a new empirically founded notion of intersubjectivity connoted first and foremost as the mutual resonance of intentionally meaningful sensorimotor behaviours. We believe that these empirical findings have important bearings on decision making and problem solving by revealing their intrinsic social and embodied quality.

Thirty years of empirical research on mirror neurons have shown that the perceptual functions of the human motor system may be linked with its evolutionarily retained relevance in planning and coordinating behavioural responses coherent with the observed action of others (for a recent review, see Bonini et al., 2022; see also Bonini et al., 2023). The picture, however, is more complex than originally thought. Recent studies employing chronically implanted multiple recording devices revealed that in macaques' lateral and mesial premotor areas, besides 'classic' mirror neurons there are neurons exclusively mapping the actions of others while lacking motor responses during action execution. Two recent studies are particularly relevant for issues pertaining to decision making and problem solving. Haroush and Williams (2015) used a joint-decision paradigm to study mutual decisions in macaques. The study revealed in the premotor dorsal region of the anterior cingulate cortex (dACC) neurons encoding the monkey's own decision to cooperate intermingled with neurons encoding the opponent monkey's decisions when they were yet unknown. The problem space, we might say, includes a reserved slot for the anticipated decisions and actions of the other agent. Another recent study by Grabenhorst et al. (2019) showed that macaques' amygdala neurons derive object values from conspecifics' behaviour observation (that is, from the other agents' observed actions towards a particular object) which the system then uses to anticipate a partner monkey's decision process. The present evidence suggests that other-related neuronal processing is co-activated with neurons encoding self-related processes in an extended network of brain areas encompassing multiple domains, from motor actions, sensations, and emotions to decisions and spatial representations, in multiple animal

species. As recently proposed by Bonini et al. (2022), when individuals witness the action of others, they face different options that are known to recruit the main nodes of the human mirror neurons network: 1) faithfully imitating or emulating the observed action, 2) avoiding doing so, or 3) executing a complementary or alternative action. Both the environmental context and the contemporary state of the observer (i.e., knowledge, motivation, emotion, skill-level etc.) profoundly shape the way in which an observed action affects his/her own motor system.

As Bonini et al. (2023) recently argued, "Although the concept of shared coding grounds the history of mirror neuron literature, our recent perspective emphasizes the role of agent-based coding as a means of linking sensory information about others (i.e., *via* other-type neurons) to one's own motor plans (i.e., self-type neurons). The inherently predictive nature of the motor and visceromotor systems, which hosts this neural machinery, enables the flexible preparation of responses to others depending on social and nonsocial contexts." Furthermore, pioneering studies capitalizing on hyperscanning techniques that go beyond the traditional "one-brain" approach, suggest that interbrain synchronies could guide social interaction by having self-related neurons in Subject 1 controlling behaviour and, in turn, causing the activity of other-selective neurons in the brain of Subject 2, processes which finally lead to an adaptive behavioural response by activating self-related neurons (Bonini et al., 2022).

Social neuroscience, therefore, shows us that the ability to understand others as intentional agents does not exclusively depend on propositional competence, but it is in the first place dependent on the relational nature of embodied behaviour. According to this hypothesis, it is possible to directly understand others' behaviour by means of the sensorimotor and visceromotor equivalence between what others do and what the observer can do. Thus, intercorporeity becomes the primordial source of knowledge that we have of others, informing interaction and providing an important source for evaluating problem spaces.

Empirical research has also demonstrated that the human brain is endowed with mirror mechanisms in the domain of emotions and sensations: the very same neural structures involved in the subjective experience of emotions and sensations are also active when such emotions and sensations are recognized in others. For example, witnessing someone expressing a given emotion (e.g., disgust, pain, etc.) or undergoing a given sensation (e.g., touch) recruits some of the viscero-motor (e.g., anterior insula) and sensorimotor (e.g., SII, ventral premotor cortex) brain areas activated when one experiences the same emotion or sensation, respectively. Other cortical regions, though, are exclusively recruited for one's own and not for others' emotions, or are activated for one's own tactile sensation, but are actually deactivated when observing someone else's being touched (for review, see Gallese, 2014; Gallese and Cuccio, 2015).

The recent research that we have cited thus suggests that our ability to interact with others in decision-making and problem-solving contexts is not exclusively or primarily the result of individual neurons that simply mirror others' behaviour, but is rather based on more complex neural networks that are constituted by a variety of cell types, distributed across multiple brain areas, coupled to the body, and attuned to selective aspects of the physical and social environment. Our own planning and problem solving involve behavioural responses that depend on the behaviours of others. To put it simply, it is not the brain *per se*, but the brain–body, by means of its interactions with the world of which it is part, that enacts our cognitive capacities. The proper development of this functional architecture of brain–body-environment scaffolds the more cognitively sophisticated social cognitive (including linguistic and conceptual) abilities that constitutes our rationality (Cuccio and Gallese, 2018; Gallese and Cuccio, 2018).

6. Conclusion

Our brief review of Subjective Expected Utility (SEU) decision making showed some of its limitations. Newell and Simon's approach to problem solving offers an alternative that reflects the concept of bounded cognition. We argued that this alternative fits well with some of the more recent research in embodied cognition. The role of embodied cognition and environmental feedback is fundamental in the pragmatic activity which we called enactive problem solving. This approach emphasizes bodily interaction with environmental affordances that form the problem space where solutions can be found. Explanations of such processes require an approach that emphasizes the enactive system of brain–body-environment. We highlighted the importance of specific brain processes (the mirror mechanisms) which contribute to this system in ways that facilitate complex social interactions. Only through the enactive interaction of the problem solver with environmental (including social and cultural) affordances is it possible to construct the complex solutions that characterize human design efforts.

A more detailed theory of enactive problem solving will depend to some extent on resolving some problems in the philosophy of mind and embodied cognition – basic issues that have to do with notions of information processing, computation, body-environment couplings, affordances, and how these may or may not involve representational processes of different kinds. In the meantime, linking the concepts of bounded rationality with embodied-enactive cognition should be taken as a pragmatic proposal (which itself would be an enactive problem solving approach) that could inform future experimental designs that may ultimately contribute to resolving the more theoretical problems.

Author contributions

RV: contribution about the critique to decision making and the proposal of enactive problem solving. SG: contribution about embodied cognition and enactivism. VG: contribution about embodied simulation and mirror neurons. All authors contributed to the article and approved the submitted version.

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

Abrahamson, D. (2021). Enactivist how? Rethinking metaphorizing as imaginary constraints projected on sensorimotor interaction dynamics. *Constr. Found.* 16, 275–278.

Bonini, L., Rotunno, C., Arcuri, E., and Gallese, V. (2022). Mirror neurons 30 years later: implications and applications. *Trends Cogn. Sci.* 26, 767–781. doi: 10.1016/j.tics.2022.06.003

Bonini, L., Rotunno, C., Arcuri, E., and Gallese, V. (2023). The mirror mechanism. Linking perception and social interaction. *Trends Cogn. Sci.* 27, 220–221. doi: 10.1016/j.tics.2022.12.010

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychol. Rev.* 50, 255–272. doi: 10.1037/h0060889

Brunswik, E. (1952). *The Conceptual Framework of Psychology*. Vol. *1*. University of Chicago Press. Chicago, IL.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.* 62, 193–217. doi: 10.1037/h0047470

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments* (*2nd*). University of California Press. Berkeley, CA.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Chiel, H., and Beer, R. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosci.* 20, 553–557. doi: 10.1016/S0166-2236(97)01149-1

Clark, A. (1997a). *Being There*. Cambridge, MA: MIT Press.

Clark, A. (1997b). "Economic reason: the interplay of individual learning and external structure" in *The Frontiers of the New Institutional Economics*. eds. J. Drobak and J. Nye (Cambridge, MA: Academic Press), 269–290.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press. Oxford.

Colombetti, G. (2014). *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA: MIT press.

Crippen, M., and Schulkin, J. (2020). *Mind Ecologies: Body, Brain, and World*. Columbia University Press. New York, NY.

Cuccio, V., and Gallese, V. (2018). A Peircean account of concepts: grounding abstraction in phylogeny through a comparative neuroscientific perspective. *Phil. Trans. R. Soc. B* 373:20170128. doi: 10.1098/rstb.2017.0128

De Jaegher, H. (2018). "The intersubjective turn" in *The Oxford Handbook of 4E Cognition*. eds. A. Newen, L. Bruin and S. Gallagher (Oxford: Oxford University Press), 453–468.

Dewey, J. (1910). *How We Think*. D C Heath. Lexington, MA.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenol. Cogn. Sci.* 4, 429–452. doi: 10.1007/s11097-005-9002-y

Edwards, W. (1992). *Utility Theories: Measurements and Applications*. Heidelberg: Springer

Felin, T., and Koenderink, J. (2022). A generative view of rationality and growing awareness. *Front. Psychol.* 13:807261. doi: 10.3389/fpsyg.2022.807261

Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind*. Oxford University Press, Oxford.

Gallagher, S. (in press) *Embodied and Enactive Approaches to Cognition*. Cambridge: Cambridge University Press.

Gallagher, S., and Lindgren, R. (2015). Enactive metaphors: learning through full-body engagement. *Educ. Psychol. Rev.* 27, 391–404. doi: 10.1007/s10648-015-9327-1

Gallagher, S., Mastrogiorgio, A., and Petracca, E. (2019). Economic reasoning in socially extended market institutions. *Front. Psychol.* 10:1856. doi: 10.3389/fpsyg.2019.01856

Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Phil. Trans. Royal Soc. London B* 358, 517–528. doi: 10.1098/rstb.2002.1234

Gallese, V. (2014). Bodily selves in relation: embodied simulation as second-person perspective on intersubjectivity. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 369:20130177. doi: 10.1098/rstb.2013.0177

Gallese, V., and Cuccio, V. (2015). "The paradigmatic body. Embodied simulation, intersubjectivity and the bodily self" in *Open MIND*. eds. T. Metzinger and J. M. Windt (Frankfurt: MIND Group), 1–23.

Gallese, V., and Cuccio, V. (2018). The neural exploitation hypothesis and its implications for an embodied approach to language and cognition: insights from the study of action verbs processing and motor disorders in Parkinson's disease. *Cortex* 100, 215–225. doi: 10.1016/j.cortex.2018.01.010

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi: 10.1093/brain/119.2.593

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci.* 8, 396–403. doi: 10.1016/j.tics.2004.07.002

Gallese, V., Mastrogiorgio, A., Petracca, E., and Viale, R. (2021). "Embodied bounded rationality" in *Routledge Handbook on Bounded Rationality*. ed. R. Viale (London: Routledge)

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Psychology Press, London.

Gigerenzer, G. (2022). Embodied heuristics. *Front. Psychol.* 12:711289. doi: 10.3389/fpsyg.2021.711289

Gigerenzer, G., and Gassamaier, W. (2011). Heuristic decision making. *Annu. Rev. Psychol.* 62, 451–482. doi: 10.1146/annurev-psych-120709-145346

Gigerenzer, G., and Todd, P. M.the ABC Research Group. *Simple Heuristics that make us Smart*. New York, NY: Oxford University Press (1999)

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395

Grabenhorst, F., Báez-Mendoza, R., Genest, W., Deco, G., and Schultz, W. (2019). Primate amygdala neurons simulate decision processes of social partners. *Cells* 177, 986–998.e15. doi: 10.1016/j.cell.2019.02.042

Haroush, K., and Williams, Z. M. (2015). Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cells* 160, 1233–1245. doi: 10.1016/j.cell.2015.01.045

Hertwig, R., and Herzog, M. S. (2009). Fast and frugal heuristics: tools of social rationality. *Soc. Cogn.* 27, 661–698. doi: 10.1521/soco.2009.27.5.661

Hertwig, R., and Hoffrage, U.ABC Research Group. (2011). *Social Heuristics that Make us Smart*. New York, NY: Oxford University Press.

Hutto, D. D., and Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. Cambridge, MA. MIT Press.

James, W. (1890). *The Principles of Psychology*, *2*, New York: Dover Publications, 1950.

Lakoff, G., and Núñez, R. (2000). *Where Mathematics Comes From*. New York: Basic Books.

Lejarraga, T., and Hertwig, R. (2021). How experimental methods shaped views on human competence and rationality. *Psychol. Bull.* 147, 535–564. doi: 10.1037/bul0000324

Mac Lane, S. (1981). Mathematical models: a sketch for the philosophy of mathematics. *Am. Math. Mon.* 88, 462–472. doi: 10.1080/00029890.1981.11995299

Malafouris, L. (2013). *How Things Shape the Mind: A Theory of Material Engagement*. Cambridge, MA, MIT Press.

Malafouris, L. (2021). How does thinking relate to tool making? *Adapt. Behav.* 29, 107–121. doi: 10.1177/1059712320950539

Merleau-Ponty, M. (2012) in *Phenomenology of Perception*. ed. D. A. Landes (London: Routledge)

Mousavi, S., and Tideman, N. (2021). "Beyond economists' armchair: the rise of procedural economics" in *Routledge Handbook of Bounded Rationality*. ed. R. Viale (London: Routledge)

Newell, A., and Simon, H. A. (1971). Human problem solving: the state of the theory in 1970. *Am. Psychol.* 26, 145–159. doi: 10.1037/h0030806

Newen, A., Bruin, L., and Gallagher, S. (Eds.). (2018). *The Oxford Handbook of 4E Cognition*. Oxford University Press. Oxford.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973. doi: 10.1017/S0140525X01000115

Overmann, K. A. (2016). The role of materiality in numerical cognition. *Quat. Int.* 405, 42–51. doi: 10.1016/j.quaint.2015.05.026

Overmann, K. A. (2017). Thinking materially: cognition as extended and enacted. *J. Cogn. Cult.* 17, 354–373. doi: 10.1163/15685373-12340012

Overmann, K. A. (2018). Constructing a concept of number. *J. Numer. Cogn.* 4, 464–493. doi: 10.5964/jnc.v4i2.161

Peirce, C. S. (1931) *Collected Papers of Charles Sanders Peirce: Science and Philosophy and Reviews, Correspondence, and Bibliography*. Cambridge, MA: Harvard University Press

Rietveld, E., and Kiverstein, J. (2014). A rich landscape of affordances. *Ecol. Psychol.* 26, 325–352. doi: 10.1080/10407413.2014.958035

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141. doi: 10.1016/0926-6410(95)00038-0

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Dover. *2nd*

Schurz, G., and Hertwig, R. (2019). Cognitive success: A consequentialist account of rationality in cognition. *Top. Cogn. Sci.* 11, 7–36. doi: 10.1111/tops.12410

Simon, H. A. (1947) *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York: Macmillan (1947)

Simon, H. A. (1978). "Information-processing theory of human problem solving," in *Handbook of learning & cognitive processes*. ed. W. K. Estes (London: Routledge).

Simon, H. (1981). *The Sciences of Artificial*. Cambridge, MA: The MIT Press

Simon, H. A. *Models of Bounded Rationality. Volume 1: Economic Analysis and Public Policy. Volume 2: Behavioural Economics and Business Organization*. (1982), Cambridge, MA: MIT Press.

Simon, H. A. (1986). Rationality in psychology and economics. *J. Bus.* 59, S209–S224. doi: 10.1086/296363

Stapleton, M. (2013). Steps to a "properly embodied" cognitive science. *Cogn. Syst. Res.* 22-23, 1–11. doi: 10.1016/j.cogsys.2012.05.001

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology and the Sciences of Mind*, Cambridge, MA: Harvard University Press.

Viale, R. (2021). "Psychopathological irrationality and bounded rationality. Why is autism economically rational?" in *Routledge Handbook on Bounded Rationality*. ed. R. Viale (London: Routledge)

Viale, R. (2023a). "Enactive problem solving: an alternative to the limits of decision making" in *Companion to Herbert Simon*. eds. G. Gigerenzer, R. Viale and S. Mousavi (Cheltenham: Elgar)

Viale, R. (2023b) Explaining social action by embodied cognition: from methodological cognitivism to embodied cognitive individualism. In N. Bulle and IorioF. Di (eds.) *Palgrave Handbook of Methodological Individualism*. London: Palgrave McMillan.

Viale, R. (2023c). "Artificial intelligence should meet natural stupidity. But it cannot!" in *Artificial Intelligence and Financial Behaviour*. eds. R. Viale, S. Mousavi, U. Filotto and B. Alemanni (Cheltenham: Elgar)

Von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ.

Weiss, D. J., and Shateau, J. (2021). The futility of decision making research. *Stud. Hist. Phil. Sci.* 90, 10–14. doi: 10.1016/j.shpsa.2021.08.018

Wheeler, M. (2005). *Reconstructing the Cognitive World*. Cambridge, MA: MIT Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322

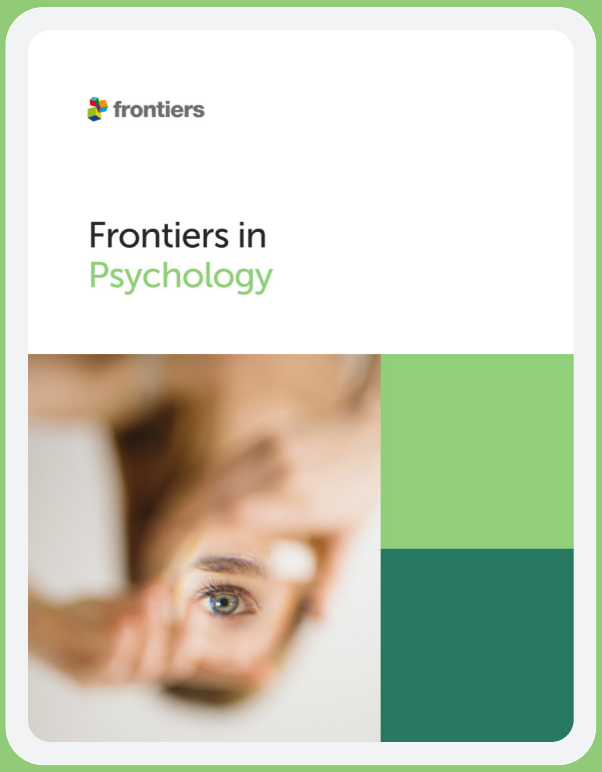
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact