ANALYTIC PHILOSOPHY AND 4E COGNITION

CONCEPTUAL ANALYSIS, EMBODIMENT AND SITUATEDNESS

Edited by
Manuel Heras-Escribano

ROUTLEDGE

Analytic Philosophy and 4E Cognition

This volume represents the first comprehensive collection of essays dedicated to exploring the conceptual and methodological intersections and tensions between analytic philosophy and the embodied–embedded approach to cognitive science, commonly referred to as "e-cognition."

Following an introductory chapter by the editor, which situates the discussion within its broader philosophical landscape, the contributors address a range of themes that traverse both analytic philosophy and 4E-cognition. These include skillful coping, habit formation, the nature and status of representations, consciousness, communication, and the social and political implications of embodied and situated approaches. The volume also examines how various theoretical traditions—such as ecological psychology, teleosemantics, enactivism, the Pittsburgh School, and intentional realism—engage with and apply these ideas.

Analytic Philosophy and 4E Cognition: Conceptual Analysis, Embodiment, and Situatedness will appeal to advanced students and scholars in analytic philosophy, philosophy of mind, and philosophy of psychology, as well as those working in cognitive science with an interest in embodied and situated cognition.

Manuel Heras-Escribano is Profesor Titular at the University of Granada, Spain.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Analytic Philosophy and 4E Cognition

Conceptual Analysis, Embodiment and Situatedness

Edited by
Manuel Heras-Escribano

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2026
by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2026 selection and editorial matter, Manuel Heras-Escribano; individual chapters, the contributors

The right of Manuel Heras-Escribano to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

The Open Access version of this book, available at www.taylorfrancis.com, has been made available under a Creative Commons Attribution (CC BY) 4.0 International license.

Any third party material in this book is not included in the OA Creative Commons license, unless indicated otherwise in a credit line to the material. Please direct any permissions enquiries to the original rightsholder.

For Product Safety Concerns and Information please contact our EU representative GPSR@taylorandfrancis.com. Taylor & Francis Verlag GmbH, Kaufingerstraße 24, 80331 München, Germany.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-041-09289-6 (hbk)
ISBN: 978-1-041-09290-2 (pbk)
ISBN: 978-1-003-64935-9 (ebk)

DOI: 10.4324/9781003649359

The research tasks for this volume have been generously supported by the Grant CNS2022-136195 ("Toward an Ecological Approach to the Natural Origins of Content: From Direct Perception to Social Norms (ECOCONTENT)") funded by MICIU/AEI/10.13039/501100011033 and by the European Union Next Generation EU/PRTR. It has also been supported by the Ramón y Cajal Fellowship, Grant RYC2022-036688-I funded by MICIU/AEI /10.13039/501100011033 and by ESF+.

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
AGENCIA ESTATAL DE INVESTIGACIÓN

Typeset in Sabon
by codeMantra

Contents

List of contributors *vii*
Preface *ix*

Introduction 1

1 Analytic Philosophy and 4E Cognition: friends or foes? 3
MANUEL HERAS-ESCRIBANO

PART 1
The analytic tradition through the 4E perspective 11

2 Capturing the ordinary 13
ALVA NOË

3 Ecological psychology as critical direct realism 25
CARL B. SACHS

4 Ecological psychology and the mirror of nature 43
VICENTE RAJA AND ANTHONY CHEMERO

5 Ecological psychology and the behaviorist bogeyman 54
MIGUEL SEGUNDO-ORTIN AND INÉS ABALO-RODRÍGUEZ

6 Being a good gadfly: radical enactivism's positively revisionary approach to cognitive science 77
DANIEL D. HUTTO

7 **Neo-Pragmatism and the natural origins of content** 99
DANIEL MARTÍNEZ MORENO AND MANUEL HERAS-ESCRIBANO

8 **Eliminativism is at the heart of E-cognition** 113
LUIS H. FAVELA

PART 2
From skills and habits to language and communication 127

9 **Skillful coping in the metaverse: on the challenges of immersion** 129
MARTA PÉREZ-VERDUGO

10 **The organismic turn. Teleosemantics after 4E** 143
XABIER E. BARANDIARAN AND TIAGO RAMA

11 **Communication beyond Inferentialism and Individualism** 160
GLENDA SATNE

12 **Linguistic relativity and embodiment** 181
FERNANDO MARTÍNEZ-MANRIQUE

PART 3
Ethical and political implications 195

13 **Situated agency: a Wittgensteinian exploration** 197
ANNEMARIE KALIS

14 **Recognition and the grounding of normativity** 211
LAURA MOJICA

15 **Embodied, embedded, enactive, extended… and exclusionary? Toward an inclusive E-Cognition for cognitive diversity** 226
VÍCTOR FERNÁNDEZ-CASTRO AND MIGUEL NÚÑEZ DE PRADO-GORDILLO

16 **Adaptive preferences and extended cognition** 246
GLORIA ANDRADA

Index *263*

Contributors

Inés Abalo-Rodríguez is Profesora Ayudante Doctora at Universidad Francisco de Vitoria, Spain

Gloria Andrada is Ramón y Cajal Research Fellow at the Instituto de Filosofía, CSIC, Spain

Xabier E. Barandiaran is Profesor Agregado at the University of the Basque Country-Euskal Herriko Unibertsitatea, Spain

Anthony Chemero is Distinguished Research Professor at the University of Cincinnati, United States

Luis H. Favela is Associate Professor, Indiana University Bloomington, United States

Víctor Fernández-Castro is Profesor Permanente Laboral at the University of Granada, Spain

Daniel D. Hutto is Senior Professor of Philosophical Psychology, University of Wollongong, Australia

Annemarie Kalis is Professor in Theoretical Philosophy at Utrecht University, the Netherlands

Fernando Martínez-Manrique is Profesor Titular de Universidad at the University of Granada, Spain

Daniel Martínez Moreno is FPU Researcher at the University of Granada, Spain

Laura Mojica is Juan de la Cierva-Formación Research Fellow at the University of the Basque Country-Euskal Herriko Unibertsitatea, Spain

Alva Noë is Professor of Philosophy at the University of California, Berkeley, United States

Miguel Núñez de Prado-Gordillo is Juan de la Cierva-Formación Research Fellow at the University of Granada, Spain

Marta Pérez-Verdugo is FPI Researcher at the University of the Basque Country-Euskal Herriko Unibertsitatea, Spain

Vicente Raja is Ramón y Cajal Research Fellow at the University of Murcia, Spain

Tiago Rama is Researcher at the University of the Republic, Uruguay

Carl B. Sachs is Professor of Philosophy at Marymount University, United States

Glenda Satne is Associate Professor at the University of Wollongong, Australia

Miguel Segundo-Ortin is Ramón y Cajal Research Fellow at the University of Murcia, Spain

Preface

The contributions gathered in this volume stem from the international workshop Analytic Philosophy and e-Cognition, held at the Carmen de la Victoria in Granada, Spain, in June 2024. Several of these essays originated as presentations delivered during the event, while others emerged as thoughtful elaborations inspired by the intellectually stimulating and collegial exchanges that unfolded over the course of four enriching days.

I would like to extend my deepest gratitude to all participants, whose engagement and generosity of spirit helped cultivate an atmosphere of genuine camaraderie and open dialogue. I am equally indebted to my fellow organizers—Lorena Lobo, Víctor Fernández, and María Rueda—for their indispensable collaboration and dedication in making the event possible. Finally, I wish to express sincere appreciation to the University of Granada and to the staff of Carmen de la Victoria for providing an incomparable setting in which to reflect, discuss, and share ideas.

The research tasks for this volume have been generously supported by the Grant CNS2022-136195 ("Toward an Ecological Approach to the Natural Origins of Content: From Direct Perception to Social Norms (ECOCONTENT)") funded by MICIU/AEI/10.13039/501100011033 and by the European Union Next Generation EU/PRTR. It has also been supported by the Ramón y Cajal Fellowship, Grant RYC2022-036688-I funded by MICIU/AEI /10.13039/501100011033 and by ESF+.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Introduction

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1 Analytic Philosophy and 4E Cognition

Friends or foes?

Manuel Heras-Escribano

1.1 Entering the E approaches to cognition

The 1990s marked a significant turning point in the philosophy of mind and cognitive science with the emergence of the so-called 4E approaches to cognition: embodied, embedded, enacted, and extended. These frameworks collectively rejected the traditional view that cognition is confined to the brain and instead proposed that cognitive processes are distributed across the brain, body, and environment. Notably, some of the most influential 4E contributions came from within analytic philosophy itself. Figures such as Susan Hurley, Andy Clark, and David Chalmers engaged with these radical reconceptualizations of mind while still operating within the analytic tradition, bringing with them its methodological rigor and conceptual clarity. Their work helped redefine long-standing debates on intentionality, mental content, and consciousness, thanks to proposals such as the extended mind hypothesis and the inclusion of sensorimotor contingencies in the picture.

Among these contributions, Alva Noë's *Action in Perception* stands out as a pivotal text. Published in 2004, it argued against the idea that perceptual experience is generated solely by internal representations in the brain. Instead, drawing on the ecological psychology of James and Eleanor Gibson and the sensorimotor theory of perception developed by enactivist approaches, Noë proposed that perception is an activity performed by the whole organism in interaction with its environment. This enactive view implies that the mind is not something hidden inside the skull but is actively constituted in the engagement with the world. He was one of the first authors to relate these embodied and situated approaches to analytic debates and ideas such as the notion of content or reference.

However, despite these overlaps and shared insights, a systematic integration of analytic philosophy and the 4E paradigm has been slow to materialize. This is surprising, given the resources analytic philosophy can offer to this project and the philosophical affinities it shares with 4E cognition when seen through the right interpretive lens. This chapter aims to propose some possible directions towards bridging that gap by examining the historical and conceptual intersections between analytic philosophy and 4E cognition,

DOI: 10.4324/9781003649359-2
This chapter has been made available under a CC BY 4.0 license.

showing how key post-analytic developments resonate deeply with embodied and situated accounts of mind.

1.2 Logicism, representationalism, and the Myth of the Analytic-Cognitivist Alliance

A common assumption is that analytic philosophy, with its roots in Fregean logic and the early philosophy of language, naturally aligns with cognitivism, i.e., the view that mental processes are defined as computational manipulations of symbolic representations.[1] In contrast, embodiment and situatedness found explicit support in traditions such as phenomenology or pragmatism,[2] while analytic philosophy implicitly supported cognitivism (the general framework of the mind that gave rise to the cognitive sciences); but this unsophisticated, broad-brush distinction just served as a revival in other areas for the analytic-continental divide that was still alive in the field of philosophy at the ending of the 20th century. This narrative, however, oversimplifies the evolution of analytic philosophy and the antecedents of embodied and situated cognition. While it is true that early analytic thinkers were concerned with formal languages and abstract semantics, the movement soon diversified, giving rise to what has been called by some authors as post-analytic philosophy.[3]

1 It is easy to see why many foundational works in the philosophy of cognitive science—particularly those written from a cognitivist perspective, such as those by Jerry Fodor and Paul Thagard—are often associated with analytic philosophy. This alignment stems from their shared emphasis on formal logic, a feature they also inherit from Chomskyan linguistics. Frege's contributions to logic and the semantics of language, when interpreted through a logicist lens, helped lay the groundwork for understanding cognition as symbolic computation—a view central to classical cognitive science. In this framework, cognitive processes are conceived as the manipulation of internal representations via computational operations on symbolic structures. Fodor's modularity of mind hypothesis exemplifies this tradition, building on the tight link between logic and language by positing the existence of an innate "language of thought".

This conceptual foundation also informed David Marr's influential three-level framework for analyzing cognitive systems: the computational level (specifying what the system does and why), the algorithmic level (describing the representations and processes it uses), and the implementational level (detailing how these processes are physically realized). Traditional cognitive science, grounded in this analytic-logical lineage, thus characterizes mental activity primarily as internal information processing—computation over symbolic representations carried out within the brain.

2 Works such as Varela, Thompson, and Rosch's *The embodied mind* explicitly mentioned Merleau-Pontyan phenomenology as a precursor of his idea of embodiment, but also Hubert L. Dreyfus' early critique of AI and cognition as computation finds support in this author and in Heidegger's hermeneutic phenomenology. On the other side, ecological psychology has been proposed by its founders and experts as a development of pragmatism (in particular, James' version).

3 For an interesting case that combines representational cognitive science and Heideggerian and Wittgensteinian nonrepresentational ideas on language, check Raja and Chemero's chapter in this volume.

Post-analytic thinkers such as Ludwig Wittgenstein, Gilbert Ryle, and later Wilfrid Sellars and Donald Davidson emphasized the primacy of use, normativity, and social practices over formal logico-semantic explanations that appealed to representational content in order to explain linguistic meaning and mind. Linguistic meaning, they argued, does not arise from internal mental symbols but from our public and interactive engagements with the world (something that undoubtedly resonates with embodiment and situatedness). This perspective reoriented the analytic tradition away from a narrow focus on logical reduction and internal representations toward a broader view of cognition as materially located and action-oriented.

Take, for example, the case of Gilbert Ryle. Beyond his well-known idea of the ghost in the machine, he was a vocal critic of the idea that mental states are inner causes of behavior or that they consist of hidden representations. His last letter to Daniel Dennett, written in 1976, offers a striking critique of the emerging computationalist and representationalist paradigm of the mind championed by figures like Jerry Fodor. Ryle writes:

> For reasons that I've forgotten, I'm anti-Fodor. But your review leaves me wondering 1) what on earth these 'representations' are supposed to be and do. Do I have them? Do I need them? Is their extension identical with that of Locke's less pompous 'ideas'? 2) What does 'internal' mean? Locke's usual 'inner'? If I run through the Greek alphabet a) in a sing-song; b) muttered; c) under my breath; d) merely 'in my head', is only d) properly 'internal'? So when I mutter or intone 'kappa' audibly is this noise not a 'representation' of an item in the Greek alphabet? (On p13 [of the typescript] we hear about 'representations of rules'. Sort of snapshots or echoes? Pinkish ones, or gruff ones?) Or if after dictating again and again a rule of grammar or chess, etc, the rule-wording goes running through my head by rote (like a maddening popular song), is that wording (or any word in it) a 'representation' of the rule–or of any part of it (if rules have parts)? From your review it seems that Fodor beats Locke in the intricacy of his 'wires-and-pulleys', when what was chiefly wrong with Locke was the (intermittent) intricacy of his 'wires-and-pulleys'! (…) In brief, I'm not persuaded that Fodor's book is about anything. It certainly seems not to be about (what interests me) thinking (= pondering, trying to get somewhere, being perplexed, baffled, stimulated, etc.) Such thinking is precisely not giving oneself 'information.' It's what one does, often in vain, when one is, e.g., without the wanted information. It's hunting, not swallowing; it fails or sometimes succeeds. Cognitive psychology sounds to me like the later days of phlogiston-theory! It looks as if F[odor] (or? D[aniel]D[ennett]!) take unexamined some bogus notion of 'internal' and then excogitate hypotheses about the ways in which postulated things, happenings, etc in this 'internal'

> region can go proxy for things, happenings etc not in this internal region–at least not in my private one.[4]

This extended passage cuts to the heart of the 4E critique of classical cognitive science.[5] Gilbert Ryle's critical remarks on Jerry Fodor's representationalist framework anticipate and substantively align with the core commitments of 4E cognitive science—namely, the embodied, embedded, enacted, and extended approaches to cognition. First of all, Ryle's probing skepticism toward the ontological and explanatory status of internal representations—framed through a series of rhetorical questions that expose their conceptual ambiguity—prefigures one of the principal challenges mounted by 4E theorists against classical cognitive science. In fact, here is no widespread assumption of what representations are in scientific terms, hence current efforts in the specialized literature to still find the holy grail of the cognitivist explanation of cognition as the formation and manipulation of representations.[6] Second, Ryle's insistence on understanding thinking not as the passive reception or internal manipulation of informational content, but as an active, often effortful, contextually situated process—"hunting, not swallowing"—resonates deeply with enactivist and ecological models of cognition, which emphasize sensorimotor engagement, environmental attunement, and the primacy of skilled action as key. This is why in embodied and situated approaches, cognition is not understood as internal computation but as adaptive behavior of the organism as a whole.[7]

Moreover, Ryle's interrogation of the internal/external distinction echoes 4E critiques of the Cartesian heritage that treats the mind as an isolated, inner realm. His refusal to treat thought as an inert, symbol-driven process enclosed within the skull aligns with the ecological and enactive views that cognition is fundamentally constituted by the dynamic interplay between an organism and its environment. In questioning the very coherence of positing internal representations as explanatory primitives, Ryle anticipates

4 Retrieved from: https://ejap.louisiana.edu/ejap/2002/RyleLett.pdf.

5 Ideas such as the centrality of agency and the first-person perspective, the importance of organismal explanations, the idea that thinking is a doing rather than a having of thoughts… All these ideas are precursors of embodied and situated approaches.

6 Although someone might find controversial the claim that there is no widely shared notion of representation in psychology, neuroscience, or cognitive science, a brief look at contemporary publications might persuade the reader that this is the case. Take, for example, Nick Shea's book *Representation in Cognitive Science* published in 2018 or Schmortchkova, Dolega, and Schlicht's volume entitled *What are mental representations?* published in 2020 as recent examples that the notion that works as the bedrock for a scientific naturalization of the mind is still far from being clearly defined.

7 This might lead to the open debate by which embodied and situated cognition is a version of behaviorism. I consider that it depends on what we understand by behaviorism, since Watsonian stimulus-response behaviorism is not the only behaviorism in town. Look to the chapter of Miguel Segundo-Ortin and Inés Abalo-Rodríguez in this volume.

contemporary concerns about representationalism's epistemic opacity and lack of ecological plausibility. His emphasis on norm-governed, embodied practices as the true site of mental activity foreshadows the 4E shift away from computationalism toward a more relational, action-oriented account of mind. Thus, Ryle's critique does not merely anticipate 4E themes superficially; it engages foundational issues that 4E cognition would later systematize, offering an early and incisive philosophical resistance to the internalist assumptions of classical cognitive science.

1.3 Ecological psychology as a precursor to 4E cognition

Before the 4E framework was fully articulated in the 1990s, many of its central insights were already present in ecological psychology. James and Eleanor Gibson's theory of direct perception stands as a foundational precursor to both enactivism and embodied cognition. His ecological framework was developed from the 1960s to the 1980s to explicitly confront both cognitivism and behaviorism and offered a vast in vivo amount of experimental data in humans to explain how we perceive and learn to perceive affordances. His notion of affordances—possibilities for action specified by the environment—provided a new way to conceptualize perception not as passive reception but as active engagement. This ecological approach was already embodied, nonrepresentational, and situated before the inception of 4E cognition in the 1990s.

Crucially, the Gibsons rejected the idea that perception depends on constructing internal models of the world (in their view, they opposed "enrichment" theories of the stimulus information). They emphasized instead the lawful regularities in the environment that agents can detect through movement and active exploration. This embodied engagement with a richly structured world allows organisms to perceive directly the opportunities for action that matter to them.

This ecological perspective laid the groundwork for many 4E theories. The enactive approach, for instance, adopts Gibson's emphasis on movement and interaction but extends it by drawing from phenomenology and systems theory. In fact, they embrace a notion that the Gibsons rejected: the idea of sensation. The Gibsons considered that sensations where problematic since they are the product of, let's say, "filtering the world through the senses", and proposed the idea of ecological information (Not information processing, but the informational structures of the environment—not internally processed; instead, they directly guide action during real-time exploration and the perception of affordances). For the Gibsons, the senses were not really senses but perceptual systems that extended through the body and included action whereas enactivists accepted sensations and tied them to action, leading to the idea of sensorimotor contingencies. The mastery of sensorimotor contingencies is what leads to cognition, according to enactivists. Authors such as Francisco Varela, and Evan Thompson, based on their emphasis on

sensorimotor contingencies, argued that cognition is a form of sense-making: organisms enact or bring forth their world through embodied activity.

Similarly, the embedded and extended mind theses can be traced back to Gibson's insight that cognition is shaped by and often realized through interaction with structured environments. Tool use, linguistic practices, and cultural artifacts—all central topics in extended mind research—can be understood as affordances in the Gibsonian sense: environmental structures that enable and constrain cognitive activity. However, the initial understanding of the extended mind as it was proposed by Clark and Chalmers was still too cognitivist: it did not deny mental representations and endorsed functionalism. In fact, Clark and Chalmers' extended mind merely leads functionalism to its ultimate consequences: if Otto's notebook can functionally replace Otto's hippocampus for finding the way to the museum, what matters is the *functional role* of the piece (the brain, the notebook), not its material irreplaceable contribution to Otto's mental states. On the contrary, ecological psychology departs from the organism-environment coalition and understands that cognitive capacities (in particular, perception and action) heavily depend on the reciprocal affections and irreplaceable contributions of that particular history of interactions. This idea has been adopted by enactivism as well.

Moreover, ecological psychology challenges the very dichotomy between internal and external that underlies much of classical cognitive science. Gibson's view renders this distinction moot by showing that perception and cognition are inherently relational: they emerge through attunement to a world that is already meaningful for the organism. In this respect, ecological psychology resonates with Wittgenstein's notion of public language-games and with Ryle's emphasis on intelligent practices of agents via applying shared criteria correctly. Just as meaning is not a mental shadow of words but something enacted in use, so too is cognition not an inner shadow of behavior but something enacted in skilled activity.

In this sense, ecological psychology is not just a historical antecedent to 4E cognition but an ongoing source of insight. It grounds the 4E perspective in an empirically robust framework that links perception, action, and cognition without resorting to internal representations. It also helps counter the notion, still common in some corners of analytic philosophy, that 4E accounts are vague or speculative. On the contrary, ecological psychology offers a rigorous scientific basis for the embodied and situated view of mind.

1.4 Thinking is not having thoughts, is something we do in the world with others

To think is not to manipulate symbols in the head but to act meaningfully in the world. This view was shared by authors as diverse as Ludwig Wittgenstein, Gilbert Ryle, Martin Heidegger, Maurice Merleau-Ponty, John Dewey, and James Gibson. This is the core insight that unites post-analytic

philosophy with 4E cognition. Figures like Ryle in the analytic tradition anticipated the embodied turn by rejecting the reification of mental states and insisting on the public, performative, and situated character of thought. The same ideas were anticipated by pragmatists such as John Dewey and ecological psychologists like James and Eleanor Gibson. Contemporary 4E approaches extend this vision, using tools from mainstream sciences, but also enhancing the conceptual corpus of available resources for creating an innovative approach to the mind.

The time is ripe for a renewed dialogue between analytic philosophy and 4E cognition. This chapter has tried to sketch the outlines of such a conversation, showing how, in historical terms, their shared concerns with normativity, practice, and interaction can lead to a more grounded understanding of mind. In doing so, we move beyond the myth of internal representations and toward a philosophy that is as dynamic, embodied, and situated as the minds it seeks to understand.

The rest of the volume is divided into three sections: the first one analyzes the role and impact of analytic philosophy from an embodied and situated perspective; the second delves into the ideas of intentionality and language, and the third analyzes the political and ethical consequences of adopting a mixture of embodied-situated ideas and analytic tools for dealing with key societal issues such as freedom, the impact of digital environments in our analog bodies, neurodiversity, or social recognition.

Bringing together the careful conceptual analysis of analytic philosophy with the dynamic, interactive models of 4E cognition allows us to ask new questions and revisit old ones with fresh insight. It allows us to see thinking not as an activity sealed off in the head, but as something we do with our bodies, our tools, our communities—and the world around us.

This volume aspires, with due modesty, to offer a small yet sincere contribution to an ongoing conversation far richer and deeper than any single work could encompass.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Part 1

The analytic tradition through the 4E perspective

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

2 Capturing the ordinary

Alva Noë

> Die Philosophen haben die Welt nur verschieden interpretiert, es kommt aber darauf an, sie zu verändern.
>
> —Karl Marx

What is habit anyway? There has been a tendency within enactive philosophy to think of habits as something like the body's rules. Habits, thought of this way, shape our understanding of how to go on, what to do, how to do it, how things are done, what feels right. To think of habit as operative in our experiential or perceptual or linguistic lives is to think of these activities, or of ourselves, as put together a certain way, as organized, as governed or ruled over by certain automaticities and ways of unthinkingly carrying on. To acquire a new habit—as, for example, when one incorporates a new technology into one's repertoire—is for one to get organized, get designed, anew.

This way of thinking about habit—influential no doubt in enactive circles—points to analytic philosophy's striking legacy. For habit, on this way of thinking, functions as something like a proxy for the fixed rules and norms governing language and conceptuality, precisely as this has been understood in the analytic tradition. Analytic philosophy's drive to make explicit the ways of thinking and talking that are licensed by *ordinary* understanding, by *commonsense*, or by *reason,* finds their analogue, in philosophy after the enactive turn, in our effort to bring out the ways in which habitual doing, skillfulness, and know-how enable human experience.

First philosophy, in the analytic tradition, is philosophy of language; we must delimit the bounds of sense so that we can resist crossing over into the metaphysical dark side. And there is a similar orientation, perhaps, to be discerned in the enactive tradition. Still now first philosophy is the philosophy of habit, or maybe, the philosophy of the body understood as *habit incorporate*. The objective is to unveil the scope and limits of our primordial, habitual, ways of engaging with and being in the world.

Read this way, enactive approaches to language, consciousness, perception, etc. can seem almost like a modern-day form of logical behaviorism, unweaving the ways in which the inner, experience, is actualized in

DOI: 10.4324/9781003649359-4
This chapter has been made available under a CC BY 4.0 license.

movement, action, interaction, and engagement. This is precisely what some early critics of enactive philosophy charged (e.g. Block 2001). And this served also as a reminder that logical behaviorism is fatally flawed insofar as it seems dedicated to the idea that we can view experience as a logical construct of its interior causes and external behavioral effects, as Putnam (1965) had argued.

But enactive philosophy is no logical behaviorism, and enactive philosophy's break with analytic philosophy is more decisive than it may seem. The enactive starting point, its insight, at least as I understand this and have sought to develop it in my own writing, is that we *make* our experience—experience is something we do or accomplish—but precisely, and this is crucial, in the absence of *stable* habits or *uncontroversial* know-how. This is not to say that there are no habits or skills, no rules in that sense. It is just to say that these are always, from the beginning, problems.

So, for example, I analyze perceptual presence in terms of skill-based availability. *To be there* is to be, in effect, in reach. But remember that presence is always fluctuating, partial, fragile, and indeterminate, it is never final, and precisely because the same is true of habit, of can-do, of know-how, of skill.

In this chapter, I reflect further on philosophy in the analytic tradition and on the question of how to situate the enactive approach in relation to this tradition. But my real goal is to try to bring out what I think is the still poorly appreciated radical promise of the enactive moment in philosophy. And this has everything to do with habit, with the ordinary, and with a phenomenon that I call Entanglement.

2.1 The Logical conception of language

It will be helpful, as a way of better framing the topic, to recall the way logicians think about formal language.

A formal language, the sort of systems that logicians work with, consists of a finite number of primitive or atomic symbols and a set of rules or procedures for determining, for any string of symbols, whether that string is also a symbol, whether it is, in the terminology of logicians, a well-formed formula. If it is, then good; if it isn't, well then, it's prohibited by the rules. And so for meaning. There are assignments of meanings or "semantic value" to every primitive symbol, and there are rules for determining, given the meaning or semantic value of each symbol, what the meaning or semantic value of each well-formed formula is. If a sign lacks a proper assignment, or if the signs are combined illegally, then what you have is not so much meaningless language, as non-language.

This conception of language—I will call it the logical conception of language, or the LCL—has been taken for granted by many thinkers working in the analytic tradition and is, I would venture to say, its default conception. Language is generated by the rules. And what is not generated by the rules is non-language. We can see this guiding idea at work in research in some

empirical linguistics which aims, specifically, to make explicit the rules and representations, as Chomsky 1980 called them, that suffice to specify the language and that, also, therefore, can be thought of as telling you what it is you know when you know a language.

What is distinctive of the LCL is not only its conception of language as formal, that is, the idea that language is basically a system of formally specified marks or signs detached from the conditions of human life in which those signs are deployed, but further that this conception draws a sharp line between what *is* language, by dint of being generated by the rules, and what, by dint of not be generated by the language, falls outside of language.

It was against the background of precisely this LCL that golden-age analytic philosophers like Carnap, Schlick, or AJ Ayer were able to argue that some uses of language—which they linked to metaphysics, religion, ethics, and esthetics—were strictly speaking nonsensical. The impulse to utter or write such would-be pseudo-propositions stems from a kind of insensitivity to the rules of one's own language, the logical order of true language. Philosophy becomes, in this setting, a kind of language criticism, and its main commandment is something along the lines of *shut up; s*top talking and pay better attention to the rules governing what it ordinarily makes sense to say.

It was likewise against the background of the LCL that Carnap and his colleagues offered linguistic accounts of the distinctions between the analytic and the synthetic, the a priori and the a posteriori, and also between the necessary and the contingent. They argued, in effect, that these are all linguistic differences—reflecting no deeper ontological significance and that statements that appeared to be universal and absolute and knowable in ways that outstrip the resources of ordinary experience were either pieces of concealed or disguised nonsense (as with metaphysical statements and the Kantian synthetic a priori), or were show to be, on proper analysis, merely formal, merely linguistic, merely, in effect, sense*less* (as opposed to nonsense), just conventions or tautologies or whatever.

If you were trained in the analytic tradition,[1] then you know that the conception of language as *formal* came under increasing criticism in the years leading up to and following World War II (due initially entirely to the work of Wittgenstein). But soon Ryle, Austin, and others became alert to the diversity of linguistic rules and norms, their lack of uniformity, and the need for an appreciation of what came to be called the different "logics", or "grammars", or "rules" governing ordinary thought and talk.

1 My dissertation advisor was Hilary Putham. My other teachers included: Burton Dreben, Michael Dummett, Warren Goldfarb, Peter Hacker, Hidé Ishiguro, Robert May, Charles Parsons, Peter Strawson, all of whom are or were champions of analytic philosophy.

But even with the emergence of Ordinary Language Philosophy, the basic picture of the LCL was unchanged. Language has rules—they are just a bit messy or local in character—and so language continues to have sharp boundaries. According to the new version of the LCL, philosophers—when they are interested in questions about personal identity, or other minds, or fictional objects, or number, or truth, or God, or whatever—run the risk of violating the familiar, ordinary rules governing what it is correct to say and how it is correct to answer questions. They cross over into nonsense. And so, correctly construed, philosophy's job remains that of policing the limits of language, patrolling the borders of what can and what can't be said, and, importantly, getting philosophers, finally, to *shut up*, getting them to stop asking the questions that drive them across the bounds of sense.

Ordinary Language Philosophy, then, no less than Vienna Circle Logical Positivism, adhered to the idea that language is fixed by the rules and that to break the rules is, as it were, to be ejected from language into outer space; or at least it is to deserve to be so ejected.

Now, later on, the LCL did in fact come under more penetrating criticism. But here again, the attacks, though important and revealing, have done little, so far as I can tell, to wean analytic philosophy off its reliance on the LCL.

Consider first the hugely influential views of Putnam (1975) and Kripke (1972). They agreed with Grice and Strawson (1956), and Kant, as against Quine (1951), that there is a distinction to be drawn between the analytic and the synthetic, even if it is not a sharp one; but their deeper sympathy was really always with Quine. For even as they upheld the existence of the distinction, they agreed with Quine that the distinction was not of any serious philosophical importance. After all, as they argued, analytic truths might sometimes be contingent (e.g. "lemons are yellow"), and necessary truths might sometimes be a posteriori (e.g. "Water is H20").

At first glance, the position developed by Putnam and Kripke seems to make a break with the LCL. This is because, in its "externalist" commitment to the idea that meanings do not supervene on "internal" rules of usage, it seems to offer a conception of language as world-involving in a way that obliterates the sharp boundary between what is and what is not language.

But this appearance is misleading. Putnam and Kripke, in their appeal to "dubbing ceremonies", in fact do little more than revert to the initial idea, a starting assumption of the LCL, that assignments of meaning are made prior to the use of language. Crucially, for them, the fixing of meanings is done antecedently to the use of language, rather than—as Wittgenstein, and I, would say—from within language. The distinction between mere descriptions and reference-fixing descriptions is a way of re-affirming the LCL. What you are talking about when you talk about water, or gold, or whatever, is fixed not by the world, as their rhetoric would have it, not really, but rather by the language model that is fixed in place. (This basically a Wittgensteinian criticism.)

A deeper criticism of the LCL is to be found in Quine's writings. As already mentioned, Quine attacked the idea that there is a sharp or clear or steady distinction between empirical or synthetic truths, on the one hand, and those, the logical or the analytic ones, whose truth is somehow guaranteed by meanings themselves, on the other. In his framework, it no longer even really makes sense to speak of "misusing" language, as distinct from being irrational, or making false statements, for there is no longer a clear conception of language's rules, norms, or meanings, to be contrasted with "substantive" questions of truth and falsehood. For Quine, would-be empirical statements like "There are brick houses on Elm Street" and would-be analytic statements like "Bachelors are unmarried men" are, finally, epistemically, and modally, on the same footing: both kinds of statement may be falsified, in principle, or, alternatively, both kinds upheld in the face of any and all evidence come what may. For Quine, then, there are not two kinds of statement here after all, but one. If there are differences in how we tend to use these statements, these are of superficial pragmatic significance at best.

And what goes for statements, for Quine, goes for the things we talk about, and think about, as well. Whether there *are* brick houses, bachelors, angels, atoms, parking meters, ducks, or machine guns is a matter of how we best choose to distribute truth and falsehood among the network of statements of our language; and how we do that will be guided by our desire to keep things simple and practically effective.

So Quine really does seem finally to be done with the LCL. Precisely, in contrast with the critical standpoint of Logical Positivism and Ordinary Language Philosophy, Quine denies that there is any standpoint from which we can delineate the bounds of sense—in Strawson's phrase—and with respect to which we might choose to criticize those who transgress the bounds. For Quine, such talk no longer makes any sense.

Now Quine's criticism is important and worth much more discussion than I give here. But I am unpersuaded that it actually rises to the level of breaking with the LCL.

First, as Grice and Strawson persuasively argued, to show that there is no *sharp* or *uncontested* distinction between the analytic and the synthetic is not the same as showing that there is no distinction at all. In particular, it is to leave open the possibility that the very fact that this distinction *is* contested and contestable may be critical to the kind of distinction it is.

What makes it possible to repudiate the distinction between the conceptual and the empirical, the analytical and the synthetic, in the way that Quine recommends—and this is a second point—is precisely his backsliding on what is, I think, actually the deep insight achieved by mid-20th-century analytic philosophy, namely, that we don't only ever do *one thing* with words, that language is precisely *not* one systematic organized rational fabric adjusted in light, as Quine would have it, of our singular interests

in pragmatically maximizing truth and predictive power. The people we know and love and care about are not posits, in the way that perhaps we can say that electrons are, and the being or non-being of God is a different matter yet again. By insisting, as Quine does, that we operate always within a single scheme, something like a global *theory,* he is reinstalling the very feature of the LCL that, as we will see, is most striking and, indeed, most suspect, namely, the idea that language is a fixed whole, with knowable boundaries.

There is more to be said about all this. But for now let me summarize: as far as I can tell, analytic philosophy has neither managed to free itself from the LCL, nor frame a viable alternative to it; when analytic philosophy strives to capture or recapture the ordinary, through linguistic analysis, what they have in mind is the idea that we need to recover a better understanding of the scope and limits of our fixed linguistic and conceptual framework.

2.2 Language and enaction: an alternative to the LCL

I began by introducing the LCL and illustrating its basic role in analytic philosophy's self-understanding. But there is another very striking fact about the LCL. And this is that it is completely misguided. We must reject it entirely. To do so will let us rethink philosophy itself, since the conception of philosophy, as we have seen, at least in the analytic tradition, is closely tied to this implausible conception of language. At ground, the basic problem with the LCL and the associated conception of philosophy is that it fails to come to grips with what I call the Entanglement.

To begin with, consider that language is *fragile*: one of the distinctive features of true language, as opposed to that of logical systems, is that it is always confronted by the live and immediate possibility of misunderstanding. And as a general rule, misunderstanding doesn't interrupt language, forcing us outside of it, as the LCL would have it; for misunderstanding is for us always an opportunity for more language, that is, for the distinctively linguistic activities of explaining, or clarifying, or elucidating, or justifying. Language users do not just carry on automatically, acting in accord with rules that govern them, occasionally misusing words and finding themselves then ejected into linguistic outer space. Rather, language users, from the very start, as it were, use language to make meaning in the face of misunderstanding. We define terms; we challenge another's usage; we explain what a term or word means. The range of evaluative reflection on language is very wide. We find some bits of discourse *clear*, others *murky*, some *humorous*, others *dull*, and so on. There are many distinct domains of critical reflection on talking that unfold inside language: logic, rhetoric, style, wit, sophistication, etc.[2]

2 See Strawson 1952 for the source of this idea.

Even very young children get this. One of the first uses of language that you see kids playing around with is that of asking after or offering definitions or explanations of meaning. And this is to have said nothing about all the ways in which we can criticize language ethically, politically, for its violence or power.

Second, language is a *rule-making*, and so in this sense, a *critical* or *normative*, activity. Let me explain: we like to speak of language as a rule-governed activity. But this is wrong. Language is a rule-*making* and rule-*using* activity. That is, speakers deploy rules as instruments to guide themselves, criticize the usage of others, adjudicate dispute, and, in similar ways, negotiate their dealings with each other. Language, in this sense, may be called a normative or evaluative *activity*. To use language is to be concerned with the question of how one ought to do things, how one ought to go on.[3]

Third, critically, language is *productive*. Reflection on language, responses to disputes or challenges, the need to find better ways to be in connection with others, all this *changes* language, renews it, alters it, drives it to evolve and to change. Indeed, the very act of trying to make sense of what we are doing when we are using language has the effect of changing language.

Fourth, and in a way this is just a restatement, language is *ungoverned*; there are and can be no language authorities; or, alternatively, each of us is authoritative as it is possible to be.

And so language is, in my sense, *entangled*: To be a language user is perforce to be one who takes a stand on language, who cares about usage and feels called on to offer corrections; it is to be one who copes with difference and disagreement. This is why I say language is a rule-using (or maybe even sometimes a rule-creating) activity as opposed to a rule-governed one. And this is why to be a language user is, whatever else it is, to be someone who thinks about language. Language is entangled in the sense that the second-order concerns belong to its first-order operation.[4] The object language always contains its metalanguage. This is why, as I argue further in recent work (Noë 2023), language is always in a way written, even before there was writing in history. For what is writing but a model or picture or face of language so that we may think about it. Whether we actually use graphical means for this end, if we are language users, we always reflect on language as an object of concern *and* that is the moral equivalent of writing.

To imagine speakers who just carried on, and never needed to reflect on what someone meant, or might have meant, would be to imagine something utterly unlike real human language.[5]

3 This has been a theme in the work of JC van den Herik (e.g. 2017 and 2022).

4 See Noë 2023 for further explication of this idea of "entanglement."

5 This sheds light on the question of machine language. It is because machine language does not participate in the *entanglement* described here, that machine language isn't really language.

The upshot of this fragile, productive, anarchic, normatively animated, entanglement of language is that, in a way, there is no such thing as language, at least as this has been understood in the analytic tradition according to the LCL.

Let me clarify:

First, precisely in contrast to the LCL, we are now in a position to say there are no sharp boundaries marking off what is inside language, and what is outside; there is no outside, but there is no inside either. Language is a porous, open, *field of activity*. It is never fixed. Debates about language cannot be settled by appeal to any court, not a court of reason, nor a court of nature.

Second, language is not a formal object; to think that it is to be confused by writing, that is, it is, as Husserl might have said, to be misled by the surreptitious substitution of writing, which is a graphical rendering and interpretation of speech, for speech itself. We forget that the model is one thing and the reality something else.

But our point is more general: there is no separation between language and the settings of our lives that are, as it were, linguistically animate. A human being is hungry, or full of desire, *in language*. Language is not a segregated or separate thing; it is, rather, the modality of our human living. We touch each other with our words. We hold each other back with our words. We harm each other with our words. We raise each other up with our words. With our words, we make, do, build, and enter into experience with each other. And there is even a sense in which, however quiet we may be, however intimate, however sheltered, *wordlessness* is not an option for us.

According to the enactive picture of language that emerges here—according to which language is fragile, productive, anarchic, normative, processual—there is never a court of appeal where we might settle our linguistic controversies. But it is also true that there is no end of controversy. We do language, we language; and we need also to fight over it, work to achieve it. This is entanglement.

Notice then that this authentically enactive conception of language makes a sharp break with the LCL and so from the whole analytic method in philosophy. In place of the idea that philosophy is a kind of linguistic analysis we are left with an astonishing and, I think, astonishingly radical conception of language itself as always already philosophical. To say that language is entangled is to say that language is the site of that work of clarification, criticism, illumination which, according to analytic philosophy, is philosophy's distinctive mission. The point is not that analytic philosophy got it wrong. The point rather is that, in its blindness to the entanglement, analytic philosophy can't quite understand what it is itself really doing.

But to understand this, we need to introduce a new theme into the discussion. This is the theme of resistance.

2.3 Getting beyond habits/alone together

I have been emphasizing that we make language. We enact it. But it is also important that we appreciate that none of us *invents* language. Nor is it something that we unproblematically inherit. We don't own language. We simply find ourselves saddled with language. We learn our languages as a process of adapting or conforming to local conditions. Language is a system of habits, thoughts, skills, attitudes, values, bodily gestures, and we learn these as we learn to be people, or to be straight, or gay, or as we learn to digest, or to masturbate. We learn to speak, as I will say, alone together.

The focus on skills and habits in enactive philosophy has tended to be all peaches and cream; this may be another one of analytic philosophy's legacies. Conceptual abilities and sensorimotor skills are the means whereby we achieve access to what there is. But we forget—we neglect—that ability is always also limitation, and that human mastery is always a story about enforced conforming, about the demand for accommodation, and that wherever we find distinctively *human mastery* we also always find resistance to the very skills and norms and requirements that we are forced to master. This is a consequence of the entanglement of human life. Wherever there is a downward arrow of skillful mastery and incorporation, there is also always an upward arrow of resistance.

For example, we focus so often in theories of perception on perceptual success; we think of the ability to see, for example, as our birthright, a simple matter of biological endowment, and we ignore the myriad ways in which real seeing, real perceiving, that is to say, the entering in to and the sustaining of lively *relationships* with people and places and situations, is something that we need to accomplish and that we frequently fail to accomplish.

Take a trivial case. I can't even see the words of the foreign writing. Here, my perceptual skills and my habits fail me. But this is just to say that my own make up, my history, my life, my *self*, fall short. To know what is there, to perceive it, I need to change myself, reorganize myself, resist the ways I have been organized. To read novel writing, I need to get reorganized.

It should not be surprising, then, that consciousness and its history are ethical and political domains. This is what is at stake, I think, in recent exploration in connection with what some people now call "the queer". The queer is what you cannot see, because although it *is*, and although it is there before you, you are unequipped to know it, or understand it; it is wrong somehow. It is your limitations, experiential, cognitive, and yes, ethical, that are the sources of your inability to be in a relationship, or to stay in conflict.

2.4 Back to philosophy

Analytic philosophy has tended to be the cult of Descartes. Philosophers are solo explorers; they are confined to arguments starting from what they alone can clearly and distinctly conceive.

Enactive philosophy is no cult, but if it were, we might celebrate Vico, who insisted that wit, understanding, and insight are the stuff of poetry, song, conversation, literature, and criticism, for what we know best, what we care about most, is what we ourselves make.

But that, the stuff that we make, includes ourselves. For we are ourselves not fixed in nature, once and for all, but are rather processes of becoming, and this is a process that is nourished and driven precisely by the work of trying to know ourselves, by the work of unveiling ourselves to ourselves, by the work of catching ourselves in the act of making sense, and recoiling from nonsense, or working to know what there is even when we are barely equipped to do so.

In conclusion, I would like very briefly to make two points, each of which deserves further elaboration.

First, it is not only analytic philosophy with its reliance on the LCL that fails to accommodate entanglement. Hubert Dreyfus (e.g. in the papers collected in 2014), on behalf of existential phenomenology, has insisted on a sharp contrast between first-order engagements with tasks or activities and the interruption of such activities for purposes of reflection or self-monitoring. When we are in the flow, we just act; reflection happens only when there is breakdown.

I am sympathetic with Dreyfus that we must be vigilant to ward off an intellectualism or a cognitivism that holds that human activity only rises to the level of action when it is accompanied by deliberate psychological acts of detached evaluation and contemplation.

But it is instructive to notice that Dreyfus's view conforms most perfectly to the artificiality of the LCL. Dreyfus's opposition of *flow* and *breakdown* corresponds perfectly to the logician's conception of what is inside and what is outside the bounds of language. The use of language to adjudicate and regulate and indeed to reflect on language is one of language's fundamental *first-order* modes. To worry about language, to reflect on it, to take up the writerly attitude to language, is *not* to interrupt language, but to enact it. Language contains its own meta-theory; or better, language contains, always, and from the start, the problem of *how to go on?* as well as that of *what's going on?* Reflection on and argument about language, second order though they may be, are already contained within language as a first-order phenomenon.

Tripping, arguing, adjudicating disputes, innovating, explaining, articulating, trying better to express—these are *ready-to-hand* modalities of ordinary, everyday language use. Criteria of correctness, questions about how to go on,

or about what is or is not grammatical, dealing with misunderstanding, these are activities that we carry on, and that we fight about, inside of language, and they do not require us to shift, as the logician might have it, to a language external meta-activity of setting up the grammar.

Second, if you think of analytic philosophy as the style of philosophy dominant in English-language universities for the last hundred years or so, a style of philosophy that tends to be characterized by scientism, by the preëminence of the article instead of the book, by an indifference to history, by a self-satisfaction in its ignorance of work conducted in other philosophical traditions, by its confidence that philosophy and its work can unfold outside of politics and the demands of value, and finally, by its fetishizing of clarity, then, it turns out, the only truly great analytic philosopher, the very source of analytic philosophy, was no analytic philosopher at all. I am thinking, of course, of Wittgenstein. Space does not permit me to say more about this, but I'll mention that I suspect that Wittgenstein's philosophical achievement is, in good measure, to have appreciated the entanglement in many of its most important dimensions.

The problem with analytic philosophy is that it in its naïve devotion to explicitness and clarity, it refuses to acknowledge the ways in which what matters to us resists clarity, reduction, explanation. What analytic philosophy, with its blindness to the entanglement, has been unable to get hold of is the fact that philosophy, in its effort to capture the ordinary, is always aspiring not just to getting clear, but to freeing us from the ways habit, culture, technology seem to make us up.

References

Block, Ned. 2001. "Behaviorism revised," a commentary on O'Regan, J. K. and A. Noë's "A sensorimotor approach to vision and visual consciousness." *Behavioral and Brain Sciences* 24(5): 883–917.

Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press, 1980.

Dreyfus, Hubert. 2014. *Skillful Coping: Essays on the Phenomenology of Everyday Action*, edited by Mark Wrathall. New York: Oxford University Press.

Grice, Herbert Paul and Strawson, Peter Frederick. 1956. "In defense of a dogma." *Philosophical Review* 65(2): 141–158.

Kripke, Saul. 1972. "Naming and necessity". In Davidson, Donald and Harman, Gilbert, eds., *Semantics of Natural Language*. Dordrecht: Reidel: 253–355, 763–769.

Noë, Alva. 2023. *The Entanglement: How Art and Philosophy Make Us What We Are*. Princeton, NJ: Princeton University Press.

Putnam, Hilary. 1965. "Brains and behavior." In Richard J. Butler, ed., *Analytical Philosophy*, vol. 2, Oxford: Blackwell, pp. 1–19.

Putnam, Hilary. 1975. "The meaning of 'meaning.'" *Minnesota Studies in the Philosophy of Science* 7: 131–193.

Quine, Willard Van Orman. 1951. "Two dogmas of empiricism." *The Philosophical Review* 60(1): 20–43.
Strawson, Peter Frederick. 1952. *Introduction to Logical Theory*. London: Methuen.
Van den Herik, Jasper C. 2017. "Linguistic know-how and the orders of language." *Language Sciences* 61: 17–27.
van den Herik, Jasper C. 2021. "Rules as resources: An ecological-enactive perspective on linguistic normativity." *Phenomenology Cognitive Science* 20: 93–116.

3 Ecological psychology as critical direct realism

Carl B. Sachs

3.1 Introduction

Despite being contemporaries and eminent in their respective disciplines, there is no evidence that Wilfrid Sellars and James J. Gibson knew of each other's work. I regard this as a missed opportunity for what could have been a productive encounter between one of the 20th century's greatest philosophers of mind and of science and one of the 20th century's great philosophical psychologists. Both were deeply American thinkers—pragmatist, skeptical, and anti-dogmatic. And both were heirs, in their own ways, to the pragmatism of James and Dewey, the neobehaviorism of Tolman and Hull, the Gestalt psychologists, and Wittgenstein. Both made important contributions to an anti- or post-Cartesian scientific metaphysics of mind—Sellars most notably in his critique of what he called "the Myth of the Given" and Gibson most notably in his discovery of affordances as what can be directly perceived by sentient, mobile animals.

More importantly for present purposes, both Sellars and Gibson were influenced by the early 20th-century debates about direct realism. These debates took shape in reaction against 19th-century Anglophone idealism, which came to be seen as incompatible with a robust scientific and progressive worldview. Two rival positions arose: new realism and critical realism. The new realists (Montague, Perry, and Holt) held that veridical perception directly takes in events, objects, and meanings in the world. The critical realists (Drake, Santayana, Roy Wood Sellars) insisted that sensations causally mediate and guide perception. Though Wilfrid Sellars is usually seen as taking up his father's critical realism (as he himself admits) and Gibson is usually seen as taking up Holt's direct realism, I want to suggest that their positions are closer than their inheritances would suggest.

I shall begin with a brief discussion of what Sellars calls "an adequate critical direct realism" in the metaphysics of perception. This position is grounded in Sellars's rejection of phenomenalism and his interest in a scientific metaphysics of perception—both of which he shares with Gibson. I will then consider whether an ecological approach to perception, as

DOI: 10.4324/9781003649359-5
This chapter has been made available under a CC BY 4.0 license.

characterized by Gibson, would count as a kind of critical direct realism. The issue here is not whether Gibson's view is a form of direct realism, but whether it is a form of *critical* direct realism. I shall argue that ecological psychology is a neglected alternative to Sellars's position: a version of critical direct realism that *rejects* the necessity of sensations for perception. Yet, the supposed necessity of sensations for perception is itself crucial for the cognitivist enterprise, since it is thought that some sort of intellectual processing is necessary for transforming sensations into perception. However, I shall argue that Sellars's version of cognitivism can not only survive the rejection of sensationalism about perception but also is largely compatible with ecological psychology, despite Gibson's own criticisms of the cognitive revolution as he understood it.[1]

3.2 Sellars's critical direct realism

A guiding motif of Sellars's philosophical project was his commitment to embracing as many forms of realism as possible: direct realism in perception, scientific realism, and even moral realism. More importantly, he recognized that his campaign against instrumentalism in philosophy of science required a critique of phenomenalism in philosophy of perception. This is because construing the objects posited by scientific theories as mere devices for predicting future experiences has often been regarded as but a short distance away from construing the objects of perceptual encounter as mere devices for predicting future sensations. And this was a line of thought that had been nicely exploited, first by Berkeley, but also by John Stuart Mill, C. I. Lewis, and Bertrand Russell. It was not a position that any intellectually adequate naturalism in Sellars's day could afford to ignore.

In his 1959 "Phenomenalism", Sellars begins his critique of phenomenalism with a characterization of what he calls "direct realism", which holds that (for example) "seeing that a leaf is green is not a matter of seeing that it looks green and inferring from this, together with the circumstance of perception, that is green" (Sellars 1963a, p. 61). That is, direct realism denies that perception *begins* with claims that are framed in terms of what things *look* like and then infers what things *are*. Instead, color predicates (and presumably predicates referring to states of other sensory modalities) are attributed to physical objects as they exist at particular times and places. The alternative to direct realism, which Sellars calls

1 It should be noted that Millikan (2000, 2004, 2007) also draws extensively on both Sellars and Gibson. However, to the best of my knowledge, she does not engage with Sellars's philosophy of perception with respect to Gibson's psychology of perception.

phenomenalism, fails for the interesting logical reason that it cannot even be coherently formulated:

> if the reformulation from the language of physical objects to the language of sense contents were carried out step by step it would not only be an endless regress, but it would involve a *circulo in definiendo*, 'eye,' for example, being explicated in terms of 'eye'.
>
> (80)[2]

Despite this wide-sweeping critique of phenomenalism, there is one theme from some versions of phenomenalism that Sellars retains. It is the idea that sensations lack intentionality: there is nothing that the sensation of red is *about*, *refers to*, or *stands for*. Not only is it the case that "having a sensation is *not* a conceptual fact" (although knowing that one has a sensation *is* a conceptual fact) but that this requires saying, *contra* Aristotle, that the difference between sense and intellect is that "between a 'raw material' which involves *no* consciousness of anything *as thus and so* on the one hand, and any consciousness of something *as thus and so* on the other" (74). Sellars reiterates this point in "Being and Being Known", written a year after "Phenomenalism":

> sense is a cognitive faculty only in the sense that it makes knowledge possible and is an essential element in knowledge, and that of itself it knows nothing. It is a necessary condition of the intentional order, but dos not of itself belong to this order … sensations have what I shall call a *pseudo-intentionality* which is easily mistaken for the genuine intentionality of the cognitive order.
>
> (Sellars 1963b, p. 46)

It must be noted that although Sellars is right to say that sensations do not belong to the intentional order, and also right to say that sense does not belong to the intentionality *of the cognitive order* (as Sellars would construe it), it does not follow that sense, as contrasted with intellect, does not exhibit its own kind of non-cognitive, or better, non-epistemic, intentionality.

The conclusion to be reached in the critique of phenomenalism is that we should provisionally assert that "physical objects are really and directly perceived, and that there is no more basic form of (visual) knowledge than

2 In "Physical Realism" (1954), Sellars puts the point in terms of predicting future sense-impressions from past and present sense-impressions, and argues that the prediction cannot go through without relying on assumptions about the perceiver's own body as a physical object that is causally affected by other physical objects.

seeing physical objects and *seeing that* they are, for example, red and triangular on this side" (Sellars 1963a, p. 87). The question is now what is meant by "direct" in "direct realism"? Sellars's point here is less psychological than epistemological: "to say that someone directly knows that-p essentially involves the fact that the idea that-p occurred to the knower in a specific way" (88). More specifically, it involves the idea that the knower can be in a general kind of condition under which people can and should be taken as authoritative about their own experiences, together with the particular fact that the knower is presently in a condition that belongs to that general class (EPM VII. 36–38), that is, that the knower is taking up a position in the logical space of reasons—a position that is *noninferential* though not *presuppositionless*.

The direct realist as now characterized is someone who affirms that among the physical things, there are some that are noninferentially perceivable under appropriate conditions, such that (for example), "a pink ice cube is a directly perceived, public, cold, solid, smooth, pink physical object having the familiar thermal and mechanical causal properties of ice" (89) such that it (1) appears to standard perceives as being pink and cubical, but also (2) is responsible for the fact that there appears to these perceivers that there is a pink and cubical physical object in front of them and also (3) causes these perceivers to have impressions of a pink cube. In short, direct realism is here characterized as the position that there are physical objects with publicly available perceptible properties, such that these properties have the causal disposition to bring about sense impressions (in normal perceivers under standard conditions), where those sense impressions are to be construed as analogous to the perceptible properties of the physical things. The direct realist is someone who affirms that her sense impressions of something pink, icy, and cubical in her visual field have been brought about by the bodily presence of a pink cube as a physical object in her proximity. She is entitled to assert, "that is a pink ice cube", not only because she is having pink-ice-cube sense impressions, but also (and just as importantly) because she knows both that she is perceiving the pink cube under standard environmental conditions and also that her own sensory systems are functioning more or less optimally.

At this point, having defended direct realism (as he understands it), Sellars now insists that "our direct realism be sufficiently critical". The first step is to abandon the abstractive theory of concept-formation, which holds that we acquire the concept *red* by noticing red things. Instead, we must say that "the coming to see something as red is the culmination of a complicated process which is the slow building up of a multi-dimensional pattern of linguistic responses" (90), which includes a long process of behavioral dispositions subjected to social sculpting. Thus, while directly perceiving *that* something is red coincides with having acquired the empirical concept *red*, that empirical concept is entangled in the whole

conceptual framework of physical objects and one's own body as existing in space and time.

At this point, Sellars then takes a decisive step:

> we must take the *second* step towards an adequately critical direct realism. This step consists in the recognition that the direct perception of physical objects is mediated by the occurrence of sense impressions which latter are, in themselves, thoroughly non-cognitive. Step *three*: this mediation is causal rather than epistemic. Sense impressions do not mediate by virtue of being known.
>
> (91)

One might be confused at this point: if we directly perceive physical objects, then why introduce sense impressions again at all? Why not banish them from the lexicon altogether, once we have seen that phenomenalism is incoherent?

Sellars insists on the reality of sense impressions for two reasons. The first is that he thinks of sense impressions as *posits*:

> entities postulated by a theory (at first common-sensical, then more and more refined) the aim of which is to explain such general truths as that when people look in mirrors in front of which there is a red object, there seems to them to be a red object 'behind the mirror', and other facts of this kind.
>
> (91)

And we should be *realists* about sense impressions for two reasons. The first is that we ought to be scientific realists generally about the entities posited by our best scientific theories. The second is that the theory of sense impressions is (supposedly) a good theory. It is a good theory because it explains various kinds of misperception, such as illusions and hallucinations. In those kinds of cases, the sense impressions that are usually brought about by physical objects under standard conditions to normal perceivers have been brought about under abnormal environmental conditions (the bent stick in water, the reflected object that appears to be behind the mirror) or under conditions where the perceiver's sensory systems are not functioning normally (Macbeth seeing the floating dagger).

Thus, while we ought to insist that we do indeed "directly perceive" physical objects, this "directly" means precisely that we do not infer the existence of physical things from anything more directly known, such as sense-contents. On the contrary, Sellars's critical direct realism *inverts* phenomenalism: we do not first observe sense-contents and then infer physical objects, but rather we first perceive physical objects (as a consequence of a long period of socialization in the presence of physical objects) and then infer that there are sense impressions. More precisely, we posit the existence of sense impressions

as causally mediating perception, and we ought to be realists about sense impressions because we ought to be scientific realists in general.[3]

It is precisely at this point, however, that the ecological psychologist is entitled to insist that the sense impression theory of perception and misperception is, *contra* Sellars, not a good scientific theory at all, and that it should be replaced with a better theory: an ecological approach to the study of perceptual systems. Yet, the ecological psychologist can still be a critical direct realist up to this point: they can allow that phenomenalism is a nonstarter and that we do indeed perceive physical things (though this needs substantial qualification). Nor would the ecological psychologist deny the role of linguistic training in coming to use empirical concepts such as *red* or *heavy*. And most importantly, for my current purposes, the ecological psychologist would agree with Sellars that direct realism must be reconciled with scientific realism: the explanatory project of empirical psychology of perception stands or falls with the adequacy of its explanation of non-veridical perception, and not just veridical perception. Where the ecological psychologist will disagree with Sellars is whether there is any need to posit sensations as the causal mediators of perception, whether veridical or non-veridical. However, it must be stressed that the ecological psychologist's successor-concept for sensations is *also* not epistemic or cognitive—at least not in Sellars's sense.[4]

3.3 Gibson's rejection of sensation-based perception

Ecological psychology, as it has come to be known, is said to have begun with the psychological research conducted by James and Eleanor Gibson. Here, I shall focus on some theoretical statements that Gibson articulated on his way toward ecological psychology. Specifically, I want to consider why Gibson thought that a sensation-based theory of perception should be rejected on both conceptual and empirical grounds.

In his 1960 presidential address "The Concept of the Stimulus in Psychology" (Gibson, 1982a), Gibson articulates the central unanswered problem of sensation-based theories of perception: "sensations are specific to receptors, not to objects in the world. And this is the age-old puzzle of sensation-based theories of perception. Sensations must be supplemented. But how?" (p. 348). Beginning with Müller's groundbreaking work, sensory physiologists had

3 See Levine (2007) and O'Shea (2024) for more systematic assessment of Sellars's critical direct realism. It should be noted that Sellars himself *abandons* critical direct realism (1963a, pp. 95–105) on the grounds that physical objects, as described by microphysical theories such as quantum mechanics, cannot have the directly perceptible sensory qualities that critical direct realism ascribes to them. Ultimately, Sellars thinks, all such sensory qualities must be re-categorized as states of sensory consciousness of sentient organisms. See Egan (2025, pp. 109–141) for a contemporary defense.

4 Withagen and Chemero (2012) underscore the difference between perceiving affordances and classifying objects. It is the latter which most concerned Sellars, especially in his contention that "[t]o reject the Mythof the Given is to reject the idea that the categorial structure of the world – if it has a categorial structure – imposes itself on the mind as a seal imposes an image on melted wax" (Sellars 1981a, Section 45, p. 12, emphasis original).

defined sensations in terms of the proximal triggering of specialized cells embedded in specialized tissues such as the retina and cochlea. But the more fine-grained the physiological and anatomical detail became, the greater the challenge of seeing how the triggering of transducers could initiate the causal processes that resulted in perception of physical objects. There seemed to be no easy way to transform the proximal (activity of sensory receptors) into the distal (perception of objects distant in space and/or time). The alternative, Gibson urges, is to simply replace sensations with structured environmental information:

> The conception of structured *array* of ambient light (or an array of contacts, vibrations, or substances) is entirely different from the notion of stimuli that impinge on receptors. Information about the environment consists of the invariants of structure in a continuous flow. … The array *consists* of contrasts and transitions, not of stimuli, and not of groups, patterns, or series of stimuli. … The concept [of a stimulus] applies to a passive receptor, not to an active perceptual system; it belongs to physiology at the neural level not at the level of homeostasis.
>
> (p. 349)

Gibson's central contention is that we should not conflate neural physiology with the organism-environment relation. Physiology and ecology are distinct sciences because they investigate different levels of reality, involving different scales of spatiotemporal resolution. To adopt an ecological approach to psychology is to say, at the outset, that psychology is distinct from physiology; it is a science of mobile, sentient animals as they interact with their environments.

Gibson's critique of sensation-based theories of perception is based on this more general critique that psychologists have been looking to physiology as the criterion for what makes psychology a legitimate science, whereas they should look to ecology. Once that is done, it becomes clear why we should not hope to explain perception in terms of sensations. As he puts it in his 1963 "The Useful Dimensions of Sensitivity" (Gibson 1982b):

> The variables of sensory discrimination are radically different from the variables of perceptual discrimination. The former are said to be dimensions like quality, intensity, extensity, and duration, dimensions of hue, brightness, and saturation, of pitch, loudness, and timbre, of pressure, warm, cold, and pain. The latter are dimensions of the environment, the variables of events and those of surfaces, planes, objects, of other animals, and even of symbols. … Having sensations is not perceiving, and this fact cannot be glossed over. Nevertheless, perceiving unquestionably depends on sensing *in some meaning of that term*. That is, it depends on sensitivity or the use of the sense organs. To observe, one must sense. The question I wish to raise is whether or not it is true that to observe one must have sensations.
>
> (p. 351)

It is of the utmost importance that by distinguishing between both *sense* as capacity and *sensing* as activity from *sensations* as the states produced by that activity, Gibson has taken the first step away from Sellars's conflation of sense and sensation. And this is crucial for distinguishing Sellars's correct thought that sensations lack intentionality from his more problematic thought that there is no intentionality to sensing at all.

In order to distinguish sensing from the mere having of sensations, Gibson found it necessary to formulate a different characterization of the senses. This led to his formulation of the senses as *perceptual systems*:

> Sensitivity is one thing, sensation is quite another. The first meaning refers to the effects of stimulation in general. The second refers to conscious impressions induced by certain selected variables of stimulation. … in the first meaning sensory *inputs* are prerequisite to perception, but in the second meaning sensory *impressions* are not prerequisite to perception. In other words the *senses* are necessary for perception but *sensations* are not. … it might be better to call the senses by a new term such as *esthetic system*s.
>
> (p. 361)

In understanding the senses as esthetic (or perceptual) systems, we come to understand that what matters is that sentient animals have modality-specific sensitivity to environmental information. The problem of perception is that of explaining how actively sought sensitivity to environmental information becomes useful for behavior, not how passively triggered sensations are processed into perceptual experiences.[5] This requires a quite different understanding of the relation between perception and sensation, based not on sensory physiology but on ecology and cybernetics:

> An entirely different picture of the senses has emerged. For this to happen, we had to suppose that their sole function was not to yield sensations. Instead of mere receptors, that is receivers and transducers of energy, they appear to be systems for exploring, searching, and selecting ambient energy. … [there is] the modification of stimulation by reactions of exteroceptive system, and … the modification of reactions by stimulation of the proprioceptive system. The latter is familiar nowadays under the name of feedback, that is, the neural loops essential for the control of behavior. … The organism has two kinds of feedback,

5 An approach to visual illusions on these lines: "The postulates of stimulus information and stimulus ecology, however, suggest ways in which the various illusions can be, for the first time, classified into types and subtypes of misperception, with the reasons therefore. … illusions will be treated as special cases of perception, not as phenomena which might reveal the laws of the subjective process of perception" (pp. 365–366).

> not one. There are two kinds of action, in fact, one being *exploratory* action and the other *performatory* action. … this new picture of the senses includes attention as part of sensitivity, not as an act of the mind upon the deliverance of the senses.
>
> (pp.366–367)

In appealing to the nowadays familiar term "feedback", Gibson is alluding to the relatively new science of his time of *cybernetics*.[6] It was the cyberneticists, beginning in the early 1940s, who emphasized the importance of feedback loops for the control of behavior. The use of cybernetic concepts for "the neural loops essential to the control of behavior" suggests possible lines of influence, one which might be the famous 1959 paper, "What the Frog's Eye Tells the Frog's Brain" (Lettvin et al. 1959), which also suggests a distinction between perception and sensation.[7]

The idea of describing animal behavior in terms of feedback loops is certainly not new—that was already the central thesis of Rosenblueth, Wiener, and Bigelow in their "Behavior, Purpose, and Teleology" (1943) as well as theorists in cognitive ethology and adjacent sciences. What Gibson appears to add is the idea that there are two distinct feedback loops, not just one. One feedback loop, what he calls "exploratory action", is between actions of the perceptual system and the kinds of information available for detection by perceptual systems (e.g. pricking up ears, converging or focusing with eyes). The other feedback loop, what he calls "performatory action", is between bodily movements and the body-generated information about the posture, orientation, movement, etc. of the body as specified by tactile, visual, vestibular stimulation. Interestingly, Gibson does not appear to say much about how these two loops interact within the animal, including its brain—an oversight that shall be remedied by incorporating into Gibson's account some ideas from Sellars's own use of cybernetics.

I shall conclude this section by underscoring how Gibson understood his synthesis of ecological and cybernetic thinking to undermine all the shared assumptions behind classical theories of perception, both "empiricist" and "rationalist". These shared assumptions include the following: (1) the sense

6 For contemporary readers unfamiliar with the history of cybernetics, I recommend Dupuy (2009) and Kline (2015).

7 "The operations thus have much more the flavor of perception than of sensation, if that distinction has any meaning now. That is to say that the language in which they are best described is the language of complex abstractions from the visual image. We have been tempted, for example, to call the convexity detectors 'bug perceivers.' Such a fiber (operation 2) responds best when a dark object, smaller than receptive field, enters that field, stops, and moves about intermittently thereafter. The response is not affected if the lighting changes or if the background (say a picture of grass and flowers) is moving, and is not there if only the background, moving or still, is in the field. Could one better describe a system for detecting an accessible bug?" (p. 1951).

organs transmit signals from the world; (2) perception is an internal operation or processing of this sensory input; (3) satisfactory perception is a representation of the world that corresponds to the world; (4) perception of the world is separate from bodily awareness because of the different kinds of sense organs involved. (p. 371). Gibson is quite clear about how he is departing from the covert mentalism that has been smuggled into mechanistic behaviorism: "Whereas all the classical theories are based on neural *inputs* and constructive operations on these inputs (sensations) the new theory is based on neural *loops* and their hypothetical capacity to resonate to invariants over time (information)" (p. 373). We can put Gibson's point now as follows: a scientific metaphysics of perception does not postulate mental operations performed over the passive deliverance of the senses, but rather a "resonance" between neural cybernetics and patterns of environmental information—patterns unfold over time and that can be detected by mobile sentient animal. Yet this counts as critical direct realism by Sellars's criteria: we directly perceive physical objects, the directness of this perception is *epistemic* rather than *causal*, and direct perception is causally mediated. The crucial difference is that Gibson proposes a circular causal loop (or rather two such loops) as a scientific explanation of direct perception, rather than a linear causal process whereby objects cause sensations that elicit conceptual responses.[8]

3.4 Sellars's embodied embedded neurocognitivism

I suggested that Sellars should have welcomed Gibson's critique of sensation-based theories of perception and accepted, or at least taken great interest in, Gibson's alternative. There are two reasons for this. The first is that Sellars shared with Gibson a general commitment to both direct realism in perception and to scientific realism in philosophy of science. It was precisely on that basis that Sellars accepted the reality of sensations: because positing sensations as causally mediating perception was taken to be a good scientific explanation of perception. If ecological psychology can offer a better scientific explanation of perception and misperception, then Sellars ought to accept it, given his general philosophical commitments.

The second reason why Sellars should have taken an interest in Gibson's theory of perception is that Sellars was also greatly influenced by cybernetics and took seriously the role of the environment in sustaining neural feedback loops as necessary for cognition.[9] This dimension of Sellars's thought has usually been overlooked because of how Sellars presented it. I am referring

8 On "the passive deliverance of the senses" as "the fourth dogma of empiricism," see O'Donovan-Anderson (1997). One may think that in rejecting 'the passive deliverance of the senses', Gibson thereby evades what Sellars calls 'the Myth of the Given'. Whether or not he does so is beyond the scope of this essay, but for an argument that Gibson does succeed in avoiding the Myth of the Given, see Wilkinson and Chemero (2025).

9 For the importance of cybernetics in Sellars's philosophy of mind, see Sachs (2018), Sachs (2022), and Huebner (2018).

to his somewhat difficult conception of what he called "picturing". Though picturing has been either ignored or criticized by most 20th-century Sellarsians, more recent work is reviving the centrality of this concept to Sellars's scientific metaphysics of mind. Here, I shall argue that correctly understood, picturing is a theory of *necessarily embedded and embodied neurocognitivism*. (In what follows, I shall call this "2E neurocognitivism".[10])

As I understand it "necessarily embedded and embodied neurocognitivism" involves the following commitments and caveats:

1 An explanation is neurocognitivist if a cognitive function is explained in terms of how those functions are implemented by neural structures.
   a Caveat 1: The structure-function relation can be one-many, many-many, or many-one. It does not have to be one-to-one.
   b Caveat 2: This does not entail that *all* cognitive functions have a neural implementation or that *every aspect* of a cognitive function has a neural implementation.

2 At least, some cognitive operations consist of computations performed over representations.
   a Caveat 1: computations can be digital, analog, or perhaps neither.
   b Caveat 2: representations can be symbols, indices, or icons.

3 Cognitive functions are assigned to neural structures on the basis of how those functional structures contribute to the realization of an organism's goals in its environments.
   a Caveat 1: this account is based on a goal-constitutive rather than etiological account of function.
   b Caveat 2: the organism-environment relationship is the necessary context in which cognitive functions can be identified and assigned to (classes of) neural structures.

I shall argue that all of these commitments and caveats are at work in Sellars's account of picturing. To do so, I shall turn to his example: a robot that has been designed to explore an environment.

Suppose a robot has been designed to explore exoplanets with terrains and/or atmospheres inhospitable to its creators. Consequently, it is equipped with a variety of scanners that allow it to detect regularities and irregularities in its

10 The account of neurocognitivsm here is largely indebted to Piccinini (2020, 2022). What of the other two Es – extendedness and enaction? I set aside extendedness because I agree with Rowlands (2010) and Gallagher (2017) that the extended mind is premised upon functionalism, rather than being a serious alternative to it. I set aside enaction because I regard enactivism as a philosophy of nature rather than a proposal for a non-functionalist or non-cognitivist scientific approach to mind; see Gallagher (2018), see also Meyer and Brancazio (2022) and Heras-Escribano (2023).

environment. As it receives information from its transducers—information about electromagnetic radiation, seismic activity, thermal gradients, etc.—those signals are transformed into a single format that allows information from different formats be collated. The internal states of the robot therefore stand in highly abstract isomorphic relations with the features of the environment with which they reliably covary. To be effective, the robot must also correlate this information with information generated by its internal sensors that convey the position, direction, and speed of its own body relative to its environment. As the robot navigates various environments, it thereby constructs an increasingly reliable and accurate map of its environment that can guide further exploration.[11]

What Sellars would stress about this robot is that *it does not matter* if the information in its feedback and feed-forward-driven cycles of transducers, processors, and effectors is represented as terms, predicates, and sentences that we associate with a natural language:

> while we can talk about the items on the tape as 'sentences' and assimilate them by analogy (and with hesitation) to the logical order, we can also consider the states of the robot in mechanical and electronic terms; and the point I wish to make is that in these terms it makes perfectly good sense to say that as the robot moves around the world the record on the tape contains an ever more complete and perfect map of its environment. In other words, the robot comes to contain an increasingly adequate and detailed *picture* of its environment in a sense of 'picture' which is to be explicated in terms of the logic of relations.
>
> (Sellars 1963b, p. 53)

We do not need to attribute to the robot anything like a language—not even a "Language of Thought"—in order to appreciate the functional role of its representational states. None of its states are sentences and the transitions between states are not inferences. It does not have propositional attitudes and its internal states lack intensions. Its representations reliably *covary* with the represented features of the environment that it can reliably detect, some of which it can also manipulate. Regardless, it does have internal states with representational functions: it *pictures* its environments.

In calling this relation "picturing", Sellars acknowledges a debt to Wittgenstein's *Tractatus*. This might give pause to those who think that the Wittgenstein of the *Tractatus* has been wholly superseded by the Wittgenstein of *Philosophical Investigations* and *On Certainty*. However, it must be emphasized that Wittgenstein's idea of picturing is based on his reading of Heinrich

11 This paragraph is my rewriting of a hypothetical robot in Sellars (1963b), though I have also drawn upon "After Meaning" in *Naturalism and Ontology* (1979) and "Mental Events" (1981).

Hertz. Giving up on, or radicalizing, the Fregean conception of sense would not, by itself, remove the need for what Wittgenstein inherits from Hertz. Once this point is appreciated, we can see that Sellars could very well have good reasons for his conviction that the philosophy of language developed by the late Wittgenstein does not obviate the need for an account of picturing. As a committed naturalistic, Sellars further realizes that picturing must be naturalized. Hence, the picturing items cannot be described as propositions, as they were in the Tractatus; they must be physical items that stand in causal relations to the pictured items. Our thoughts *in rerum natura* are states of a complex representational system that is not itself essentially linguistic in structure.[12]

At this point, one may wonder whether I have transgressed the spirit of ecological psychology by introducing Sellars's robot with its internal representations. Gibson was not unaware of the cognitive revolution, but he regarded it as seriously confused. The criticism is worth noting in detail:

> *Information*, as the term is used in this book (but not in other books), refers to specification of the observer's environment, not to specifications of the observer's receptors or sense organs. … The term *information* cannot have its familiar dictionary meaning of *knowledge communicated to a receiver*. … The world does not speak to the observer. … The assumption that information can be transmitted and the assumption that it can be stored are appropriate for the theory of communication but not for the theory of perception. … [in Shannon's theory] a sender and receiver, a channel, and a finite number of possible signals were assumed. … although psychologists promptly tried to apply it to the senses and neuropsychologists began thinking of nerve impulses in terms of bits and the brain in terms of a computer, the applications did not work. … The information for perception, unhappily, cannot be defined and measured as Claude Shannon's information can be.
>
> (Gibson 2015, pp. 231–232)

Put otherwise, information *as defined by communication theory* is a wholly separate concept from the information *as defined by an ecological approach to the psychology of perception*. Only by conflating these two wholly different concepts do we arrive at the misbegotten (by Gibsonian lights) idea that the brain is a computer, i.e. that it receives information from the world (via its senses) and that it processes that information. As I read Gibson, the root

12 Put otherwise, the early Wittgenstein does conflate content and covariation. Sellars's key breakthrough is the realization that a socially normative account of content (as developed by the late Wittgenstein) does not eliminate the need for an account of covariation as well, but it does require that covariation be described as contentless. Where Sellars differs from radical enactivism is that he has a positive account of representations; see Christias (2024) for a contemporary defense.

of the problem is that the relation between organisms and their environments is nothing at all like the relation between senders and receivers: the perception of the world is not a communication with it.

If this critique were granted, would it undermine my attempt to bring Sellars into productive conversation with Gibson? I do not think so. The critique is not just that the cognitive revolution rests on a misapplication of communication theory, but that what is communicated is *contentful*: it is senders and receivers who are *speaking* to each other. Here, it is crucial to distinguish, as Hutto and Myin do (2013), between *content* and *covariation*:

> [t]here is consensus that s's being F 'carries information about' t's being H if the occurrence of these states of affairs covary lawfully, or reliably enough. But anything that deserves to be called content has special properties—e.g., truth, reference, implication—that make it logically distinct from, and not reducible to, mere covariance relations holding between states of affairs. … In yet other words, it is important to distinguish the notion of information-as-covariance from its richer cousin semantic or intentional information—the kind of contentful information (the message) that some communications convey. … covariation in and of itself neither suffices for nor otherwise constitutes or confers content, where content minimally requires the existence of truth-bearing properties.
>
> (Hutto and Myin 2013, pp. 66–67)

The distinction between information-as-covariation and information-as-content bears directly on both Gibson's critique of the cognitive revolution and Sellars's account of picturing. For what Gibson is complaining about is precisely the conflation of content and covariation: a mobile sentient animal achieves resonance (covariation) with ambient environmental information, and that is nothing at all like communication (contentful messages transmitted between senders and receivers).[13] But Sellars's account of picturing is also an account of information-as-covariation, and not an account of content—precisely because by Sellars's own lights of what counts as content, content is governed by rules in a language game. But the solitary robot exploring exoplanets is not playing a language game with its environment, nor is it playing one by itself. Its representations are only covariations, not content. To use Hutto and Myin's term, the Sellarsian robot is an example of a basic mind, one without content. To use Sellars's own distinction, it *pictures* the environment but it does not *signify* it.

13 This does not by itself show that Gibson is wholly innocent of the covariation-content conflation. But see Segundo-Ortin, Heras-Escribano, and Raja (2019) for an explanation as to why ecological psychology does not conflate covariation and content.

Previously, I alluded to Sellars's distinction between "sense" and "the intellect". In that context, I noted that Sellars denies that sense has genuine intentionality; on Sellars's account, all genuine intentionality, cognitive intentionality, belongs to the intellect alone. As a consequence, Sellars construes picturing as a scientific theory of the intellect:

> *qua* belonging to the real order the intellect *pictures* the world, i.e. is related to the real order as the electronic state of the anthropoid robot is related. … But what sort of thing is the intellect as belonging to the real order? I submit that as belonging to the real order it is the central nervous system, and that recent cybernetic theory throws light on the way in which cerebral patterns and dispositions picture the world. … there is no absurdity in the idea that what we know *directly* as *thoughts* in terms of *analogical* concepts may *in propria persona* be neurophysiological states.
>
> (Sellars 1963b, p. 59)

Given Sellars's assumption that only the intellect displays genuine intentionality, it follows that picturing is the intellect *in rerum natura*. But that assumption depends on Sellars's conflation between *sense* as a capacity and *sensations* as states produced in the exercise of that capacity. Although Sellars was right to say that *sensations* lack intentionality, it does not follow that *sensing* lacks intentionality.

Instead, by drawing upon Gibson's account of the senses as perceptual systems, in which sensing necessarily involves sensitivity to environmental information, we can synthesize Sellars's embodied-and-embedded neurocognitivism with Gibson's ecological cybernetics. One result is that it is intentionality as such, whether sensitive or cognitive, which pictures the environment.[14] Recall that Gibson does not articulate how exploratory actions and performatory actions are coordinated within the animal or by the animal's brain. Sellars, by embedding a more computational version of cybernetics in his account of a hypothetical robot, allows us to venture the following suggestion: the function of neural computations performed over neural representations is to coordinate the feedback loop running from exploratory actions to exteroceptive stimulation with the feedback loop running from performatory action to proprioceptive stimulation. And since neural representations thus construed are extensionally specified covariations and not intensionally specifiable contents,

14 This is nevertheless compatible with Sellars's argument that cognitive intentionality cannot be a relation between mind and world, even if sensitive intentionality is; I develop a closely related view in Sachs 2014. What matters is keeping distinct the kinds of relationality involved in covariation and in content: it is covariation, not content, that is a relation between mind and world. One could nevertheless accept that sensing pictures the world by virtue of being what Hutto and Myin (2017) call "Ur-intentionality" (104–114).

there does not seem to be anything in the Sellarsian approach to neural representations that should worry ecological psychologists. Sellars and Gibson are therefore both compatible (neither contradicts the other) and complementary (each contributes what the other lacks but needs).

3.5 Conclusion

Sellars and Gibson are often seen as late representatives of two opposing philosophical-psychological movements: critical realism and new realism. I have argued that Sellars's own more nuanced position, what he calls "critical direct realism", could describe Gibson as well. And while they clearly disagree about the necessity of sensations for perception, it would be fully consistent with Sellars's larger philosophical commitments if he were to replace a sensation-based theory of perception with an information-based theory. Doing so entails revising his account of picturing from being an account of the intellect to being an account of intentionality *tout court*, including the active intentionality of the senses as perceptual systems.

The late 20th and early 21st centuries in philosophy of cognitive science were marked by a debate between the defenders of mainstream cognitive science, or "cognitivism", and the rise of 4E cognitive science. Sellars is often looked to as a precursor of cognitivism due to his influence on Dennett, Churchland, Millikan, and (to a lesser extent) Fodor. Likewise, Gibson is often looked to as an influence on 4E cognitive science, notwithstanding the debates between ecological psychologists and enactivists. But I have argued that there is room for a more nuanced position that takes seriously what both Sellars and Gibson have to offer. Correcting Sellars with Gibson removes the need for a sensation-based theory of perception, which is the basis for the fourth dogma of empiricism (cf. O'Donovan-Anderson 1997); correcting Gibson with Sellars shows that a computational account of neural contributions to cognition addresses the question as to how the exploratory and performatory loops causally interact within the organism.[15] Importantly, this can be done while also abiding by the covariation/content distinction, which both Gibson and Sellars accept. It is my hope that this provisional synthesis will prove to be yet one more step along the route that will lead to a satisfactory scientific metaphysics of mind-in-the-world.[16]

15 The account sketched here also suggests that one could incorporate neuroscience into ecological psychology without giving up on the concept of neural representations altogether; but see Favela (2024) for a non-representational ecological neuroscience.

16 A previous version of this paper was presented at "Analytic Philosophy and E-Cognition" held at the University of Granada, June 25–27 2024. I am grateful to the organizers and participants for their encouragement and insightful criticisms.

References

Christias, D. (2024) "Contentless Representationalism? A Neglected Option Between Radical Enactivist and Predictive Processing Accounts of Representation". *Mind and Machines* 34 (1): 1–21.

Dupuy, J. (2009) *On the Origins of Cognitive Science: The Mechanization of the Mind*, trans M. B. DeBevoise. The MIT Press.

Egan, F. (2025) *Deflating Mental Representation*. The MIT Press.

Favela, L. (2024) *The Ecological Brain: Unifying the Sciences of Brain, Body, and Environment*. Routledge.

Gallagher, S. (2017) *Enactivist Interventions: Rethinking the Mind*. Oxford University Press.

Gallagher, S. (2018) "Rethinking Nature: Phenomenology and a Non-reductionist Cognitive Science". *Australasian Philosophical Review* 2 (2): 125–137.

Gibson, J. (1982a) "The Concept of the Stimulus in Psychology". In E. Reed and R. Jones (Eds.) *Reasons for Realism: Selected Essays of James J. Gibson* (pp. 333–349). Routledge.

Gibson, J. (1982b) "The Useful Dimensions of Sensitivity". In E. Reed and R. Jones (Eds.) *Reasons for Realism: Selected Essays of James J. Gibson* (pp. 350–373). Routledge.

Gibson, J. (2015) *The Ecological Approach to Visual Perception*. Taylor and Francis.

Heras-Escribano, M. (2023) "Ecological Psychology, Enaction, and the Quest for an Embodied and Situated Cognitive Science". In M.- O. Casper and G. Artese (Eds.) *Situated Cognition Research* (pp. 83–102). Springer.

Huebner, B. (2018) "Picturing, Signifying, and Attending". *Belgrade Philosophical Annual* 1 (31): 7–40.

Hutto, D. and E. Myin (2013) *Radicalizing Enactivism: Basic Minds without Content*. The MIT Press.

Hutto, D. and E. Myin (2017) *Evolving Enactivism: Basic Minds Meet Content*. The MIT Press.

Kline, R. (2015) *The Cybernetics Moment: Or Why We Call Our Age the Information Age*. Johns Hopkins University Press.

Lettvin, J., H. R. Maturana, and W. S. McCulloch (1959) "What the Frog's Eye Tells the Frog's Brain". *Proceedings of the IRE* 47 (11): 1940–1951.

Levine, S. (2007) "Sellars's Critical Direct Realism". *International Journal of Philosophical Studies* 15 (1): 53–76.

Meyer, R. and N. Brancazio (2022) "Putting Down the Revolt: Enactivism as a Philosophy of Nature". *Frontiers in Psychology* 13: 948733.

Millikan, R. (2000) *On Clear and Confused Ideas: An Essay About Substance Concepts*. Cambridge University Press.

Millikan, R. (2004) *Varieties of Meaning*. The MIT Press.

Millikan, R. (2007) *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford University Press.

O'Donovan-Anderson, M. (1997) *Content and Comportment: On Embodiment and the Epistemic Availability of the World*. Rowman and Littlefield.

O'Shea, J. (2024) "Critical Direct Realism? New Realism, Roy Wood Sellars, and Wilfrid Sellars". *Topoi* 43 (1): 135–145.

Piccinini, G. (2020) *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford University Press.

Piccinini, G. (2022) "Situated Neural Representations: Solving the Problems of Content". *Frontiers in Neurorobotics* 16: 846979.
Rosenblueth, A., N. Wiener, and J. Bigelow (1943) "Behavior, Purpose, and Teleology". *Philosophy of Science* 10 (1): 18–24.
Rowlands, M. (2010) *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. Bradford Books.
Sachs, C. (2014) *Intentionality and the Myths of the Given*. Routledge.
Sachs, C. (2018) "In Defense of Picturing: Sellars's Philosophy of Mind and Cognitive Neuroscience". *Phenomenology and the Cognitive Sciences* 18 (4): 669–689.
Sachs, C. (2022) "A Cybernetic Theory of Persons: How and Why Sellars Naturalized Kant". *Philosophical Inquiries* 10 (1): 97–123.
Segundo-Ortin, Heras-Escribano, and Raja (2019) "Ecological Psychology is Radical Enough: A Reply to Radical Enactivists". *Philosophical Psychology* 32 (7): 1001–1023.
Sellars, W. (1954) "Physical Realism". *Philosophy and Phenomenological Research* 15 (1): 13–32.
Sellars, W. (ed.). (1963a) "Phenomenalism". In *Science, Perception, and Reality* (pp. 60–105). Ridgeview Press.
Sellars, W. (ed.). (1963b) "Being and Being Known". In n *Science, Perception, and Reality* (pp. 41–59). Ridgeview Press.
Sellars, W. (1979) *Naturalism and Ontology*. Ridgeview Press.
Sellars, W. (1981) "Mental Events". *Philosophical Studies* 39 (4): 325–345.
Sellars, (1981b) "Lever of Archimedes". *The Monist* 64 (1): 3–36.
Wilkinson, T. and A. Chemero (2025) "Affordances, Phenomenology, Pragmatism, and the Myth of the Given". *Phenomenology and the Cognitive Sciences* 24 (1): 85–101.

4 Ecological psychology and the mirror of nature

Vicente Raja and Anthony Chemero

4.1 Introduction

It should not be a surprise that Jerry Fodor's *The Language of Thought* (1975; henceforth *LoT*) contains several passages critical of the work of Eleanor and James Gibson. Fodor was a leading architect of the installation of cognitivism as the dominant view in philosophy and psychology; the Gibsons were resisters. Fodor's cognitivist position made computational manipulations of representations the center of the cognitive sciences; the Gibsons were anti-representationalists in the pragmatist tradition. What is surprising is that Richard Rorty, himself an avowed anti-representationalist in the pragmatist tradition, takes Fodor's side in his *Philosophy and the Mirror of Nature* (1979; henceforth *PMN*). The purpose of this chapter is to make sense of this. To preview, Rorty did not want his devastating critique of representationalist epistemology to run into resistance from proponents of the newly popular cognitive sciences, which led him to endorse a strong divide between epistemology and psychology. He argues that Fodor's new representationalist cognitive psychology was entirely unrelated to issues in epistemology. The Gibsons were opposed to this division, intending their ecological approach to perception as both an epistemological and a psychological position. We will argue that Rorty made a mistake here.

4.2 Rorty's critique of the Gibsons

Rorty's critical comments on the work of Eleanor and James Gibson occur in Chapter 5 of *PMN* "Epistemology and Empirical Psychology". In the first four chapters, Rorty recounts a series of arguments against the epistemology-focused philosophy that emerges from Cartesian and Kantian conceptions of the Modern era.

> The aim of the book is to undermine the reader's confidence in "the mind" as something about which one should have a "philosophical" view, in "knowledge" as something about which there ought to be a

DOI: 10.4324/9781003649359-6
This chapter has been made available under a CC BY 4.0 license.

> "theory" and which has "foundations," and in "philosophy" as it has been conceived since Kant.
>
> (PMN, 7)

Following thinkers like Wittgenstein, Ryle, and Malcom, Rorty argues that the philosophical conception of the mind as an invisible carrier of representations of the external world is both historically contingent and intellectually bankrupt. Following Quine and Sellars, Rorty argues against foundationalist epistemology, the idea that there is some basic set of mental features that grounds all our knowledge. This is, in our view, very convincing. In Chapter 5, which we find less convincing, Rorty aims to separate the new, representationalist sciences of the mind from the traditional epistemological questions that Rorty rejects. In particular, some philosophers have argued that mental representations proposed by cognitive scientists could answer traditional philosophical questions by serving as the foundation for knowledge (see the essays collected in Kornblith 1985). Against Quine (1969), Rorty argues that epistemology cannot be naturalized as psychology: the former is normative, while the latter is about natural causal processes.

Rorty's first mention of Eleanor and Janes Gibson appears in a footnote to a passage in which he argues that the degree to which cognitive abilities are innate, a question of great importance in the Cartesian-Kantian view of the mind, is in fact philosophically unimportant.

> The notion that it is important to discover what is "innate" comes out in such questions as "Does all knowledge (information is the contemporary term) come through the sense organs or is some knowledge contributed by the mind itself?" (J. J. and E. J. Gibson, "Perceptual Learning: Differentiation or Enrichment?" *Psychological Review* 62 [1955], 32.) Gibson and Gibson take this Kantian question with entire seriousness, and urge that, pace Hume and Helmholtz, perceptual learning is not unconscious inference from memory-traces, but simply "increased sensitivity to the variables of the stimulus array" (p. 40). Yet it is very difficult to imagine how experiment could help decide between this view and, say, Gregory's neo-Helmholtzian interpretation of standard experiments in perceptual learning. Cf. R. L. Gregory, *Eye and Brain* (New York and Toronto, 1966), especially such passages as at p. 11: "The senses do not give us a picture of the world directly; rather they provide evidence for checking hypotheses about what lies before us." See Fodor's discussion of Gibson, which I cite and briefly discuss in section 4.
>
> (*PMN,* 249, note 29)

Rorty thinks that taking "this Kantian question" seriously confuses a modern epistemological question with an empirical question. Empirical methods,

Rorty argues, cannot settle epistemological questions, which are ill posed in any event. This is, we think, ironic because the experiments that the Gibsons describe in this paper do in fact provide evidence concerning the origin of knowledge. In the experiments, the Gibsons showed participants a series of cards with pictures of spirals and asked them to match the other cards to a sample card. They found that, without feedback, participants got better at matching cards to the sample over multiple passes through the deck. This suggests that there is sufficient information in the light reflected off the cards to match them correctly and that no mental enrichment of the information or reinforcement of correct matches was necessary. It also suggests that traditional mind representing reality that Rorty criticizes on philosophical grounds is also empirically otiose.

The second criticism that Rorty makes is aimed specifically at James Gibson's 1966 book *The Senses Considered as Perceptual Systems* (*SCPS* hereafter). He quotes from and endorses Fodor's critique of James Gibson's view of perception.

> Fodor rightly says that if we are to have anything like a "psychological problem of perception" we must have some such model in mind. He criticizes Gibson's suggestion that we could avoid "the problem of how the (presumed) stimulus invariants are detected" by "distinguishing between the stimulus for the sensory transducers (viz., physical energies) and the stimulus for the perceptual organs (viz., abstract invariants)" by saying:
>
> … this way trivialization lies. If one is allowed to use the notion of a stimulus so as to distinguish the input to the retina (light energy) from the input to the optic system (patterns of light energy which exhibit invariances relevant, e.g., to the explanation of perceptual constancies), why not also talk about the stimulus for the whole organism (viz., perceptibles)? Thus, the answer to "How do we perceive bottles?" would go: "It is necessary and sufficient for the perception of a bottle that one detect the presence of the stimulus invariant bottle.
>
> … What this shows, I think, is not that the psychological problem of perception is a muddle, but that stating the problem requires choosing (and motivating) a proprietary vocabulary for the representation of inputs. I have argued that the vocabulary of values of physical parameters is appropriate on the plausible assumption that sensory transducers detect values of physical parameters and that all perceptual knowledge is mediated by the activity of sensory transducers.
>
> (*PMN*, p. 224–5, quoting *LOT*, p. 49n.)

To make sense of this, it is useful to see the context in which these quotes from *LoT* appear. First, though, it is worth noting that in the last line of the quoted passage, Fodor discusses "perceptual knowledge", which suggests

that Fodor does not agree with Rorty about the strict separation of epistemology and psychology. We will see in Section 4.3 that neither do the Gibsons.

In general, Fodor's critique of the Gibsons does not find its place in the realm of epistemology but in the realm of psychological theories. He is skeptical of the separation between stimulus and stimulus information—aka ecological information—at the core of Gibsonian psychology. The Gibsons make this distinction and then define perception with respect to the detection of stimulus information. As Fodor puts it:

> For Gibson, perception involves the detection of invariant (typically relational) properties of impinging stimulus arrays. He apparently assumes that any percept can be identified with such an invariant if only the relevant property is sufficiently abstractly described. But, though Gibson denies that percepts are constructed from conscious sensory data, he does apparently hold that the presence of the relevant stimulus invariant must be inferred from the information output by sensory transducers.
>
> (Fodor 1975, p. 49)

Here, Fodor makes inaccurate claims regarding the Gibsonian position. For instance, the ecological theory would not accept that the detection of the invariants of stimulus information involves any kind of inference. Fodor claims:

> Thus, even for psychologists who think of perceptual distinctions as distinctions between (abstract) stimulus invariants, the problem of how such invariants are themselves detected needs to be solved; and it appears that solving it requires postulating the same sorts of inferences from inputs that empiricist theories assumed. The difference is mainly that contemporary psychologists do not assume that the computations, or the data over which they are defined, must be consciously accessible.
>
> (Fodor 1975, p. 50)

This directly clashes with the general position within ecological approach to perception that our perceptual states are not the product of any inferential process but the outcome of a direct process of information detection. It is true that an account of the way such a detection occurs is important for the ecological approach—and there are several contemporary works on that regard (e.g., Favela 2023; Raja 2018, 2021, 2024)—but whatever the this account is, the Gibsons and the Gibsonians would not be happy with the appeal to any form of inference.

Despite these inaccuracies at the outset, Fodor gets the general idea of Gibsonian psychology—i.e., that perception is based on the detection of

ecological information—mostly correctly. And he goes on and criticizes it by complaining about the unclear character of stimulus information:

> The status of the claim that there are stimulus invariants corresponding to percepts is unclear. On one way of reading it, it would seem to be a necessary truth: Since 'perceive' is a success verb, there must be at least one invariant feature of all situations in which someone perceives a thing to be of type *t*; viz., the presence of a thing of type *t*. On the other hand, it is a very strong empirical claim that, for any type of thing that can be perceived, there exists a set of physical properties such that the detection of those properties is plausibly identified with the perception of a thing of that type. This latter requires that the distinction between things of type *t* and everything else is a physical distinction… The issue is whether there are physical kinds corresponding to perceptual kinds and that, as we have been saying all along, is an empirical issue. My impression of the literature is that the correspondence fails more often than it holds; that perception cannot, in general, be thought of as the categorization of physical invariants, however abstractly such invariants may be described.
>
> (Fodor 1975, p. 48, Note 15)

After this complaint, Fodor introduces the charge of triviality that can be read in the passage cited by Rorty (see above). Overall, Fodor's critique seems to rest on one main point: there is something trivial in the postulation that, at some level of description, there will always be a property of stimulation that informs for some object in the environment and, eventually, there must be a property of stimulation that informs for *all* objects of the environment. This point is, like the one regarding the inferential needs of Gibsonian psychology, somewhat inaccurate. The Gibsons claimed no such thing.

4.3 What the Gibsons actually claim

An interesting aspect of Rorty's critique of the Gibsons is the veiled accusation of Kantianism. When Rorty claims the Gibsons take the Kantian question "with entire seriousness", there is a lurking charge: Rorty really thinks that the Gibsons are taking the Kantian framing *too* seriously, and that they shouldn't. This accusation would have likely been poorly received by the Gibsons, as James Gibson, for instance, has been explicit several times about the dangers of "falling in the arms of Immanuel Kant" (Gibson 1967a, p. 10). But the accusation is also interesting because it could go both ways: the Gibsons could claim Rorty takes the Kantian framework with entire seriousness when he proposes a sharp dichotomy between epistemology and cognitive science, which exactly matches the dichotomy between rational psychology (or epistemology) and experimental psychology Kant proposes in the *Critique of*

Pure Reason (1781). Unlike Rorty's Kantianism, the Gibsons could say, they reject such dichotomy and offer psychological theory that actually informs epistemological issues.

Fun as it is throwing accusations of Kantianism at each other's heads; it does not strike us as especially productive. So, it is better if we don't put the accusation of Kantianism in the mouth of the Gibsons and it would have been better if Rorty did not make the accusation in the first place and had focused on the actual reasons why the Gibsons were rejecting the epistemology-psychology dichotomy. Or, at least, to pay attention to the reasons why they thought the ecological approach to perception and action was indeed relevant for epistemology—and even ontology (see Gibson 1967b).

The main interest of both James and Eleanor Gibson can be read both in terms of psychology and in terms of epistemology. On the one hand, they were mostly interested in how (visual) perception and learning *work*. In this sense, they are providing a psychological theory. A theory that explains how perception (and perceptual learning) happens in humans and animals. This is the context in which James Gibson, for instance, refers to mainstream theories of perception:

> All kinds of metaphors have been suggested to describe the ways in which sensory inputs are processed to yield perceptions. It is supposed that sensation occurs first, perception occurs next, and knowledge occurs last, a progression from the lower to the higher mental processes.
> (Gibson 1979, p. 240)

Even though he refers to knowledge, the focus of this quote is the causal process involved in perception and higher cognitive processes. In the same way, when he postulates his alternative to the mainstream theories of psychology, Gibson proposes a psychological theory with the different components that build up the perceptual process. For instance, he claims:

> To perceive is to be aware of the surfaces of the environment and of oneself in it. The interchange between hidden and unhidden surfaces is essential to this awareness. These are existing surfaces; they are specified at some points of observation. Perceiving gets wider and finer and longer and richer and fuller as the observer explores the environment. The full awareness of surfaces includes their layout, their substances, their events, and their affordances.
> (Gibson 1979, p. 244)

In this quote, we can identify some of the main concepts of the Gibsonian theory of perception: the surfaces of the environment, the occluding edge (i.e., interchange between hidden and unhidden surfaces), specification, exploration, and affordances. There are other places to explain this theory in detail (e.g., Chemero 2009; Heras-Escribano 2019; Segundo-Ortin & Raja

2024; Turvey 2018). For what matter to us, the Gibsonian theory of perception dispenses with the need for mental representations insofar as it regards stimulus information as rich enough to support perception in a pragmatic way: if organisms are able to explore their environment enough, they will find stimulus information (aka ecological information; see Segundo-Ortin et al. 2019) that specifies the environment in the context of their current and subsequent *actions*.

This is a psychological theory and, as such, it is the aim of Fodor's critique. When Fodor makes the charge of triviality, he is attacking the ecological idea of stimulus information as one of the pieces of the psychological explanation of perception. According to him, postulating stimulus (ecological) information, such as invariants, is either a truism or just does not solve the inferential problems of perception it is supposed to solve (Fodor 1975). Beyond whether this is a fair critique or not (and we think it is not very powerful; see Baggs & Raja 2024), it clearly belongs to the realm of the psychological theories and not to the realm of epistemology (see the previous section).

On the other hand, the Gibsons were interested in human/animal awareness of the environment and its development. In this sense, they are well within the epistemological camp: they provide a framework for human/animal knowledge and how it can be. This is explicit, for instance, in Gibson (1979) when critiquing the mainstream (mostly Kantian) theories of perception and knowledge:

> The error lies, it seems to me, in assuming that either innate ideas or acquired ideas must be applied to bare sensory inputs for perceiving to occur. The fallacy is to assume that because inputs convey no knowledge they can somehow be made to yield knowledge by "processing" them. Knowledge of the world must come from somewhere; the debate is over whether it comes from stored knowledge, from innate knowledge, or from reason. But all three doctrines beg the question. Knowledge of the world cannot be explained by supposing that knowledge of the world already exists. All forms of cognitive processing imply cognition so as to account for cognition.
>
> (p. 241)

In this paragraph, there is a clear connection between the psychological theory of perception and epistemological considerations regarding knowledge. Ultimately, the position of the Gibsons is that the ecological theory of perception is the only workable way to understand how different organisms get to be aware of their environment. The mainstream options that need to appeal to some form of prior non-perceptual knowledge to process stimualiton in order for perception to work are, according to them, a dead end (see Warren 2021). Only a theory like the ecological one, which provides a way to have meaningful environmental information for the organism, is able to provide

an account of perceptual knowledege. In this sense, the Gibsons are indeed detailing not only a psychological theory but an epistemological one: they are setting a *norm* regarding what we can talk about as be a secure path to knowledge and what is indeed a dead end. Ecological psychology, therefore, solves the epistemological question.

Now, the kind of knowledge the Gibsons were thinking of when they developed their theory of perception is not of the kind of knowledge Fodor has in mind and, certainly, not of the kind of knowledge that should bother Rorty. What we know about the environment by detecting stimulus information of the kind Gibson (1979) describes and Fodor (1975) criticizes is not a set of categories or discursive elements, but its *affordances*—i.e., the opportunities for action that surround us (Chemero 2009; Heras-Escribano 2019; Segundo-Ortin & Raja 2024). Thus, and *contra* Fodor, the Gibsons never thought one would find an invariant for a bottle. A bottle is a thing that falls under a specific category within some specific contexts—"bottle" in English, "botella" in Spanish. The Gibsons were not thinking about such kind of discursive knowledge when building up their ecological theory of perception. On the contrary, and in a deeply pragmatic spirit, they were targeting action-oriented knowledge. In this context, they would say that there is no invariant for the bottle but there are invariants for the actions we can perform with respect to it: approach it, grab it, drink from it, etc. These are the affordances of the bottle we can *know* according to ecological psychology.

It is quite straightforward to see the influence of pragmatism in the sense of knowledge used by the Gibsons. To know is to know how to act. And perception delivers this kind of knowledge. It seems that Rorty, as a neo-pragmatist himself, should be happy with this move away from "knowledge as representation" to "knowledge as action". However, he agreed with Fodor and not with the Gibsons. In the following, we will consider a few reasons for this fact.

4.4 Why Rorty should have been a Gibsonian

We got to a point in which we have Jerry Fodor criticizing ecological psychology as a psychological theory and Richard Rorty accepting this criticism and extending it to ecological psychology as an epistemological theory. The core of this is that Rorty endorses Fodor's critique of the Gibsons as trivializing the problem of the stimulus for perception. For one thing, Rorty leaves off Fodor's sentence following the material he quoted to the effect that the Gibsons' views have "a curiously Rylean sound" while having spent the prior few chapters endorsing arguments by Ryle. So, the Gibsons' position seems to us to be the one that Rorty has already endorsed. For another, and as we already noted, Fodor misrepresents the Gibsons' position on this. What the Gibsons do in their ecological theory of perception is distinguishing between stimulation and information. Stimulation of modality-specific sensor cells is not necessary for modality-specific perception; stimulation of sensor cells is

not sufficient for perception. When Fodor parodies the Gibsons' views by saying "Thus the answer to 'How do we perceive bottles?' would go: 'It is necessary and sufficient for the perception of a bottle that one detect the presence of the stimulus invariant bottle'", he gets it wrong. From the Gibsons's perspective the answer to the question of how we perceive bottles is that we detect the informational invariant that specifies bottles. And more concretely, we detect the informational invariant that specify the kind of actions the organism can do (i.e., affordances) with the object called "bottle" in English and "botella" in Spanish. Rorty is thus endorsing Fodor's inaccurate account of the Gibsonian theory. He should have, instead, endorsed the Gibsons' nonrepresentational scientific approach to perception which could have buttressed his critiques of epistemology, by trivializing epistemological questions. If perception and knowledge are direct, as the Gibsons have it, there is no need for epistemology: again, ecological psychology solves the epistemological question from experimental psychology.

If Rorty was indeed misled by Fodor's account of Gibsonian psychology that could be reason enough for him taking sides with the philosopher and not the psychologists. However, there might be some other complementary reasons for what, in our opinion, is a Rortyan mistake. It seems clear to us that Rorty, the neo-pragmatist, should have aligned his own views with those of the Gibsons. As we have already noted, they propose a theory of perception that involves a kind of knowledge based on affordances that is completely alien to any form of mirroring of nature. On the contrary, the perceptual knowledge offered by ecological psychology is action-based and pragmatist at its heart. Otherwise, Rorty, an analytic philosopher deeply invested in the linguistic turn (see Rorty 1967), might have struggled to see this point. As for many other philosophers who were victims of the infamous linguistic turn, Rorty reduced experience to language and based all his developments from that point of view. For him, if knowledge was anything, it was discursive knowledge. And, of course, epistemology was only concerned with that kind of knowledge that was intrinsically pernicious—i.e., intrinsically representational. However, James Gibson made a clear-cut distinction between perceptual knowledge and discursive (or linguistic) knowledge:

> Perceiving is the simplest and best kind of knowing. But there are other kinds, of which three were suggested. Knowing by means of instruments extends perceiving into the realm of the very distant and the very small; it also allows of metric knowledge. Knowing by means of language makes knowing *explicit* instead of *tacit*. Language permits descriptions and pools the accumulated observations of our ancestors.
>
> (Gibson 1979, p. 251; emphasis is ours)

Unlike the explicit, linguistic knowledge Rorty focuses on, affordances are the kind of tacit, perceptual knowledge Gibson is pointing out in this

passage. Affordances lie in the same space of knowledge, as *percepts* as William James understood them or those interactions that are ready-to-hand, as Martin Heidegger pointed out. That perceptual knowledge is not explicitly tied to language and is usually nonrepresentational and action-oriented. In other words, it is knowledge in the best pragmatist (or even radical empiricist) tradition. A tradition that Rorty supposedly endorses and that he should have endorsed in the case of Gibsonian psychology. Perhaps his investment in analytic philosophy and the linguistic turn is one of the reasons that precluded him from making the right choice.

4.5 Conclusion

The reader might have been left with a sense of paradox. On the one hand, the Gibsons propose a psychological theory with epistemological consequences. On the other hand, Gibsonian psychology dissolves epistemological questions by virtue of solving it outside the representational framework inherited from Modern philosophy. Thus, the epistemological reach of ecological psychology effectively does away with epistemology. At least with the epistemology Rorty is warning us against: the one that is based upon the Kantian point of view that knowledge is fundamentally discursive (i.e., made of judgments). Ecological psychology, evenly standing on the shoulders of pragmatists (James and Dewey) and phenomenologists (Heidegger and Merleau-Ponty), provides an empirical route to understand and justify knowledge that is independent from (and perhaps prior to) discursive considerations and, therefore, is freed from the representational problems Rorty correctly identified. This is precisely the reason why we think Rorty should have sided with the Gibsons in his discussion of Fodor. By dismantling the distinction between epistemology and psychology (or cognitive science), the Gibsons did not help to legitimate epistemology but to dissolve it. And Rorty should have been happy about it.

Acknowledgments

Vicente Raja is supported by a "Ramón y Cajal" Fellowship (Grant# RYC2023–045124-I) funded by MCIN/AEI/10.13039/501100011033. Anthony Chemero is supported by the Taft Research Center at the University of Cincinnati.

References

Baggs, E., & Raja, V. (2024). The affordance hypothesis. https://doi.org/10.31234/osf.io/xu4wk

Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press.

Favela, L. (2023). *The ecological brain*. Routledge.

Fodor, J. (1975). *The language of thought*. Thomas Crowell Company.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton-Mifflin.

Gibson, J. J. (1967a). Autobiography. In E. S. Reed & R. Jones (Eds.), *Reasons for realism* (pp. 7–22). Lawrence Erlbaum. (Re-published 1982)
Gibson, J. J. (1967b). New reasons for realism. *Synthese*, 17, 162–172.
Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton-Mifflin.
Heras-Escribano, M. (2019). *The philosophy of affordances*. Palgrave.
Kant, I. (1781). *Critique of pure reason*. Cambridge University Press (Edited by P. Guyer & A. Wood in 1997).
Kornblith, H. (ed.). (1985). *Naturalizing epistemology*. MIT Press.
Raja, V. (2018). A theory of resonance: Towards an ecological cognitive architecture. *Minds and Machines*, 28(1), 29–51. https://doi.org/10.1007/s11023-017-9431-8
Raja, V. (2021). Resonance and radical embodiment. *Synthese*, 199(S1), 113–141. https://doi.org/10.1007/s11229-020-02610-6
Raja, V. (2024). The motifs of radical embodied neuroscience. *European Journal of Neuroscience*, 60(5), 47384755. https://doi.org/10.1111/ejn.16434
Rorty, R. (1967). *The linguistic turn*. The University of Chicago Press.
Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton University Press.
Segundo-Ortin, M., Heras-Escribano, M., & Raja, V. (2019). Ecological psychology is radical enough: A reply to radical enactivists. *Philosophical Psychology*, 32(7), 1001–1023. https://doi.org/10.1080/09515089.2019.1668238
Segundo-Ortin, M., & Raja, V. (2024). *Ecological psychology*. Cambridge University Press.
Turvey, M. T. (2018). *Lectures on perception: An ecological perspective*. Routledge.
Warren, W. H. (2021). Information is where you find it: Perception as an ecologically well-posed problem. *I-Perception*, 12(2). https://doi.org/10.1177/20416695211000366

5 Ecological psychology and the behaviorist bogeyman

Miguel Segundo-Ortin and Inés Abalo-Rodríguez

5.1 Introduction

The status of behaviorism is paradoxical: though often portrayed as a relic from the past in contemporary psychology textbooks, it remains a central figure within "the demonology of modern psychology" (Costall, 2004, p. 179; Barrett, 2015). A telling example is the frequent dismissal of radical embodied theories of cognition—this is, theories that challenge the received assumption that cognition is a form of computation—as modern-day revivals of behaviorism.

The typical argument goes as follows: (i) behaviorism collapsed because of its own theoretical and explanatory limitations and was replaced by more capable computational models of the mind; (ii) radical embodied theories are essentially behaviorism in new clothes; therefore, (iii) these theories must suffer from the same flaws. From this perspective, radical embodied approaches appear as a regression to the pre-cognitive revolution era.

For instance, O'Brien and Opie argue that the attempt to explain intelligent behavior without invoking representations and computations "has been tried before, and it does not work. Back then the scheme was known as 'behaviorism' […] but the two ideas are of one piece" (2015, p. 724). Similarly, Adams (2018) rejects all non-representational accounts of cognition, claiming that intelligent behavior results from internal processes involving "representations in the mind of the organism that represent desirable outcomes and possible strategies for achieving that outcome. […] That was what the cognitive revolution was all about—a movement away from behaviorism" (p. 4). Similar claims have been advanced by Aizawa (2015), Block (2001), and Spaulding (2011), to name a few examples.

In this chapter, we focus on one such radical embodied theory: ecological psychology (Chemero, 2009; Turvey, 2019; Blau & Wagman, 2023; Segundo-Ortin & Raja, 2024; Heras-Escribano, 2019). After briefly introducing ecological psychology in Section 5.2, we devote Section 5.3 to assessing whether it can truly be considered a form of behaviorism. However, since behaviorism is not a unified framework, we have limited our comparative analysis to four major forms of behaviorist theory: Watson's methodological

DOI: 10.4324/9781003649359-7
This chapter has been made available under a CC BY 4.0 license.

behaviorism, neobehaviorism, philosophical behaviorism, and radical behaviorism. Our goal is to clarify how ecological psychology aligns with certain behaviorist principles, while also emphasizing its distinctive contributions to the study of perception, action, and other cognitive abilities. Finally, Section 5.4 explores how ecological psychology might benefit from engaging more deeply with contemporary Behavior Analysis (BA).

5.2 Ecological psychology in six core principles

Ecological psychology emerged in the mid-20th century with the pioneering works of James and Eleanor Gibson (J. J. Gibson, 1966; E. J. Gibson, 1969; J. J. Gibson, 1979). Over the past four decades, the research tradition has evolved through extensive empirical and theoretical development (Blau & Wagman, 2023; Segundo-Ortin & Raja, 2024; Turvey, 2019; Chemero, 2009). For instance, ecological psychologists have adopted new methodologies—such as Dynamical Systems Theory and non-linear analysis—and have expanded the framework to previously neglected domains, including motor rehabilitation (Silva et al., 2019), sensory substitution (Lobo et al., 2014), cognitive neuroscience (Raja, 2021), or social cognition (Marsh et al., 2009; Richardson et al., 2007).

In what follows, we summarize ecological psychology in six core ideas:

Perception is based on the detection of information, not the reception of stimuli. Traditionally, theories of perception begin with the assumption that perception begins with sensory states (e.g., retinal images) elicited by stimuli. However, it is well-known that such sensory states are "impoverished" or ambiguous regarding their causal origin. This means that the same state at the observer's sensory receptors (the same retinal image) can arise from different stimuli (a house vs. a picture of a house; a large but distant tree vs. a small but nearby tree, etc.). As a result, theories that begin with sensory data must explain how this ambiguity is resolved. A classical but still popular solution posits that the brain acts as an "intuitive statistician" (J. J. Gibson, 1957, p. 33), combining sensory data with some prior knowledge to infer the most probable cause for the sensory state. However, this creates another challenge: inferential theories must explain the origin of the prior knowledge required for perception.

Skeptical that this explanation can be provided, ecological psychologists distinguish between stimulus and "stimulus information" (or simply information), arguing that perception begins with detecting information, not the reception of stimuli. The distinction is subtle but important: whereas the first depends on the existence of an ambient energy capable of stimulating our sensory organs—namely, light, sound, etc.—, the second is contingent on the existence of a structured ambient energy array the individual can explore and attend selectively.

For illustration, think of what happens when the light emanating from a bulb propagates into a furnished room. As the ambient light is reflected from

the surfaces of the objects, it gives rise to different patterns, textures, gradients, shadows, and so on. This distribution of light constitutes an "ambient optic array". Importantly, because the structuring of the ambient light occurs according to physical laws, "there is *only one* situation that could produce this distribution of light and *only one* distribution of light that could have been produced by this situation" (Blau & Wagman, 2023, p. 38, emphasis original; see also Segundo-Ortin et al., 2019). Put differently, since a particular configuration in the environment (α) lawfully generates a unique ambient optic array (β), the occurrence of β non-ambiguously corresponds to, or "specifies", the presence of α. This structured array of light provides information for perceiving the environment.

Thus, whereas inference-based (computational) approaches assume that perception begins with ambiguous sensory states, ecological psychologists believe that perception is a matter of detecting specific patterns of information that exist in the topology of an ambient energy array.

Perception is direct (non-inferential). If the environment provides specific, non-ambiguous information that can be detected by the individual, the need for internal models and knowledge-based inferences disappears. Individuals do not need to re*construct* the world internally to infer what the world is like; rather, they only need to detect the patterns of information that specify its properties to gain perceptual access to it.

Perception is active. To understand this idea, consider again what happens when you stand still in an illuminated room. This point of view offers a rich array of information about certain features of the environment. However, perceiving the environment requires more than simply being present; you must detect the information by actively scanning the ambient optic array and directing your attention to the different informational variables. Perception is active in the sense that it involves selectively attending to (or "picking up") certain patterns of information over others.

There is another sense in which perception is active. After spending some time in the same position, it becomes clear that a single viewpoint does not reveal all there is to perceive. In contrast, certain features of the room can only be accessed through movement and interaction. By changing your position and engaging with objects, you generate transformations in the ambient optic array, which, in turn, reveal new informational patterns that were previously unavailable. Fully perceiving the environment, therefore, depends on the continuous interplay between perception and motor behavior.

Perception is embodied. As shown in the previous example, perception involves the activity of large "perceptual systems" that encompass the sensory organs, the brain, and the entire body-environment system (J. J. Gibson, 1966). For instance, speaking about visual perception, James Gibson wrote that "one sees the environment not with the eyes but with the eyes-in-the-head-on-the-body-resting-on-the-ground" (1979[2015], p. 195).

Perception is a skill. If perception requires the active exploration of the environment, it follows that someone can be more or less skilled to it—implying that perception requires learning (E. J. Gibson, 1969, 1991). For example, someone trying to perceive the length of a rod through dynamic touch must discover how to manipulate it to reveal the relevant information. Likewise, one can learn that motion parallax aids in depth perception.

Consistent with the rejection of inferential explanations of perception, ecological psychologists also reject the view that perceptual learning involves building better internal models of the world. Instead, they see perceptual learning as a process of increasing differentiation. Accordingly, one improves one's capability to perceive the different features of the world as they increasingly learn how to detect the informational variables that are specific to them.

We perceive affordances. Affordances are opportunities for action that a situation or environmental setting offers to an individual with the required bodily morphology and capabilities. Affordances highlight the embodied character of perception too, and they imply a complementarity between the environment and the perceiver.

Research consistently shows that individuals can reliably perceive whether a surface is *walkable* or a gap is *passable*, even when they cannot estimate physical dimensions like height or width (Higuchi et al., 2011; Thomas & Riley, 2014; Wagman & Stoffregen, 2020). This implies that the perception of affordances takes precedence over, and is independent of, the perception of the physical (individual-independent) properties of the environment.

Affordances can be perceived directly because of the specific relationship between the structured energy and the structuring environment. Therefore, rather than engaging in complex computations to infer what the environment affords, individuals can become aware of what they can do by detecting the structured patterns that convey the relevant information. For instance, a driver can smoothly adjust braking pressure—or decide to steer instead—by detecting optical patterns related to object expansion (Venkatraman et al., 2016).

To these core theoretical principles, we must add a methodological preference for formulating psychological explanations in the form of "lawful regularities between perception and action at the level of the animal-environment interactions—the *ecological scale*" (Raja, 2019b, p. 4). This non-reductive strategy has conducted to the elaboration of many successful explanations for perceptual-based goal-oriented activities in the form of dynamical laws built on differential equations (Warren, 2006; Richardson et al., 2008).

5.3 Which behaviorism, if any?

Having introduced ecological psychology, we now ask whether it qualifies as a form of behaviorism. Given the multiplicity of views and theories that fall under the banner "behaviorism", we won't offer an exhaustive review of the complete behaviorist tradition. Instead, we will focus on the most

representative ones and on those that are explicitly mentioned as inspirations to ecological psychology.

5.3.1 Watson's methodological behaviorism

While figures like Ivan Pavlov or Edward Thorndike helped shape early behaviorism, its official launch is often dated to J. B. Watson's 1913 "Psychology as the behaviorist views it". In his manifesto, Watson repudiated the traditional conception of psychology as the science of the mind and consciousness and dismissed introspection as a research method for being unscientific and unreliable. Instead, he argued that if psychology was to become "a purely objective branch of natural science" (Watson, 1913, p. 158), it had to focus solely on observable behavior. Accordingly, Watson set the primary goal of psychology to predict and control behavior.

Two main reasons drove this shift in focus. First, Watson considered the production of measurable and reproducible results to be a signature mark of natural sciences but believed that introspective psychology failed to meet the standard. Such results, he argued, could only be obtained if psychologists focused on studying behavior and the environmental stimuli that preceded it.

Even though Watson's position is often misrepresented as a denial of consciousness or subjective experience, Barrett (2012) notes that his goal was not to reject experience itself, but to eliminate terms like "mind" or "consciousness" from scientific psychology. Hence, Watson's behaviorism can be understood as "a normative theory about the scientific conduct of psychology" (Graham, 2023, p. 4), grounded in the view that the mind was inaccessible to the empirical sciences.

Second, he argued that behavior is fully explainable in terms of stimulus-response learning and that referring to mental states "adds nothing to what psychology can and should understand about the sources of behavior" (Graham, 2023, p. 4). In line with this view, he emphasized the role of environmental influences in eliciting behavioral responses and claimed that complex personality traits could be explained by the acquisition of conditioned reflexes after years of conditioning. This "radical environmentalism" (Leahey, 2018, p. 347) is at the root of his famous claim that given a dozen healthy infants, he could train them to become "any type of specialist I might select—doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief" (Watson, 1930, p. 104).

Although ecological psychology is rarely identified explicitly as Watsonian behaviorism, O'Brien and Opie (2015, p. 724) suggest that rejecting computational explanations leaves only explanations in terms of "the history of stimulus-response events". If true, this would place ecological psychology squarely within Watsonian behaviorism. We argue, however, that this is mistaken.

For starters, recall that ecological psychologists reject the idea that perception-action coupling can be explained by isolated stimuli. Instead, perceiving affordances requires the existence of structured energy arrays that

convey information. In addition, these arrays can undergo transformations and changes, revealing new information in the form of flux and invariant patterns. As J. J. Gibson explains: "We must learn to conceive an array not as mosaic of stimuli but as a hierarchy of forms within forms, and a flux not as a chain of stimuli but as a hierarchy of sequences within longer sequences" (1960[1982], p. 343).

Because ecological psychology does not treat stimuli as the basis for perceiving affordances, it avoids the objection that "moment-by-moment stimuli are simply too impoverished to account for the richness, variety, and specificity of the behaviors that animals exhibit" (O'Brien & Opie, 2015, p. 724). Ecological psychologists can agree that moment-by-moment stimuli are too impoverished to guide action, while also rejecting the assumption that stimuli are all the perceiver has. In contrast, the ambient energy array contains enough information to guide action.

The ecological conceptualization of perception as active does not sit well with the S-R formula either. As explained before, the information needed for perception is not passively received but actively obtained by the perceiver. In this sense,

> the surroundings with respect to which organisms behave are not to be understood as collections of so many triggering stimuli, and the behaving organisms are not to be understood as collections of so many conditioned and unconditioned reflexes. Rather than simply reacting to triggering stimuli, organisms cleverly exploit the information about their surroundings and their movements to control their actions both retrospectively ("after the fact") and prospectively ("before the fact").
> (Turvey, 2019, p. 376)

Consequently, perception and action cannot be explained in terms of acquired reflexes. In contrast to Watson's behaviorism, ecological psychologists place agency at the center of their theory (Segundo-Ortin, 2020; Segundo-Ortin & Kalis, 2022), and see the organisms as purposive beings, actively seeking affordances to meet goals and needs.

5.3.2 Neobehaviorism

Neobehaviorism[1] appeared in the 1930s as a response to the limitations of Watson's behaviorism. Although neobehaviorists agreed with the emphasis on observable data and rejected introspection, they believed that "mental" constructs could have a role in scientific psychology as long as they were operationally defined and measurable (Feest, 2025).

1 For a lack of a better word, we take the term "Neobehaviorism" from Staddon (2021) to integrate the diversity of theories that, recognizing themselves as behaviorists, departed significantly from methodological and radical behaviorism.

The first name we must mention is Edwin B. Holt. Since the influence of Holt on James Gibson has been extensively documented already (Heft, 2001; Charles, 2011), we will not go into much detail here. Rather, we will highlight three key connections between Holt's neobehaviorism and ecological psychology.

First, Holt was a Jamesian radical empiricist, neutral monist, and functionalist,[2] credited for having created "molar behaviorism" (Holt, 1915) in response to Watson's molecular theory. Whereas "molecular behaviorists" believed that the appropriate strategy to study behavior was to break it into the smallest components, e.g., muscular reflexes caused by triggering stimuli linked by association and ultimately describable in physiological terms, molar behaviorists viewed behavior as a goal-oriented, coordinated totality.

Consistent with this view, Holt rejected the view that behavior (and psychological phenomena in general) could be understood in terms of reflex arcs. Instead, he "took *activity* to be the defining quality of all psychological phenomena" (Heft, 2012, p. 194) and characterized psychology in active terms—as "out-reaching, outgoing, inquiring, and examining, and grasping" (Holt, 1931, p. 41). This is mirrored by Gibsonians when they give the mainstage of their theory of perception to "the analysis of the activities of feeling, tasting, smelling, listening, and looking" (Reed & Jones, 1982, p. 282).

Second, Holt advocated a lawful approach to behavior, in which activity reflects regular relations with aspects of the objective world (1915, p. 370). As explained by Raja (2019b), this fits hand in globe with the ecological emphasis on describing task-specific perception-action couplings in terms of dynamical laws.

Third, Holt introduced the idea of the "recession of the stimulus" (Holt, 1915). As Heft explains, when we break away from a molecular approach and conceive of actions as integrated functional wholes directed at concrete objects or states of affairs, the stimulus "*recedes in significance* as a referent and as a basis for explanation of the action" (2012, p. 200). Interpreted in ecological terms, behavior is controlled in relation to a series of affordances that are relevant to my current goals and intentions, whereby these affordances are specified by complex information instead of simple stimuli.

While ecological psychology aligns well with Holt's molar behaviorism, the connections with other neobehaviorists, such as Edward C. Tolman and Clark L. Hull, are more tenuous. Let's examine both authors in turn.

2 Here, we refer to psychological functionalism, not functionalism as it is used in the philosophy of mind. Psychological functionalism adopts a Darwinian approach to psychology and affirms that psychological or cognitive capacities "result from the active adaptation of the organism to its environment […] cognitive skills are for the control of intentional behavior, so they must be based on a history of interactions with the environment" (Heras-Escribano, 2019, pp. 24–25). In contrast, philosophical functionalism "is the doctrine that what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the way it functions, or the role it plays, in the system of which it is a part" (Levin, 2016, p. 1).

Although Tolman agreed with Watson in believing that much of what is important to human psychology can be understood by studying learning in non-human animals, he did not share the belief that simple stimulus-response learning could account for complex behaviors. In addition, he championed molar behaviorism too.

His major contributions concerned the study of "latent learning"—learning in the absence of explicit reward—, the operationalization of "purposes" or "demands" as functions of behavior, and the introduction of "cognitive postulations", internal representations of possible outcomes of actions (Tolman, 1932).

A famous example of the latter is cognitive maps. Cognitive maps were introduced to explain the "insightful" decisions rats made when navigating mazes, including their ability to find alternative routes and shortcuts when their preferred routes were blocked (Tolman & Honzik, 1930). Tolman argued that these and other similar behaviors could only be explained if we assumed that, after being allowed to explore the maze, "something like a field map of the environment gets stablished in the rat's brain" (1948, p. 142). Accordingly, "it is this tentative map, indicating routes and paths and environmental relationships, which finally determines what responses, if any, the animal will finally release" (1948, p. 192), instead of a sum of previously learned reflexes.

Although he offered little detail on how such maps were represented and used in the brain, the idea inaugurated a long tradition of research that aimed to explain behavior by positing intervening (mental) variables, or representations, between stimulus and response. This justifies that Tolman usually features "somewhere in between Watsonian behaviorism and what is now called cognitive psychology" (Staddon, 2021, p. 31).

In contrast, it is illuminating to read what J. J. Gibson wrote about the strategy of supplementing the stimulus-response formula with intervening variables:

> we [James and Eleanor Gibson] converge in the developing belief that the weakness of the stimulus-response formula in American psychology lies on the side of the stimulus, not on that of the response […] We have no patience with the attempt to patch up the S-R formula with hypothesis of mediation. In behavior theory as well as in psychophysics you either find causal relations or you do not.
>
> (1967[1982], p. 12)

As this brief quote indicates, the Gibsons were very critical of the explanatory strategies in psychology that assume that brains must compensate for what it is not in the environment. Instead, they pursued an explanatory strategy that looked for causal explanations not between stimulus and response, but between stimulus information and goal-oriented activities.

Moreover, James Gibson (1958[1982], 1977[1982]) criticized Tolman's cognitive maps explicitly, and believed that it was possible to offer a more parsimonious explanation of his empirical findings if we paid attention to higher-order patterns of stimulation that rats discovered while exploring the maze: "Place learning is a primitive kind of cognition, akin to perception, that involves visuo-motor activities, opaque geometry, reversible occlusion, the cluttered environment, and a whole new set of problems for psychology" (1977[1982], p. 293). Echoing this view, Warren (1998, 2006, 2019) and Heft (1983, 2013) have advanced ecological explanations for perceptually controlled navigation and way-finding that do not necessitate cognitive maps.

Like Tolman, Hull sympathized with Watson's attacks on introspection too. However, his interest in Gestalt psychology and a visit to Kurt Koffka convinced him that Watson's behaviorism needed refinement (1952b, p. 154). During his career, Hull pursued a formalist, "axiomatic" psychology, proposing a mathematical, hypothetico-deductive approach to the acquisition and forging of hierarchies of habits (including reasoning habits) (Feest, 2025, pp. 58–59). His main theoretical goal was the formulation of behavioral laws that could help us predict how organisms would learn to behave given particular classes of internal and external stimuli, and in the context of particular needs or "drives" (Hull, 1952a). His emphasis on habits was rooted in his mechanistic view of the mind. As Leahey explains, Hull "sought to explain purpose and cognition as the result of mindless mechanical processes describable in logico-mathematical equations" (2018, p. 357).

Although ecological psychologists share an interest in lawful behavioral explanations, they are critical of the view of cognitive systems as machine-like (Reed, 1996; Turvey, 2019). Neither perception is elicited by stimuli, nor affordances trigger behavioral responses. Instead, affordances are opportunities for action that individuals seize depending on their needs and intentions.

Finally, even though Hull had a significant influence on Eleanor Gibson during her formative years—James Gibson wrote that "[d]own deep she [Eleanor] is a Hullian, as I am a Holtian" (1967[1982], p. 12)—she is nonetheless clear that she was interested in understanding Hull's psychology, particularly his ideas on generalization and differentiation, in the context of American functionalism (E. J. Gibson, 1991, p. 4; 2001, p. 26), rather than adopting his mechanistic commitments.

5.3.3 Philosophical behaviorism

Recently, Schlosser (2020) has suggested that ecological psychology is "in line with philosophical behaviorism" (p. 278). His justification for this claim boils down to the fact both propose accounts of cognition that "do not require the ascription of mental representations" (p. 278). Following his lead, this section examines whether ecological psychology can truly be classified as a form of philosophical behaviorism.

In a nutshell, philosophical behaviorism is a theory about the meaning of mental concepts. According to it,

> when we attribute a belief, for example, to someone, we are not saying that he or she is in a particular internal state or condition. Instead, we are characterizing the person in terms of what he or she might do in particular situations or environmental interactions.
>
> (Graham, 2023, p. 5)

Mental terms like "belief", "desire", "intention", and the like refer to (families of) behavioral dispositions, not to specific states in the subjects' brains or minds.

Philosophical behaviorism is commonly associated with authors like Carl Hempel (1935), Gilbert Ryle (1949[2000]) or Ludwig Wittgenstein (1953), although it is debatable whether Ryle or Wittgenstein really endorsed philosophical behaviorism—at least in the crude, textbook version presented above (Leahey, 2018, p. 370).

For instance, Ryle is famous for writing extensively about the "grammar" or "logical behavior" or mental terms, as well as for criticizing the "the dogma of the Ghost in the Machine", which he identifies with the belief that our behaviors can be causally explained in terms of a series of mental events, which are private and "live on a floor of existence defined as being outside the causal system to which bodies belong" (1949[2000], p. 65). The view that mind and body are fundamentally different but somehow interact creates both metaphysical issues (the mind-body problem) and epistemic issues (the problem of other minds). One way to solve these issues is to offer a dispositional analysis of mental terms, arguing that statements containing such terms "can be translated, without loss of meaning, into subjunctive conditionals about what the individual will do in various circumstances" (Tanney, 2022, pp. 24–25).

As suggested before, the reading of Ryle as a philosophical behaviorist, albeit commonplace, has not gone undisputed (e.g., Heras-Escribano & Pinedo, 2014). We won't address this topic here. Instead, our goal is to determine whether Schlosser is right when saying that there is a continuity between philosophical behaviorism and ecological psychology.

We hold that this is not the case and that, in fact, the two projects differ significantly. To begin with, ecological psychologists do not generally subscribe to the dispositional analysis of mental (or folk-psychological) terms. Instead, they favor explanations of (at least some) cognitive capacities in terms of dynamical laws that connect the perception of affordances with the achievement of specific tasks. In the context of these dynamical explanations, representations are not rejected for being logically or metaphysically problematic, but for being explanatorily redundant (Chemero, 2009, p. 77).

Second, while logical behaviorism is concerned with how we use mental terms, ecological psychologists are concerned with offering a robust and empirically informed alternative to computational explanatory models of perception, motor control, and cognition. Reducing the research tradition of

ecological psychology to the mere rejection of representational explanations, as Schlosser (2020) does, is a simplification.

This said, we would like to remember something Dennett notes in his preface to *The Concept of Mind*. According to him, many emerging topics and ideas in contemporary cognitive science, including "embodied and 'situated' cognition; your mind is not your brain; skill is not represented; intelligence without representation" (1949[2000], p. xii), were already anticipated by Ryle. If Dennett is right, ecological psychology might seem compatible with Rylean philosophy, independently of whether it is behaviorist or not. We invite researchers to dig into this hypothesis.

5.3.4 Radical behaviorism

> "Gibson did for perception what Skinner did for animal learning: he handicapped a generation of workers by his blinkered and oversimplified approach" (Sutherland, 1989, p. 175).

Starting in the 1940s, Burrhus Frederic Skinner developed radical behaviorism, thus laying the groundwork for Behavior Analysis (BA).[3] Unlike other forms of behaviorism, which he believed retained traces of Cartesian dualism, Skinner sought to establish psychology as an autonomous science by making behavior itself the proper object of study, not a proxy for something else.

Contrary to common misconceptions, radical behaviorism does not deny or disregard the existence of the mental. What it rejects is the idea that the mind is fundamentally different from behavior and that it functions as an internal entity that causes action (Barrett, 2012). For Skinner, mental activity—such as thinking—is a form of behavior too, the difference being that while bodily actions are publicly observable, mental behaviors are accessible from a first-person perspective only (Skinner, 1945). Crucially, if the "mind" refers to a kind of behavior (private behavior), then, like all behavior, is shaped by context, contingencies, and learning too (Freixa i Baqué, 2003). However, not all behavior-analytic interpretations share this view; some authors, such as Baum (2011), have criticized Skinner's treatment of private events as functionally equivalent to public ones, arguing that such an equivalence risks introducing a subtle form of mentalism that undermines the naturalistic foundation of radical behaviorism.

As a result, the mental does not hold explanatory primacy over the behavioral, nor are internal constructs—such as mental representations—needed to explain action (Baum, 2017; Costall, 2004; Malone, 2009; Uttal, 1999). Private behaviors may sometimes precede public ones, but they do not cause the latter. Instead, both are products of the functional relationships that are established between an individual and the surrounding environment.

3 Radical behaviorism is the philosophical stance that considers all behavior, including private events, as valid subjects of scientific inquiry. Behavior Analysis is the empirical discipline applying these philosophical principles in both research and practice.

A key aspect of Skinner's proposal is that individuals are constantly engaged with their environment. Behavior, including private or mental behavior, is the product of this engagement, and it is always subject to modification through contingencies and learning processes (e.g., operant conditioning and classical conditioning).

Skinner developed his theory in close conjunction with an extensive empirical program, consisting of the systematic analysis of behavior (Skinner, 1938, 1953). Among his most important empirical contributions is the development of operant conditioning (Skinner, 1938, 1953). Unlike Pavlov's classical conditioning, operant conditioning is grounded in the principle that behavior is shaped and maintained by its consequences. Accordingly, behaviors increase when followed by reinforcement and decrease when followed by punishment. In this context, "contingency" refers to the specific relationship between a behavior and its consequence under defined conditions. A contingency is typically structured as a three-term relationship: (1) an antecedent—the environmental condition or stimulus present before the behavior, aka the "discriminative stimulus" (SD); (2) a behavior; and (3) a consequence—the outcome that follows the behavior, such as reinforcement or punishment. Thus, "the part of the environment called antecedent stimulus acquires significance as an opportunity to behave if and only if it is correlated with the consequent stimulus (e.g., a reinforcer)" (Covarrubias et al., 2017, p. 233).

The introduction of the notion of selection by consequences brings with it the idea that behavior, including spontaneous behavior, is governed by the effects it produces (Skinner, 1938, 1984). This perspective bears a conceptual resemblance to Darwinian logic: just as traits are selected for their adaptive value at the phylogenetic level, behaviors are selected at the ontogenetic level based on their reinforcing effects (Skinner, 1984).

Other central contributions in Skinner's work include his systematic investigation of reinforcement schedules, which revealed how different patterns of reinforcement delivery (e.g., fixed ratio, variable ratio, fixed interval, variable interval) influence the rate and persistence of behavior (Skinner & Ferster, 1957). He also advanced the empirical study of complex behavior through the principles of shaping and successive approximations, demonstrating that elaborate behavioral repertoires can be constructed by reinforcing successive steps toward the target response (Skinner, 1938, 1951, 1953, 1957). Although much of this research was initially conducted with non-human animals, Skinner emphasized the relevance of these findings for understanding and modifying human behavior, particularly in clinical settings. His work played a key role in the development of behavioral therapy techniques and interventions such as token economies. Additionally, he provided a functional, behavior-analytic framework for understanding language (Skinner, 1957), which has since been refined and extended within contemporary BA.

To sum up, Skinner's radical behaviorism defines behavior, both private and public, as inherently relational, inseparably linked to its context. In

addition, it conceives of psychology as the science that studies the relationships between environmental stimuli and individuals, with the aim of identifying the learning processes through which these relationships are established.

Claims that ecological psychology is linked to radical behaviorism are not difficult to find. On the positive side, Chemero (2009) situates both research traditions as part of the same intellectual lineage, descending from what he calls "American naturalism". In contrast, critics like Weimer (2024) argue that both frameworks disregard the contribution of the nervous system to our cognitive faculties, thus treating the organism as "an opaque black box" (p. 143) and failing "for exactly the same reasons" (p. 138).

This claim, though repeated, is plainly false on both scores. In fact, Skinner explicitly denied the accusation of black-boxing the organism (1976, p. 233) and wrote the following about "the promise of [neuro]physiology":

> New instruments and methods will continue to be devised, and we shall eventually know much more about the kinds of physiological processes, chemical or electrical, which take place when a person behaves. […] [The neurophysiologist] will be able to show how an organism is changed when exposed to contingencies of reinforcement and why the changed organism then behaves in a different way, possibly at a much later day. What he discovers cannot invalidate the laws of a science of behavior, but it will make the picture of human action more nearly complete.
>
> (pp. 236–237)

Likewise, researchers in the ecological tradition have been studying the role of the nervous system in the perception and actualization of affordances for more than a decade (van der Weel & van der Meer, 2009; van der Meer et al., 2012; Anderson, 2014; Raja, 2019a). Neither of these frameworks treats the organism as a black box; what they reject, instead, is that the laws of psychology can be eventually subsumed or replaced by the laws of neurophysiology.

This said, and despite the surprisingly few mentions of Skinnerian behaviorism in the ecological literature, there are important points of convergence.[4] To begin with, they equally advance a scientific psychology situated at the personal level, thereby legitimizing psychology as an autonomous science (Lazzeri & Zilio, 2023; Raja, 2019b). Second, both approaches conceive of the organism as continuously interacting with its surrounding environment and take this interaction as the very unit of analysis (Skinner, 1953; Pérez Álvarez, 2014; Segundo-Ortin et al., 2019).

Likewise, neither BA nor ecological psychology relies on internal constructs such as representations or "mental schema" to explain behavior.

4 Covarrubias et al. (2017) offer a similar analysis. However, they focus exclusively on J. J. Gibson (1966).

Instead, both locate the explanatory power of psychology in the relation between organism and environment, and ecological psychologists seem to agree that mental and physical behavior should not be treated as different: "[a] major lesson […] should be that in some sense everything behavioral is learned, action and cognition alike" (E. J. Gibson & Pick, 2000, p. 108).

Despite Skinner never developed a theory of perception, some of his remarks resonate with ecological views: "The visual field is the occasion for effective manipulatory action. […] The visual and the tactual properties of objects in space leads us to develop an effective repertoire in which we approach and reach for objects successfully" (1953, p. 139). Moreover, radical behaviorist's conception of perceptual learning as "learning to respond differentially" (Pérez Fernández et al., 2017, p. 224) fits well with the ecological notion of it as increased differentiation (J. J. Gibson & E. Gibson, 1955).

Interestingly, even though most empirical work in ecological perceptual learning has focused on early development, E. Gibson and A. Pick are explicit that perceptual learning is not exclusive to infancy and that "[p]otential new affordances never stop becoming available, nor do people of any age stop learning to perceive them" (2000, p. 177). Studying perceptual learning beyond infancy entails examining processes that are increasingly shaped by the individual's specific learning history, where "perceptual learning is engaged in adjusting to more and more specialized tasks: acquiring language, using many kinds of implements (spoons and crayons to name just two) and extending body actions to athletic and recreational skills" (p. 178). This individual learning history has been the central concern of BA, which explicitly investigates "individualizable" forms of learning—those that are dependent on an organism's unique, idiosyncratic learning history—and thus supports a methodology based on $n = 1$ designs.

Taken together, these parallels reveal a productive potential for dialogue between ecological psychology and BA. Their shared emphasis on non-representational, agent-environment centeder explanations of behavior suggest that deeper integration between the two could be profitable. The next section explores this possibility.

5.4 Ecological psychology and BA: toward a productive synthesis

In recent decades, a series of ecological psychologists have noted that our relationship with the affordances of the environment does not depend on the detection of information only (Heft, 2007, 2018; Chemero, 2009; Heras-Escribano, 2019; Segundo-Ortin, 2024). For instance, Warren's biomechanical model of *climbability* (1984) does not account for why we typically don't perceive chairs as climbable while in a workplace, although our leg-to-height ratio is less than .88. Similarly, although some insects are objectively nutritious, they are not generally perceived as edible in Western cultures. Reflecting on this, Heras-Escribano notes: "our social norms and

conventions share their space with our individual perception of affordances, and sometimes our norms exert some pressure for not taking certain affordances given some social conventions" (2019, p. 175).

Costall (2012) goes further, suggesting that many of the affordances we perceive and act upon exist only within specific normative contexts. Take a green traffic light: while its color is specified in the ambient optic array, its pragmatic meaning—drive forward—is not. This affordance depends not on physical properties alone, but on a shared background of social conventions. The same logic applies to Gibson's (1979[2015], p. 130) example of a mailbox: for a metal box on the street to afford sending letters, a functioning postal system must be in place.

This raises an important question: if sociocultural norms are not directly specified in perceptual information, how do they influence what we perceive and how we act? The question seems to present ecological psychology with a dilemma: either exclude the perception of norm-dependent affordances from its scope or expand its conceptual resources to account for them. This section explores how BA may help address this challenge.

To that end, we distinguish two related problems. The first is how cultural norms shape the affordances we perceive—why we don't perceive office chairs as climbable or grasshoppers as edible, despite their physical properties. The second concerns affordances that only exist within normative contexts, such as traffic lights or mailboxes.

Addressing these problems calls for an expanded account of perceptual learning. Traditionally, Gibsonians have focused on the *education of attention*—learning to detect the most useful variable for the sought affordance. Yet, as Segundo-Ortin (2024) argues, this must be complemented by an *education of intention*—learning "what affordances are appropriate to seek and actualize given the situation" (p. 8).

Accordingly, while the physical properties and the information about an object remain constant, our perception of certain affordances depends on our socially shaped intentions:

> I do not perceive the office [chairs] as *climbable* because I do not pay attention to the information that specifies this action, but the reason why I do not pay attention to this information is that, when I am at the office, I do not intend to climb on [them].
>
> (p. 8)

Through social training, we learn what affordances we have (and have not) to pursue, and in which specific context, often without conscious reflection.

In addition, Baggs and Chemero (2021) have argued that ecological psychology needs an account of individual learning that explains "how the world can appear differently to different members of the same species, relative to their skills, abilities, and histories" (p. 2175). We believe that BA can contribute to understanding this phenomenon.

Let's begin with the first issue: why do most people not perceive office chairs as climbable regardless of their leg-to-height ratio? According to BA, the answer lies in the consequences of behavior. This framework conceptualizes "climbing a chair" as an "operant"—a behavior that is shaped and maintained by its consequences. Accordingly, one is more likely to climb the chair in situations where doing so is followed by reinforcement (e.g., successfully reaching a high shelf), and less likely to do so in the absence of reinforcement or in the presence of punishment (e.g., receiving a disapproving look from a colleague or being fired).

A similar explanation applies to why people raised in Spain are typically unwilling to eat grasshoppers. Through repeated associations, such as consistently pairing "insect" with "disgust" and lacking any pairing of "insect" with "food" certain classical associations may function as an abolishing operation. This would reduce the likelihood that the individual perceives or interacts with the grasshopper as *a piece of food*, even if they possess the physical ability to do so.

This explanation, however, does not account for our second case: affordances that exist thanks to the normative context. Concerning this issue, Bruineberg et al. (2018) argue for expanding the notion of information to include social conventions. They coin the notion of "General ecological information", defined as "any regularity in the ecological niche between aspects of the environment, x and y, such that the occurrence of aspect x makes the occurrence of aspect y likely" (p. 5237). The idea is that non-specifying variables—those based on conventional constraints or highly reliable regularities—can support the *direct* perception of affordances too.

To illustrate this, they draw on an example from Chemero (2009). Consider an unopened beer can on a table in a well-lit room. Light reflects off the can, and at any location in the room with an unobstructed view, the light will be structured in a lawful way, specifying the presence of a beer can on the table. However, since there is no physical law guaranteeing that all beer cans contain beer—this one could contain soda due to some mistake at the brewery—and no visual information specifies this because the can is opaque, the individual should not see the can as affording the possibility of drinking beer.

Bruineberg et al. (2018) deny this conclusion. According to them, because beer cans *typically* contain beer in our culture, the presence of a closed can of beer become a reliable predictor of beer. Thus, their hypothesis is that individuals with a learning history in the conventional (artificially established) constraints that make it likely that beer cans contain beer will see the unopened can as affording the possibility of drinking beer. This would constitute a case of direct visual perception based on conventions and norms, not physical laws.

Interestingly, while Bruineberg et al. acknowledge the role of the individual's learning history in the perception of norm-based affordances, they limit their analysis to the *education of attention* (p. 5233). We believe this is insufficient. The education of attention involves learning to detect the most

specific informational variables for a given affordance, yet, as we have mentioned, no variable specifies the presence of beer in this case. Therefore, this kind of perception cannot be explained through the education of attention alone.

BA can also contribute to understanding how humans, like other organisms, learn to navigate environments characterized by *artificially* established regularities. The key idea is that, through various learning processes—such as classical conditioning, operant conditioning, or the formation of stimulus equivalence classes—organisms can learn to respond to these contingently structured regularities. As a result, such regularities become informative and acquire functional control over behavior.

For example, in a given context, the sound of a bell might predict the delivery of food, as in the classical conditioning paradigm, leading a dog to salivate in response to the bell. This association is not mandated by physical principles but rather reflects an arbitrary contingency arranged by the experimenter. In a different laboratory setting, another stimulus, such as a tone of a different frequency or even a visual cue like a light, could serve an equivalent predictive function. In either case, the organism learns to respond to an artificially constructed contingency (e.g., bell-food or light-food) via classical conditioning.

A similar dynamic can be observed in the examples previously mentioned. Humans, too, can learn to respond to artificial regularities through a variety of learning processes. For instance, principles of operant conditioning help to explain how the green color of a traffic light comes to function as a "discriminative stimulus", signaling that it is safe to cross the street without risk of collision or injury. This discriminative function emerges through diverse learning histories and contextual experiences: verbal instructions (e.g., a parent telling a child, "if it's green, you can cross"), direct observation (e.g., noticing that others cross on green but not on red), trial-and-error learning (e.g., attempting to cross on red and being honked at by a driver), etc. Through these processes, individuals learn that the green light signals that the operant response—crossing the street—can be emitted in the absence of punishment (e.g., having an accident).

Analogous learning processes operate in other common situations. A metal box on the street comes to signal the opportunity to send a letter, and a label on a can signals the presence of beer. These examples underscore that human behavior is shaped not only by regularities grounded in physical laws but also by contingencies that are arbitrarily and socially constructed. Above all, these examples illustrate that humans navigate their environments not only by responding to information grounded in physical laws, but also by adapting their behavior in accordance with regularities based on social norms.

Before concluding, it is worth dedicating a few words to the role of verbal behavior, both in relation to the current discussion and to the earlier examples. BA helps us understand that, although learning can be facilitated through verbal mediation (e.g., someone might explicitly tell you: "Green

means you can cross" or "Don't climb that chair"), such mediation is not necessary for learning to occur. Verbal input is merely one of many possible sources of information. Direct interaction with the environment—without the involvement of verbal stimuli—is often sufficient for acquiring and maintaining the behaviors in question. For example, one might refrain from climbing a chair in the office simply because one has observed that others do not do so, and because previous attempts were met with disapproving looks from colleagues. Similarly, one might avoid crossing the street when the light is red because, on a previous occasion, a car honked in response. In these cases, verbal instruction is not required to shape or maintain the behavior. In the same vein, verbal self-instruction is not necessary to regulate one's own actions: there is no need to consciously think "I should not climb the office chairs"—one simply refrains from doing so.

5.5 Concluding remarks

Critics of ecological psychology often dismiss it as a modern form of behaviorism. This accusation, however, is rarely presented with sufficient clarity. First, scholars who summon the "behaviorist bogeyman" (Alksnis & Reynolds, 2021, p. 5804) against ecological psychology tend to overlook that behaviorism is not a unified theory but a diverse set of approaches, many of which are incompatible. As a result, it is often unclear which version of behaviorism ecological psychology is being compared to. Second, these critiques seldom specify which assumptions, hypotheses, or methods of ecological psychology allegedly resemble those of behaviorism.

In this chapter, we aimed to address these concerns by examining ecological psychology in relation to four major strands of behaviorism: Watson's methodological behaviorism, the neobehaviorism of Holt, Tolman, and Hull, philosophical behaviorism, and Skinner's radical behaviorism—including BA.

Moreover, we have argued for a productive, yet unexplored, synthesis of ecological psychology and BA. As we see it, whereas ecological psychology succeeds in explaining real-time perception-action couplings through the detection of information about affordances, BA offers a powerful account of how individuals learn to perceive and act upon socially and culturally structured affordances. These affordances, we have suggested, are not always grounded in physical laws alone but often emerge from normative and conventional regularities. By integrating these approaches, we can better account for how individuals come to perceive and act upon affordances that are not only physical, but also deeply cultural and normative—thus moving closer to a comprehensive science of situated human behavior.

Rather than dismissing ecological psychology as a rebranded behaviorism, we suggest treating both frameworks as partners in the broader project of developing a naturalistic, non-computational science of cognition and behavior.

References

Adams, F. R. (2018). Cognition wars. *Studies in History and Philosophy of Science Part A*, *68*, 20–30. https://doi.org/10.1016/j.shpsa.2017.11.007

Aizawa, K. (2015). Cognition and behavior. *Synthese, 194*(11), 4269–4288.

Alksnis, N., & Reynolds, J. (2021). Revaluing the behaviorist ghost in enactivism and embodied cognition. *Synthese*, *198*(6), 5785–5807. https://doi.org/10.1007/s11229-019-02432-1

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. The MIT Press.

Baggs, E., & Chemero, A. (2021). Radical embodiment in two directions. *Synthese*, *198*(9), 2175–2190. https://doi.org/10.1007/s11229-018-02020-9

Barrett, L. (2012). *Why behaviorism isn't satanism*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199738182.013.0002

Barrett, L. (2015). Why brains are not computers, why behaviorism is not satanism, and why dolphins are not aquatic apes. *The Behavior Analyst*, 1–15. https://doi.org/10.1007/s40614-015-0047-0

Baum, W. M. (2011). What is radical behaviorism? A review of jay moore's *conceptual foundations of radical behaviorism*. *Journal of the Experimental Analysis of Behavior*, *95*(1), 119–126. https://doi.org/10.1901/jeab.2011.95-119

Baum, W. M. (2017). *Understanding behaviorism: Behavior, culture, and evolution*. John Wiley & Sons.

Blau, J. J. C., & Wagman, J. B. (2023). *Introduction to ecological psychology: A lawful approach to perceiving, acting, and cognizing* (1st Edition). Routledge.

Block, N. (2001). Behaviorism revisited. *Behavioral and Brain Sciences*, *24*(5), 977–978. Cambridge Core. https://doi.org/10.1017/S0140525X01270113

Bruineberg, J., Chemero, A., & Rietveld, E. (2018). General ecological information supports engagement with affordances for 'higher' cognition. *Synthese*, *196*, 1–21. https://doi.org/10.1007/s11229-018-1716-9

Charles, E. P. (Ed.). (2011). *A new look at new realism: The psychology and philosophy of E.B. Holt*. Transaction Publishers.

Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press.

Costall, A. (2004). From darwin to watson (and cognitivism) and back again: The principle of animal- environment mutuality. *Behavior and Philosophy*, *32*(1), 179–195.

Costall, A. (2012). Canonical affordances in context. *Avant: Trends in Interdisciplinary Studies*, *3*(2).

Covarrubias, P., Cabrera, F., & Jiménez, Á. A. (2017). Invariants and information pickup in *the senses considered as perceptual systems*: Implications for the experimental analysis of behavior. *Ecological Psychology*, *29*(3), 231–242. https://doi.org/10.1080/10407413.2017.1332460

Feest, U. (2025). *Operationism in psychology: An epistemology of exploration* (1st Edition). University of Chicago Press.

Freixa i Baqué, E. (2003). ¿Qué es conducta? *International Journal of Clinical and Health Psychology*, *3*(3), 595–613.

Fuentes Ortega, J. B. F. (2019). El aprendizaje como contexto determinante de la psicología científica: Metodología biológica versus metodología psicológica. *Revista de historia de la psicología*, *40*(2), 27–41.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Prentice-Hall.

Gibson, E. J. (1991). *An odyssey in learning and perception*. MIT Press.
Gibson, E. J (2001). *Perceiving the affordances*. Taylor & Francis.
Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. Oxford University Press. https://doi.org/10.1093/oso/9780195118254.001.0001
Gibson, J. J. (1957). Survival in a world of probable objects. *Contemporary Psychology: A Journal of Reviews*, *2*(2), 33–35. https://doi.org/10.1037/005466
Gibson, J. J. (1958[1982]). Visually controlled locomotion and visual orientation in animals. In: Reed, E., & Jones, R. (Eds.). *Reasons for realism: Selected essays of James J. Gibson* (pp. 148–163). L. Erlbaum.
Gibson, J. J. (1966). *The senses considered as perceptual systems*. Greenwood Press.
Gibson, J. J. (1977[1982]). Notes on direct perception and indirect apprehension. In: Reed, E., & Jones, R. (Eds.). *Reasons for realism: Selected essays of James J. Gibson* (pp. 289–293). L. Erlbaum.
Gibson, J. J. (1979[2015]). *The ecological approach to visual perception*. Psychology Press.
Gibson, J. J., & Gibson, E. J. (1955). *Perceptual learning: Differentiation or enrichment?*
Graham, G. (2023). Behaviorism. In: Zalta, En E. N. & Nodelman, U. (Eds.). *The stanford encyclopedia of philosophy* (Spring 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/behaviorism/
Heft, H. (1983). Way -finding as the perception of information over time. *Population and Environment*, *6*(3), 133–150. https://doi.org/10.1007/BF01258956
Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Psychology Press.
Heft, H. (2007). The social constitution of perceiver-environment reciprocity. *Ecological Psychology*, *19*(2), 85–105.
Heft, H. (2012). Holt's "recession of the stimulus" and the emergence of the "situation" In Psychology. En *A new look at new realism*. Routledge.
Heft, H. (2013). Wayfinding, navigation, and environmental cognition from a naturalist's stance. En *Handbook of spatial cognition* (pp. 265–294). American Psychological Association. https://doi.org/10.1037/13936-015
Heft, H. (2018). Places: Widening the scope of an ecological approach to perception–action with an emphasis on child development. *Ecological Psychology*, *30*(1), 99–123. https://doi.org/10.1080/10407413.2018.1410045
Heras-Escribano, M. (2019). *The philosophy of affordances*. Palgrave.
Heras-Escribano, M., Noble, J., & de Pinedo, M. (2014). Enactivism, action and normativity: A Wittgensteinian analysis. *Adaptive Behavior*. http://adb.sagepub.com/content/early/2014/12/07/1059712314557364.abstract
Holt, E. B. (1915). *The Freudian wish and its place in ethics*. Henry Holt.
Holt, E. B. (1931). *Animal drive and the learning process: An essay toward radical empiricism*. Henry Holt.
Hull, C. L. (1952a). *A behavior system; an introduction to behavior theory concerning the individual organism*. Yale University Press.
Hull, C. L. (1952b). Clark L. Hull. In: Boring, E. G., Langfeld, H. S., Werner, H., & Yerkes, R. M. (Eds.). *A history of psychology in autobiography* (Vol. 4). Clark University Press.
Higuchi, T., Murai, G., Kijima, A., Seya, Y., Wagman, J. B., & Imanaka, K. (2011). Athletic experience influences shoulder rotations when running through

apertures. *Human Movement Science*, *30*(3), 534–549. https://doi.org/10.1016/j.humov.2010.08.003

Lazzeri, F., & Zilio, D. (2023). Commitments with reductive and emergent relations in behavior analysis. *Behavior and Philosophy*, *51*, 102–124.

Leahey, T. H. (2018). *A history of psychology: From antiquity to modernity*. Routledge.

Levin, J. (2023). Functionalism. In: Zalta, E. N., & Nodelman, U. (Eds.). *The stanford encyclopedia of philosophy* (Summer 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/functionalism/

Lobo, L., Travieso, D., Barrientos, A., & Jacobs, D. M. (2014). Stepping on obstacles with a sensory substitution device on the lower leg: Practice without vision is more beneficial than practice with vision. *PLOS ONE*, *9*(6), e98801. https://doi.org/10.1371/journal.pone.0098801

Malone, J. C. (2009). *Psychology: Pythagoras to present*. MIT Press.

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*(2), 320–339. https://doi.org/10.1111/j.1756-8765.2009.01022.x

Núñez de Prado Gordillo, M., Abalo-Rodríguez, I., Estal-Muñoz, V., & Froxán Parga, M. X. (2020). Cuestiones filosóficas en torno al análisis de la conducta. En *Análisis funcional de la conducta humana. Concepto, metodología y aplicaciones* (pp. 51–80). Pirámide.

O'Brien, G., & Opie, J. (2015). Intentionality lite or analog content? *Philosophia*, *43*(3), 723–729. https://doi.org/10.1007/s11406-015-9623-5

Pérez Álvarez, M. (2014). *Contingencia y drama*.

Pérez Fernández, V., Gutiérrez Domínguez, M. T., & García García, A. (2017). *Procesos psicológicos básicos*. UNED - Universidad Nacional de Educación a Distancia.

Raja, V. (2019a). From metaphor to theory: The role of resonance in perceptual learning. *Adaptive Behavior*, *27*(6), 105971231985435. https://doi.org/10.1177/1059712319854350

Raja, V. (2019b). J. J. Gibson's most radical idea: The development of a new law-based psychology. *Theory & Psychology*, *29*(6), 095935431985592. https://doi.org/10.1177/0959354319855929

Raja, V. (2021). Resonance and radical embodiment. *Synthese*, *199*(S1), 113–141. https://doi.org/10.1007/s11229-020-02610-6

Reed, E. (1996). *Encountering the world: Toward an ecological psychology*. Oxford University Press.

Reed, E., & Jones, R. (1982). *Reasons for realism: Selected essays of James J. Gibson*. L. Erlbaum.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*(6), 867–891. https://doi.org/10.1016/j.humov.2007.07.002

Richardson, M. J., Shockley, K., Fajen, B. R., Riley, M. A., & Turvey, M. T. (2008). 9—Ecological psychology: Six principles for an embodied–embedded approach to behavior. En *Handbook of cognitive science* (pp. 159–187). Elsevier. https://doi.org/10.1016/B978-0-08-046616-3.00009-8

Ryle, G. (1949[2000]). *The concept of mind*. Routledge.

Schlosser, M. E. (2020). Why behaviorism and anti-representationalism are untenable. *The Journal of Mind and Behavior*, *41*(3/4), 277–292.

Segundo-Ortin, M. (2020). Agency from a radical embodied standpoint: An ecological-enactive proposal. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01319

Segundo-Ortin, M. (2024). Socio-cultural norms in ecological psychology: The education of intention. *Phenomenology and the Cognitive Sciences*, *23*(1), 1–19. https://doi.org/10.1007/s11097-022-09807-9

Segundo-Ortin, M., Heras-Escribano, M., & Raja, V. (2019). Ecological psychology is radical enough: A reply to radical enactivists. *Philosophical Psychology*, *32*(7), 1001–1023. https://doi.org/10.1080/09515089.2019.1668238

Segundo-Ortin, M., & Kalis, A. (2022). Intentions in ecological psychology: An anscombean proposal. *Review of Philosophy and Psychology*, *15*, 69–89. https://doi.org/10.1007/s13164-022-00661-x

Segundo-Ortin, M., & Raja, V. (2024). *Ecological psychology*. Cambridge University Press.

Silva, P., Kiefer, A., Riley, M. A., & Chemero, A. (2019). Trading perception and action for complex cognition: Application of theoretical principles from ecological psychology to the design of interventions for skill learning. En *Handbook of embodied cognition and sport psychology* (pp. 47–74). The MIT Press.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. B. F. Skinner Foundation.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review*, *52*(5), 270–277. https://doi.org/10.1037/h0062535

Skinner, B. F. (1951). How to teach animals. *Scientific American*, *185*(6), 26–29.

Skinner, B. F. (1953). *Science and human behaviour*. Simon and Schuster.

Skinner, B. F. (1957). A functional analysis of verbal behavior. En *Verbal behavior* (pp. 1–12). Appleton-Century-Crofts. https://doi.org/10.1037/11256-001

Skinner, B. F. (1976). *About behaviorism*.

Skinner, B. F. (1984). Selection by consequences. *The Behavioral and Brain Sciences, 213*, 501–504.

Skinner, B. F., & Ferster, C. B. (1957). *Schedules of reinforcement*. B. F. Skinner Foundation.

Spaulding, S. (2011). Embodied social cognition. *Philosophical Topics*, *39*(1), 141–162. https://doi.org/10.5840/philtopics201139118

Staddon, J. E. R. (2021). *The new behaviorism: Foundations of behavioral science*. Routledge.

Tanney, J. (2022). Gilbert ryle. En: Zalta, E. N. (Ed.). *The stanford encyclopedia of philosophy* (Summer 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2022/entries/ryle/

Thomas, B. J., & Riley, M. A. (2014). Remembered affordances reflect the fundamentally action-relevant, context-specific nature of visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(6), 2361–2371. https://doi.org/10.1037/xhp0000015

Tolman, E. C. (1932). *Purposive behavior in animals and men*. Century.

Tolman, E. C., & Honzik, C. H. (1930). "Insight" in rats. *University of California Publications in Psychology*, *4*, 215–232.

Turvey, M. T. (2019). *Lectures on perception: An ecological perspective*. Routledge.

Uttal, W. R. (1999). *The war between mentalism and behaviorism: On the accessibility of mental processes*. Psychology Press. https://doi.org/10.4324/9781410603111

van der Meer, A. L. H., Svantesson, M., & van der Weel, F. R. R. (2012). Longitudinal study of looming in infants with high-density EEG. *Developmental Neuroscience*, *34*(6), 488–501. https://doi.org/10.1159/000345154

van der Weel, F. R., & van der Meer, A. L. H. (2009). Seeing it coming: Infants' brain responses to looming danger. *Naturwissenschaften*, *96*(12), 1385–1391. https://doi.org/10.1007/s00114-009-0585-y

Venkatraman, V., Lee, J., & Schwarz, C. (2016). Steer or brake? Modeling drivers' collision avoidance behavior using perceptual cues. *Transportation Research Record Journal of the Transportation Research Board*, *2606*(1), 16–6657. https://doi.org/10.3141/2602-12

Wagman, J. B., & Stoffregen, T. A. (2020). It doesn't add up: Nested affordances for reaching are perceived as a complex particular. *Attention, Perception, & Psychophysics*, *82*(8), 3832–3841. https://doi.org/10.3758/s13414-020-02108-w

Warren, W. H. (1984). Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 683–703. https://doi.org/10.1037/0096-1523.10.5.683

Warren, W. H. (1998). Visually controlled locomotion: 40 years later. *Ecological Psychology*, *10*(3–4), 177–219. https://doi.org/10.1080/10407413.1998.9652682

Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, *113*(2), 358–389. https://doi.org/10.1037/0033-295X.113.2.358

Warren, W. H. (2019). Non-euclidean navigation. *Journal of Experimental Biology*, *222*(Suppl_1), jeb187971. https://doi.org/10.1242/jeb.187971

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. https://doi.org/10.1037/h0074428

Watson, J. B. (1930). *Behaviorism*. Norton.

Weimer, W. B. (2024). Notes on the Gibsonian variant of behaviorism. *Journal of Mind and Behavior*, *45*(2), 137–202.

Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell.

6 Being a good gadfly

Radical enactivism's positively revisionary approach to cognitive science

Daniel D. Hutto

Feynman argues, "Science is the belief in the ignorance of experts", that "When someone says 'science teaches such and such', he is using the word incorrectly" (*The Pleasure of Finding Things Out*, p. 187). At its core, Enactivism offers us a conception of cognition that characterizes it as situated in the world through historically conditioned, spatiotemporally extended, embodied organismic activity.[1] Given this animating vision of what lies at the core of mind and cognition, it is easy to see why enactivism, which comes in several forms, sits near phenomenology, pragmatism, and Buddhism in philosophical state space.[2] These philosophical frameworks and traditions of thought are close neighbors in terms of their logical geography. This explains why long-established alliances exist between enactivists, phenomenologists, and pragmatists who often work together on specific topics of common philosophical interest.

Yet, what of analytic philosophy? While all enactivists critically engage with offerings from analytic philosophy of mind, Radical Enactivism is unique in having adapted and incorporated proposals from that tradition into the heart of its positive account of cognition. More precisely, while other enactivists have in part defined themselves by rejecting and opposing cognitivist and representationalist assumptions, Radical Enactive/Embodied

1 The best-known early use of the term *enaction* was in *The Embodied Mind*, co-authored by Varela, Thompson, and Rosch and published in 1991. That seminal work provided the platform for later articulations of enactivism, such as Thompson's (2007) and Di Paolo's (2005, 2009) developments. O'Regan and Noë (2001) created their own variant of Sensorimotor Enactivism, which was developed further in Noë (2004). In a more recent solo effort, Noë (2023) describes the basic idea of the enactive approach as follows: "Perception, thought of at maximum generality, is an organized activity of engaging with the environment, making use of skills of access (concepts, sensorimotor skills). Different sensory modalities, or even the contrast between perception and thought itself, come down to differences in the ways that we deploy our understanding. The varieties of presence correspond to different styles of knowledgeable engagement with the world." (pp. 57–58).

2 See Hutto (2025) for a discussion of these connections.

DOI: 10.4324/9781003649359-8
This chapter has been made available under a CC BY 4.0 license.

account of Cognition (REC) has forged its distinctive conception of mind by engaging directly with cognitivist proposals—not by simply opposing them but by borrowing from them and adjusting what they have to offer.

Indeed, among all the frameworks that self-identify as enactivist, Radical Enactivism alone continues to fashion itself from an eclectic array of diverse sources, selectively incorporating elements from the full sweep of cognitivist and E-theoretical approaches. It includes and integrates what it deems most viable to develop and refine its distinctive conception of cognition.

The REC has always been a hybrid creation—one that selectively appropriates and retools material taken from across the philosophical landscape.

Strangely, REC's constructive, tinkering strategy of framework-building has earned it a reputation for being essentially negative in character. Its critics consistently claim that REC fails to provide any positive alternative to cognitivism. Yet, such assessments are based on a mistaken view of the nature of REC's philosophical efforts to revolutionize our conception of cognition and of what making such a fundamental conceptual shift requires. This chapter sets out to correct those misconceptions and to demonstrate REC's transformative potential.

The action of this chapter unfolds as follows. Section 6.1—*REC's Positively Positive Framework*—sets the record straight. It shows, *pace* critics, that REC does in fact offer a positive and principled framework for reconceiving cognition—one built from a distinctive method of philosophical synthesis. Drawing on both analytic and enactivist sources, REC integrates and retools elements from across traditions to clarify what cognition is and is not. This section addresses and defuses recurring criticisms from Thompson, Baggs, and Wheeler, showing that they mischaracterize and systematically undervalue the nature of REC's contributions.

Section 6.2—*REC's Tools and Targets*—reminds the reader of the two main corrective tools in REC's philosophical workshop: the Hard Problem of Content (HPC) and the Information Processing Challenge (IPC). Their role in REC's revolutionary efforts is clarified.

Section 6.3—*Two Clarifying Case Studies*—shows what REC's analytic work can achieve by considering two case studies. The first examines Rescorla's account of Bayesian cognition; the second intervenes in the debate over the existence of memory traces, focusing on recent work by Robbins. In both cases, it is shown how and why REC's dissolution of conceptual confusion ultimately requires us to seek more viable avenues for characterizing and explaining the phenomena in question.

This chapter concludes by clarifying how REC's clarificatory method of playing the part of being a good gadfly is perfectly in tune with the sort of revisionary philosophical ambition that has always been the hallmark of progressive science. REC exemplifies how genuine conceptual progress is made: not only by discarding what came before, but also by reworking familiar materials into forms that better fit the phenomena—and better serve our real explanatory needs.

6.1 REC's positively positive framework

REC debuted in Hutto (2005) and was most substantially developed in Hutto and Myin (2013, 2017), with key contributions by others.[3] It is one of a family of radical accounts of E-cognition that maintain that at least certain forms of cognition can be wholly constituted by situated actions and interactions that do not essentially involve the manipulation of contentful mental representations at all (see also Varela et al. 1991; Gallagher 2017, 2020; Thompson 2007; Chemero 2009; Di Paolo et al. 2017, 2018).

REC openly espouses well-known revolutionary, strongly revisionist ambitions. Like all the E-approaches mentioned above, REC rejects the mainstream cognitivist account of cognition that dominates today's sciences of the mind. But unlike those accounts, it poses fundamental challenges to advocates of the cognitivist framework in the form of analyses and arguments. Specifically, in this regard, Hutto and Myin (2013) articulated the HPC and the IPC—which will be discussed in greater detail in the next section. Its focus on detailed argumentation—especially via the HPC and IPC—has led some to see REC as emphasizing critique over construction, accentuating the negative rather than foregrounding its broader positive vision.

Yet, REC has positive ambitions beyond posing challenges to classical cognitivism and the E approaches that continue to embrace it. REC, unashamedly, aspires to be the one E to rule them all (see, e.g., Hutto & Myin 2017). In that respect, it seeks to put the house of E in order.

Talk of 4E cognition (Embodied, Embedded, Enactive, and Extended) abounds. Some researchers advocate 3E cognition, swapping Embedded for Environmentally Scaffolded and omitting Extended altogether. Others speak of 5E cognition (which adds Ecological into the 4E set). This practice prompts pressing questions: how many Es do we need? Which ones, and why? How do they cohere into a viable framework?

These are pressing questions if we want to explicate a tenable E-framework, since not all Es can live together under the same theoretical roof. This is a matter of logical consistency, not mere theoretical preference. For example, those who conceive of cognition as essentially Enactive necessarily reject the idea that cognition is Extended because, given certain standard assumptions, Enactive and Extended are logically incompatible. What to do? One could temper one's commitment to Enactive, if one favors Extended. Alternatively, if one favors Enactive, as REC does, one could modify one's understanding of

3 REC was originally proposed as a critical adjustment to the conservative tendencies of sensorimotor enactivism, as set out in O'Regan and Noë (2001). It has since been defended and developed in many publications, including: Hutto and Myin (2013), Hutto and Satne (2015), Zahidi and Myin (2016), Hutto and Myin (2017), Hutto and Peeters (2018), Hutto et al. (2019), Zahnoun (2021a), Zahnoun (2021b), Zahidi (2021), Robertson (2022); Robertson (2023).

Extended—and replace Extended with Extensive (see, e.g., Hutto, Kirchhoff, & Myin 2014; Sganzerla, Hutto & Kirchhoff 2025).

This is just a quick reminder that, if we seek a coherent framework for thinking about cognition, not all Es fit together without adjustment. Thus, those who think that cognition is, at root, Enactive—as REC does—will also think that the way it is Embodied and Embedded must be understood in terms of its fundamentally Enactive character.

Relatedly, we may wish to add other Es into the cognitive mix—e.g., Enculturated, Engaged, Entangled—or some combination thereof. The existing list of Es is neither exhaustive nor complete. There are other things that might yet be included or removed from the E-cognition framework. But crucially, whatever we decide on that score, concerning the members of the list, we need to know how the various items on it hang together in a coherent framework.

These reflections put us in a better position to address questions that have been raised about how exactly we should understand REC's revolutionary ambitions. Are they primarily philosophical or scientific? Are they plausibly realizable (Gärtner & Clowes 2017)? Is such a revolution desirable—do we even want or would we benefit from such a shift in our thinking about the nature of cognition (Wheeler 2017)?

Some are put off by REC's self-professed radicality. Perhaps this is because talk of "radicalizing" tends to be associated in the popular sphere with unwelcome extremist political movements. Still, making that association would be peculiar in the context in which REC operates. For, at least in science and philosophy, radical investigations are precisely those that seek to get at something's fundamental nature and thus to discover or uncover something far-reaching or thorough about a given phenomenon.

When we reorient our thinking on that basis, it is legitimate to speak of conceptual revolutions or radical overhauls of our previous ways of thinking about the chosen topic. Radical philosophical work of this sort opposes the unchecked or uncritical acceptance of familiar assumptions.[4]

In addressing these sorts of concerns, it is important to bear in mind that those who seek to promote REC's proposed revolution are first and foremost trying to bring about a philosophical revolution—one that aims to change our thinking about the nature of mind and cognition, not by doing first-order science or even by marshaling empirical findings, though it respects those. Rather, methodologically, the proposed revolution is to be achieved through philosophical argumentation and the clarification and adjustment of background assumptions (Hutto 2009, 2017).

4 Appropriate synonyms for the ambition of REC's radical investigations include: thoroughgoing, comprehensive, exhaustive, root-and-branch, sweeping, far-reaching, extensive, profound, serious, major, rigorous.

REC seeks to effect a conceptual revolution by exposing weaknesses in prevailing conceptions of cognition and articulating a viable alternative. The guiding assumption is straightforward: if REC's arguments prove compelling, and if its proposed adjustments to and synthesis of existing frameworks prove tenable, then—barring the articulation of better alternatives—we ought to take its framework seriously. Moreover, if so, this ought to lead to changes in how experimental work in the sciences of the mind is framed, conducted, reported, and discussed. Likewise, the adoption of REC's framework ought to make a difference in other practical domains such as education, law, mental health, and professional ethics.

It is important to stress that the use of "ought" in the preceding sentences is normative, not descriptive. REC makes no claims and offers no predictions about what shape cognitive science or other practical domains will, in fact, take in response—or lack thereof—to the arguments it lays out.

This may all seem well and good, but another oft-repeated complaint has persuaded many to doubt REC's prospects of achieving its revolutionary aims.

Wheeler (2017) outlines what he believes REC would need to provide if it were to revolutionize cognitive science—and why he thinks it will likely fail. Channeling his inner Kuhn, he writes:

> A well-established explanatory framework, even one that confronts acknowledged difficulties, should be replaced only when a new, competing framework offers robust evidence of its own explanatory and predictive superiority. In short, I take Hutto and Myin's negative arguments against the idea that basic minds are contentful to be inconclusive, even if they go through. What will finally decide the issue is the positive character of radical enactivism as a conceptual and explanatory framework.
>
> (2007, p. 466)

According to Wheeler (2017), REC "currently fails to deliver [its promised] revolutionary transformation of cognitive-scientific explanation … and thus … its positive character remains fundamentally unclear" (p. 466).

Wheeler is not alone in this diagnosis. Baggs (2017), in his review of *Evolving Enactivism*, applauds REC's challenges to cognitivism but laments that it falls short of delivering a positive alternative that can serve as cognitivism's successor: In his words, "Railing against cognitivism is a noble pastime. Maybe even a heroic one… What we need, though, is a positive project with which to replace cognitivism".

Thompson (2018) echoes this concern, suggesting that to make a serious contribution to cognitive science, REC must develop its own generative framework—one that can be empirically tested. He complains: "They do not systematically construct a positive account from the ground up … they

defend their claim that basic cognition is contentless by analyzing and criticizing other theories". For him, good cognitive science "requires starting from basic theoretical and empirical issues and using them to motivate the careful construction of a positive theoretical framework with testable models".

There is a clear pattern here. The underlying, largely unvoiced, assumption behind all these criticisms is that the only way a philosophical framework can transform cognitive science is by constructing a novel, empirically generative framework that yields testable hypotheses and outperforms competitors on explanatory and predictive metrics.

How should REC respond?

First, let's consider Thompson's charge that REC fails to "construct a positive account from the ground up". At first glance, this sounds serious. But what could this possibly mean, in general, and in REC's case, in particular?

REC is a unique philosophical product—more like a blended whisky than a purist single malt inherited, derived, and adapted, like Thompson's (2007) or Di Paolo's (2005) versions of enactivism, from single source Varela-style first principles. But that doesn't imply REC doesn't have a constructed, positive account.[5]

In a post-Frege lecture interview, when asked to describe REC's method of framework-building, I invoked what I take to be an apt if monstrous, literary allusion:

> Victor von Frankenstein was steadfastly constructive. He took dead parts and, by putting them together in the right way, brought to life something novel through his synthetic work. I see myself as a philosophical tinkerer with similar ambitions.
>
> (Mölder 2022, p. 3)[6]

The following long quotation from Hutto (2011b) illustrates not only what motivates REC, but how it has constructed the most distinctive features of its framework. This passage follows on from the observation that explanatory naturalists who assume that even basic forms of cognition are contentful must supply a convincing theory of content:

> … there is every reason to doubt that this theoretical debt can be paid. A review of the situation strongly suggests that the required theory of content is not on the cards. Only a handful of representationalists have ever tried, in earnest, to pay their theoretical bills in full. The result has

5 Perhaps the concern is not that REC is not constructed, but that it is not constructed "from the ground up". But what does this mean, and why should it matter? Would it make a difference if REC's positive account was constructed "sideways on" or "from top down"?

6 See also Heras-Escribano, M. (2021) and Segundo-Ortin, M. (2020) for other synthetic framework-building efforts of this kind.

been a small clutch of well-developed information-based naturalistic theories of content. Ultimately, they all fail.

Most adopt some or other variety of inferential or conceptual role semantics, according to which what a symbol represents depends (at least partly) upon the use that the rest of the cognitive system is apt to make of that symbol …

There is an immediate worry with this strategy in that no one has yet explained how the imagined symbols, each standing for discrete concepts and which are meant to play the relevant conceptual roles, get their putative semantic properties. Unless this is done, "the semantic properties … are assumed, not explained" (Fodor 1991, p. 46). So, in effect, to go this way is to ask for a line of credit. I believe that this borrowing strategy will lead to a theoretical crisis in cognitive science to rival the economic crisis we currently face in the financial world. Continued borrowing is not the answer; it will only bankrupt future generations of researchers. And it is no good looking for a loan from the banks of Dretske, Fodor, or Millikan to escape the problem; they have all collapsed.

Despite initial optimism, many now doubt that attempts to naturalize semantic content can have any chance of success. Godfrey-Smith (2006) provides an astute assessment: "There is a growing suspicion that we have been looking for the wrong kind of theory, in some big sense. Naturalistic treatments of semantic properties have somehow lost proper contact with the phenomena" (p. 42). Nevertheless, he also acknowledges that the driving idea behind teleosemantics—that evolved structures can have a kind of 'specificity' or 'directedness'—is essentially correct; "there is an important kind of natural involvement relation that is picked out by selection-based concepts of function. But this relation is found in many cases that do not involve representation or anything close to it" (p. 60). What should we make of this?

This suggests a different strategy—that of determining what can be legitimately done with the resources we are actually known to have. With important adjustments, there is much that can be salvaged from attempts to naturalize representational content. For example, although teleosemantic accounts fail to provide an adequate basis for naturalizing intensional (with an 's') content, they are proceeding along basically the right lines. Crucially, they provide the right tools for making sense of something more modest—i.e. responses involving intentionality (with a 't'). To quote a famous Rolling Stones lyric, "You can't always get what you want, but if you try sometimes, you just might find, you get what you need."

In the place of teleosemantics we can put teleosemiotics. Teleosemiotics borrows what is best from teleosemantics and covariance accounts of information to provide a content-free naturalistic account

> of the determinate intentional directedness that organisms exhibit towards aspects of their environments (Hutto 2008, ch. 3). Yet unlike teleosemantics, it does not understand the most basic forms of directedness, such as registering, in semantic (contentful, representational) terms—they are not to be understood in terms of reference or truth conditions. In many cases, organisms act successfully by making appropriate responses to objects or states of affairs in ways mediated by their sensitivity to natural signs. But this does not involve contentfully representing those objects or states of affairs. Undoubtedly, some mental states exhibit semantic intentionality.
>
> Such mental states are properly contentful. Nevertheless, plausibly, a great deal of sophisticated, world-directed cognition exhibits intentional directedness that is not contentful in the sense just discriminated. Teleosemiotics understands on-line perceptual responding as informationally sensitive but it denies that the notion of a purely informational, non-conceptual representing is coherent. It denies that 'carrying information about' X or registering X constitutes "a way of representing X without representing it as anything" (Fodor 2008, p. 182).
>
> (Hutto 2011b, pp. 334–335)

Ironically, in an early commentary on REC's positive proposal, Thompson (2011) initially saw promise in REC's synthetic work:

> Hutto aims to open up a dialogue between analytical philosophy of mind and the enactive approach. His strategy is to show how teleosemantic theories of content need to be modified in a variety of ways that end up bringing these theories closely in line with the basic orientation of the enactive approach.
>
> (2011, p. 19)

At the time, Thompson (2011) approved of REC's effort, writing: "I welcome this dialogue, greatly appreciate Hutto's bridge-building efforts, and find myself largely in agreement with his commentary". (p. 19). Indeed, he went so far as to add:

> … Hutto's version of a modified teleosemantics—'teleosemiotics' as he calls it—would also need to move away from adaptationist views of evolution in order to find common ground with the enactive approach. I see no reason why this movement cannot happen, though I suspect the resulting teleosemiotic theories would look rather far removed from their teleosemantic ancestors.
>
> (2011, p. 20)

Assuming Thompson's (2011) characterization above is accurate, and then it is difficult to deny that REC has synthetically constructed a substantive and

novel philosophical framework. In our line of work, it really doesn't get more positive than that.

What then of the seemingly more serious charge—that REC fails to advance a theory-generating, empirically testable framework?

The first thing to say is: if that *is* truly the core business of philosophers, then we must throw in our lot with philosophical naturalists like Quine and admit, "I'm something of a scientist myself".

REC's recommendation to philosophers, by contrast, cast in a more Wittgensteinian spirit, is: "Don't give up the day job". Philosophy and science must certainly interact, but one does not thereby reduce to the other (Hutto & Satne 2018a, 2018b).

Philosophy can be conducted in many legitimate forms (Hutto 2009). One core problem with Thompson's (2018) critique of REC is that it assumes philosophical work—at least the kind worth contributing to cognitive science—must take the form of advancing a novel, positive, and testable theory. And if not that, then philosophers are to be relegated to more modest roles: refining existing models to make them more empirically tractable or testable.

This overlooks the deep value of fundamental philosophical work that aims to challenge or refine frameworks by compelling us to review and rethink instances of "musty" thinking—axiomatic assumptions taken up uncritically and left unchecked, especially when they underpin cherished constructs at the foundations of a given science.

More than that, REC's critics treat their own enactivist or functionalist proposals—many of which are far less rigorously defended—as valuable contributions to science, even when they offer no testable explanations or empirical predictions. The charge they level against REC is, thus, inconsistently applied. They demand that REC both dismantle foundational assumptions and deliver a ready-made research program of a first-order scientific kind—a burden not placed on their own frameworks.

6.2 REC's tools and targets

When it comes to its work of addressing foundational issues in mainstream cognitive science, REC's primary targets have been the deep-seated commitments to representationalism and computationalism by cognitivists—which bear all the hallmarks of such 'musty thinking'.

What is "musty" thinking? Examples abound. Here are two parade cases from scientists working on memory.

Gallistel and King (2010) tell us:

> This is a long book with a simple message: there *must be* an addressable read/write memory mechanism in brains that encodes information received by the brain into symbols (writes), locates the information when needed (addresses), and transports it to computational machinery that makes productive use of the information (reads) (p. vii, emphasis added).

These authors go on to explicate the precise way in which this framework assumption at the heart of their scientific theorizing is not—or at least, not yet—based on or justified by empirical findings. As they put it:

> Computational cognitive scientists presume the existence of an addressable read/write memory mechanism, yet neuroscientists do not know of, and are not looking for, such a mechanism.
>
> The truths the cognitive scientists know about information processing, when integrated into neuroscience, will transform our understanding of how the brain works.
>
> (Gallistel & King 2010, p. vii)

In this regard, these scientists compare their work in the cognitive sciences to the approach adopted for the molecular identification of the gene in biochemistry—that which led to the discovery of code written into the structure of the DNA molecule.

Or, to take another example of "musty" thinking from psychology, consider this passage from Endel Tulving:

> As a scientist I am compelled to the conclusion—not postulation, not assumption, but conclusion—that *there must exist* certain physical-chemical changes in the nervous tissue that correspond to the storage of information, or to the engram, changes that constitute the necessary conditions of remembering … The alternative stance… is sheer mysticism.
>
> (Endel Tulving, as quoted in Gazzaniga, 1997, emphasis added)

Here, again, we can see a scientist making a fundamental metaphysical commitment to a specific posit—as a matter of necessity. It is a commitment designed to drive the interpretation of empirical findings. Indeed, if it is taken up, it is the sort of commitment that cannot be challenged by such findings, which implies that it cannot be justified by direct appeal to those findings either.

The force of "musty" thinking in these examples should be indelibly clear—it is not a form of Inference to the Best Explanation but rather a kind of Inference from axiomatic assumptions, which is designed to dismiss and disallow any possible alternatives. In this instance, all alternatives are swept aside, and all rival ways of framing things are ruled out a priori—designating them as having irrational or unscientific status.

These kinds of commitments are not rare or incidental; they form the core pillars of the dominant explanatory framework in mainstream cognitive science. And yet, despite their centrality—or rather because of it—they evade proper scrutiny, operating as unquestionable tenets rather than assumptions that require independent justification.

To challenge such assumptions, REC has developed two principal philosophical tools: the HPC and the IPC. These bits of philosophical machinery are not intended to achieve minor theoretical refinements; rather, they are designed to expose deep incoherencies in the deep structure of the mainstream representationalist-cum-computationalist information-processing framework. REC does not merely question the adequacy of its many and various specific models—it questions the conceptual legitimacy of building cognitive science on these sorts of foundations in the first place.

The HPC and IPC function as twin RECing balls—fixed Archimedean points from which REC aims to shift how we think in and about cognitive science.

The HPC challenges explanatory naturalists to give a scientifically respectable account of how mental content fits into and arises within the natural order. It demonstrates that there are fundamental obstacles facing any theory that treats content as a primitive explanatory posit or that uses the standard resources of information theory and biological function to explain its ultimate origins.

Starting from the assumed common ground that information-as-covariance is the only notion of information with the right credentials to play a foundational role in cognitivist theorizing about the origins of content, the HPC begins by observing that covariance alone does not constitute any kind of content. That is something that nearly all committed cognitivists concede. The HPC, however, draws out the serious implications that follow from acknowledging this for the most prominent and promising naturalistic theories of content. It argues that attempts to explicate how content fits into the natural scheme of things—based on appeals to structural similarity or teleological function—fail to justify positing genuinely contentful mental states, rather than causally efficacious, information-sensitive responses.

What's left after the HPC has done its work are promissory notes, metaphysical speculation, or a retreat into fictionalism. In this light, the HPC reverses the burden: if no naturalistic theory can adequately explain how content could or should arise at the most basic levels of cognition, then the natural conclusion is that our accounts of cognition should actively explore ways to do without the assumption of such contents.

The HPC isn't a call for a better-engineered or deflated account of content—it's a call to rethink the need to posit representational content altogether, at least when it comes to understanding what happens at the ground floor of cognition.

Since its first articulation, the HPC has attracted considerable attention and generated a variety of responses. Miłkowski (2015) claimed the HPC overlooked the alleged fact that its core challenge had already been addressed—"at least in principle"—by some variant of teleofunctionalism combined with a control-theoretic account of information. Others, such as Shapiro (2014) and Matthen (2014), accept that the HPC has metaphysical

bite but argue that it can be safely shelved until the appropriate cognitive architecture is identified—on the assumption that clarity about such metaphysical matters need not, and should not, constrain or impede ongoing scientific theorizing and modeling. Still others, including Colombo (2014) and Sprevak (2013), propose fictionalist or instrumentalist strategies that bracket any realistic metaphysical commitments to content entirely.

Apart from Miłkowski's (2015) attempt to convince us that the HPC had already been answered, these views are united by their deferral: they all acknowledge the challenge posed by the HPC but they all avoid offering a straight solution. More recent efforts to address or dissolve the HPC have emerged from Ramsey (2023), Egan (2025), and Richmond (2025). Speaking of the HPC's impact on the field, Schulte (2023) observes: "The debate about radical enactivism and other anti-representationalist approaches is still in its early stages and will surely continue for decades to come" (p. 55).[7]

The HPC is the natural partner of the IPC, which targets a different but no less entrenched assumption in mainstream cognitive science: that information can be treated as a manipulable commodity—something the brain acquires, stores, transmits, and processes.

As REC first argued (Hutto 2011a; Hutto & Myin 2013), this assumption does not hold up to technical scrutiny. If information is understood in scientifically respectable terms—as a lawful, objective covariance between states of affairs—then it is, essentially, relational. It is, thus, not a substance or property that can be picked up, encoded, or shuffled around.

As shown elsewhere, if information is fundamentally a relation—not an entity—then talk of processing, storing, or manipulating information is either a metaphorical gloss or, if taken seriously, it misunderstands what is technically possible.

A classic illustration is the correlation between tree rings and tree age. With the right background knowledge and conceptual capacities in place, we can use that correlation to infer a tree's age, but the tree does not literally store information about its age in the rings that we somehow retrieve when doing so.

Likewise, to treat sensory input as transmitting objective information into the brain is to project metaphors onto cognitive activities that are, in fact, dynamic, embodied interactions between organisms and their environments. For a detailed discussion of these issues, see Hutto (2024), Zahnoun (2025), and Kirchhoff (2025).

For these reasons, REC reframes cognition in terms of information sensitivity, not information processing, taking inspiration from a distinction first

7 Some have attempted to explain why progress on this issue has proven so difficult. Taylor and Williamson (2024) argue that the enduring impasse stems from deep epistemic differences over what counts as evidence in this debate—and how that evidence ought to be treated.

introduced in Hutto (1999). By REC's lights, cognitive systems do not receive information from the world as if it were an inert commodity. Rather, if we are not to be taken in by metaphors, it should be evident that cognizers must actively seek out, register patterns, and coordinate their responses accordingly if they are to be reliably informed and act effectively with respect to the state of the world.

6.3 Two clarifying case studies

Having articulated REC's core philosophical challenges—the HPC and the IPC—we now turn to several illustrative cases in which those challenges help bring clarity to live debates in contemporary cognitive science.

The examples that follow are not chosen at random. Each targets an area where representationalist and computationalist commitments continue to exert a strong influence. In each case, REC's intervention exposes hidden assumptions and reframes the terms of debate.

This section critically evaluates two cases: Rescorla's (2024) case for endorsing representationalism in Bayesian cognitive science; and Robins' (2023) defense of informationally supported, contentful memory traces.

Each case reveals what is gained when REC's conceptual tools are applied in different domains—and what is lost when they are ignored. In each, the signature challenges of the HPC and IPC are readily visible, prompting REC-inspired rethinks that aim to move the sciences of the mind forward.

Consider Rescorla's (2024) assessment—offered, to some extent, in REC's favor—of what predictive processing explanations require. In predictive processing frameworks, prediction errors are said to be calculated by comparing incoming signals with top-down predictions via a generative model—this error is often understood to arise "between" hierarchical levels.

Although this structure can be described in Bayesian terms, Rescorla is clear that he thinks the core operation—namely, the computation of prediction error—can be captured without appeal to representational properties (see also Kirchhoff & Robertson 2018).

Contra Hohwy (2013), Clark (2016), Gładziejewski (2016), and Gładziejewski and Miłkowski (2017), Rescorla (2024) maintain that "Representational properties play no role in characterizing the 'prediction error' computation" (Rescorla 2024, p. 73). He adds, "A similar diagnosis applies to higher levels in hierarchical PPC models" (p. 73). In summary assessment, he concludes: "I agree with Hutto and Myin that we achieve no explanatory gain by glossing this talk in representational terms" (Rescorla 2024, p. 73).

It is worth noting, however, that despite this acknowledgment and even though Rescorla does not address REC's HPC "Can't Have" challenge, he is persuaded that there is at least one domain—Bayesian cognitive science—in which we simply must posit contents featuring in representational explanations.

To establish this, Rescorla advances a "Must Have" argument in response to REC's "Don't Need" challenge, even though he offers no "Can Have" reply to REC's HPC. If a compelling "Must Have" case could be made, we would have strong reasons to posit contentful representations—even if we were never in a position to say whether we'd ultimately be able to explain how we "Can Have" such content.

Put otherwise: even if an answer to the HPC forever evades us, a successful "Must Have" argument would warrant our continued use of explanations that invoke contentful representations.

Let's take a closer look at this "Must Have" move. Rescorla thinks he can demonstrate an absolute need for representationalism when it comes to explicating Bayesian cognitive science. According to his analysis, Bayesian explanations necessarily rely on representationally characterized belief-like states. Indeed, he can't imagine things being otherwise. Thus, he tells us:

> For example, how can we explain overshooting in dead reckoning unless we posit a prior that favors slower speeds? I have no idea how enactivists would interpret the 'slow speed' prior in nonrepresentational terms, let alone how the ensuing explanations would work.
>
> (Rescorla 2024, p. 74, emphasis added)

Rescorla (2024) appears to think these observations settle the matter—especially when coupled with the claim that Hutto and Myin (2017) "do not provide a developed alternative interpretation" (p. 74). However, there are a few things to note here and a few lessons to learn.

First, although it is true that Hutto and Myin (2017) did not provide a developed alternative focused on that specific case, there is plenty of room for industrious and inventive efforts to develop non-representationalist accounts of Bayesian-style cognitive science. For one thing, the slow speed prior might be bound up with cognizers being informationally sensitive in threshold, signature ways to slow speeds, in specific conditions, without their having to contentfully represent slow speeds as such.[8]

Second, it is important to understand the logic of the situation. The mere possibility of an alternative account is sufficient to undermine a "Must Have" defense of representationalism.

8 Rescorla (2024) criticises those who insist that predictive processing account of cognition must posit representations, rightly noting that "the core computational operations are formally describable in ways that do not rely on representational notions" (p. 73). Yet, he misses the irony of his own claims: for when defending the idea that Bayesian cognition requires the positing of representations he defaults to representational glosses and treats representational posits as if they were explanatorily indispensable—without demonstrating their necessity or technical viability. In short, he commits the very same error he himself flags as problematic in those other accounts.

Third, the "Must Have" defense could also be undercut by showing that we "Don't Need" realist representationalism—which can be achieved by adopting instrumentalist, fictionalist, or pragmatist stances.[9]

In the end, it is not easy to pull off a true "Must Have" defense of representationalism. For it to carry the required logical force, it must rule out the very possibility of an alternative explanation—not merely observe that one isn't readily at hand. Presumably, this is why those making the "Must Have" move tend to resort to Inference from Lack of Imagination arguments (Hutto 2007), rather than offering Inference to the Best Explanation arguments, as they should.

Let us consider another case study, concerning the debate about the existence of memory traces and the need to posit them in our best memory science. The commitment to memory traces in contemporary mainstream memory science is not, contrary to how it may appear to some, based on or justified by empirical discoveries. Nor indeed can the positing of memory traces be understood as a straightforward or pure scientific hypothesis. The conviction that memory traces are a "must" arises from a particular way of reading empirical results—a way of reading that is bound up, yet again, with a familiar set of philosophical convictions.

To illustrate the point let's look closely at Robins's (2023) project to enrich our account of the engram. She tells us:

> Recent developments in the tools and techniques available for investigating the mechanisms of memory have allowed researchers to proclaim the search [for the engram] is over. While there is ongoing debate about the justification for that claim, renewed interest in the engram is clear. This attention highlights the impoverished status of the engram concept. As research accelerates, the simple characterization of the engram as an enduring physical change is stretched thin. Now that the engram commitment has been made more explicit, it must also be made more precise. If the project of 20th century neurobiology was finding the engram, the project of the 21st must be supplying a richer account of what's been found.
>
> (p. 1)

In setting out this stall, it is important to observe that Robins (2023) is not being led solely by the empirical findings or the existing science. Instead, she is making a case for augmenting what they offer by appealing to philosophical considerations and proposals. Like Tulving, Gallistel, and King before her, she is first and foremost engaged in philosophical theorizing. Thus, she

9 See, for example, Kirchhoff, Kiverstein, and Robertson (2025) and Kirchhoff (2025) for arguments for in favor of seeing model-based cognitive science as making heavy use of idealizations.

reaches conclusions about the sort of properties the engram must have if it is to play a specific explanatory role in enabling remembering. Those observations and conclusions are not derived from the existing science but are added to it.

We can see this in Robins's (2023) plea that we must articulate a more precise, richer account of the engram. That is a philosophically inspired demand, driven by specific explanatory ambitions. As she acknowledges:

> The trace is generally understood to be the mechanism or process by which information, ideas, and experiences are preserved across time—making possible memory's diachronic nature. It is difficult to say much beyond this; the memory trace remains as elusive as it is enduring. Arguments for traces are rare; appeals to metaphor are common.
>
> (p. 2)

In promoting this notion of memory trace, Robins (2023) is engaged in precisely the kind of foundational philosophical spadework, here found at the heart of cognitive and memory science, that merits counter-philosophical attention and challenge.

Robins is no mere reporter or interpreter of scientific results. She is a philosopher offering conceptual considerations that, purportedly, give us reasons to believe in engrams with distinctive properties that enable them to play specific roles and functions. Her project is not that of simply summarizing and detailing what neuroscience has already discovered. Rather, her analyses of the relevant empirical findings are part of an attempt to give reasons for believing in the existence of engrams with said properties.

We can see this in the way she reaches back to Semon for inspiration, while not giving him the last word on the shape of the science. Rather, she tells us: "I propose that Semon's work be seen as a starting point from which a richer conception of the engram can be built" (Robins 2023, p. 10). And again: "To support these exciting new avenues of inquiry, a richer conception of the engram is required". (p. 10)

Mechanically speaking, Robins (2020) maintains that the engram undergoes both synaptic consolidation—"the initial stabilization of the engram that occurs directly after learning"—and systems consolidation, "where the engram moves from the hippocampus to frontal cortex to make room for the formation of new memories" (p. 1132). But if engrams are, as she imagines, information-bearing, content-generating entities then such explanations seem to imply that "information" is a kind of commodity that is, really, capable of being stored, retrieved, and moved about in ways that the IPC directly challenges.

The issue can be brought into sharp focus by considering the standard cognitivist readings of celebrated optogenetic studies that Robins (2023) highlights. These experiments identify specific neurons active during encoding, tag them with light-sensitive proteins, and later stimulate them to trigger behavior associated with the original experience. The result is taken as proof

that a memory's information and content have been stored and retrieved. But do these findings really establish any such thing?

Certainly, these experiments show that reactivating or suppressing a specific set of neurons is reliably implicated in modifying the behavioral responsiveness of these test subjects. What is not demonstrated by the findings themselves is that we have reason to believe that such neutrally inspired changes support the existence of information-bearing engrams that are capable of generating contentful rememberings. Hence, it is surrounding philosophical reasoning, not the findings themselves, that must persuade us of this.

This is why Robins (2023) concedes that Semon's minimal, scientized notion of the engram is not adequate for her explanatory aims. Having philosophically enriched her notion of the engram, she concludes: "The commitment to the existence of an engram is not merely a commitment to a physical mechanism" (p. 10). If engrams exist, they must do more than play a physically explicable causal role in enabling acts of remembering—they must somehow support the retention and storage of information that gives rise to remembered contents in relevant contexts. Hence, she concludes that what's missing from Semon's initial purely scientific proposal is that "the connection between the causal role and the engram's content is not made explicit" (p. 10).

It should go without saying that Robins's (2023) philosophically enriched proposal about memory traces and her arguments for believing in their existence is open to philosophical challenge. Indeed, the philosophical nature of the proposal positively invites it.

What all this shows—and what I have been at pains to stress here—is that further scrutiny of the empirical details will not settle this case. What is needed is greater clarity about the particulars of these philosophically sourced explanatory proposals that are found at the very heart of today's memory science. In particular, it is entirely appropriate to ask, in a philosophical register, whether the existence of memory traces with the assumed properties is even technically possible.

This is precisely where the HPC and the IPC once again earn their keep (for yet more detailed analyses on this score, see Hutto 2022, 2024).

6.4 Conclusion

What this chapter has sought to demonstrate is that the foundations of mainstream cognitive science are ripe for philosophical investigation and intervention—specifically, of a kind that works to remove uncritical and dogmatic tendencies of thought. To achieve that end, what we need—decidedly—is philosophical work that clarifies rather than rushes to theorize. This is precisely the approach REC supplies.

In this vein, REC might be thought of as a kind of "Analytic" Enactivism. After all, its principal, self-avowed method is RECtification: "a process through which … target accounts of cognition are radicalized by analysis and argument" (Hutto & Myin 2017, p. xviii). The analyses and arguments in

question are designed to remove unwanted assumptions and commitments in our efforts to fashion more viable overall frameworks.[10]

This is precisely what REC seeks to do when challenging the framework commitments of cognitivism, enactivism, and other E-approaches alike—namely, this is precisely what it does when it casts doubt on representationalism, computationalism, functionalism, information processing, knowledge-mediated cognition, and sense-making.[11]

The real contribution that RECers—like myself and Myin—make, as Thompson (2018) notes, is that we do our philosophical duty by being good gadflies. We proudly stand in that long Socratic tradition. Like Socrates, any RECer worth their salt might equally say: "I am that gadfly which God has attached to the state, and all day long and in all places am always fastening upon you, arousing and persuading and reproaching you" (Plato, *Apology* 30e).

In this regard, though REC indeed operates with a subtractive methodology, this need not be seen in a negative light. REC's approach can be understood as a kind of sculpting—aiming to refine and strengthen philosophical proposals. Here again, its methods echo those of Socrates, who tested and reshaped the ideas of his interlocutors. Through its rectifying work, REC seeks a better, more coherent final product—achieved by removing what is unnecessary or obscuring.

Nor is this general approach alien to analytic philosophy. As Wittgenstein (1953) famously put it:

> Where does our investigation get its importance from, since it seems only to destroy everything interesting: that is, all that is great and important? (As it were, all the buildings, leaving behind only bits of stone and rubble.)—What we are destroying is nothing but houses of cards, and we are clearing up the ground of language on which they stand.
>
> (*PI*, Section 127)

So, here's where we've landed: Isn't this the very same approach Thompson (2020) adopts when criticizing confused ways of framing our understanding of Buddhism? On that score, he writes:

> The dominant strand of modern Buddhism, known as 'Buddhist modernism,' is full of confused ideas … They need to be discarded if Buddhism is to take its rightful place as a valuable contributor to a modern cosmopolitan community.
>
> (p. 7)

10 I have long argued that it is possible to pursue philosophy with subtractive, therapeutic methods for purely clarificatory ends (Hutto 2003/2006, 2009).

11 Here it helps to recall that apart from challenging cognitivism, REC's other major aim has been to clarify and refine enactivism by *radicalising* it. Indeed, it does the same with all other valuable proposals about mind and cognition that are worthy of serious philosophical attention.

Explicating further, he adds: "I am not a Buddhist. I'm a philosopher who believes in cosmopolitanism and who also values Buddhism. I like to think of myself as a good friend to Buddhism" (p. 8).

Well, the same holds for RECers. We are philosophers, not cognitive scientists—but in helping cognitive science discard its confused ideas, we are both good gadflies and good friends to the sciences of the mind.

References

Baggs, E. (2017) Evolving enactivism: Basic minds meet content [Book Review], *Frontiers in Psychology*, 8, Article 1947. https://doi.org/10.3389/fpsyg.2017.01947

Chemero, A. (2009) *Radical embodied cognitive science*. Cambridge: MIT Press.

Clark, A. (2016) *Surfing uncertainty: Prediction, action, and the embodied mind*. New York: Oxford University Press.

Colombo, M. (2014) Neural representationalism, the hard problem of content and vitiated verdicts: A reply to Hutto and Myin. *Phenomenology and the Cognitive Sciences* 13 (2): 257–274.

Egan, F. (2025) *Deflating mental representation*. Cambridge: MIT Press.

Di Paolo, E. (2005) Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences* 4 (4): 429–452

Di Paolo, E. (2009) Extended life. *Topoi* 28 (1): 9–21.

Di Paolo, E., Buhrmann, T., & Barandiaran, X. (2017) *Sensorimotor life*. Oxford: Oxford University Press.

Di Paolo, E., Cuffari, E., & De Jaegher, H. (2018) *Linguistic bodies: The continuity between life and language*. Cambridge: MIT Press.

Feynman, R. P. (2001) *The pleasure of finding things out: The best short works of Richard P. Feynman*. London: Penguin.

Fodor, J. A. (1991) Fodor's guide to mental representation. In Greenwood, J. (ed). *The future of folk psychology* (pp. 22–50). Cambridge: Cambridge University Press.

Fodor, J. A. (2008) *LOT2: The language of thought revisited*. Cambridge: MIT Press.

Gallagher, S. (2017) *Enactivist interventions: Rethinking the mind*. Oxford: Oxford University Press.

Gallagher, S. (2020) *Action and interaction*. Oxford: Oxford University Press.

Gallistel, C. R., & King, A. (2010) *Memory and the computational brain*. Malden, MA: Wiley-Blackwell.

Gazzaniga, M. S. (1997) Why can't I control my brain? Aspects of conscious experience. In Ito, M., Miyashita, Y., & Rolls, E. T. (eds). *Cognition, computation, and consciousness* (pp. 69–79). Oxford: Oxford University Press.

Gärtner, K., & Clowes, R. (2017) Enactivism, radical enactivism, and predictive processing: What is radical in cognitive science? *Kairos. Journal of Philosophy and Science* 18 (1), 54–83.

Gładziejewski, P. (2016) Predictive coding and representationalism. *Synthese* 193, 559–582.

Gładziejewski, P., & Miłkowski, M. (2017) Structural representations: Causally relevant and different from detectors. *Biology Philosophy* 32, 337–355.

Godfrey-Smith, P. (2006) Mental representation and naturalism. In Macdonald, G., & Papineau, D. (eds). *Teleosemantics* (pp. 42–68). Oxford: Oxford University Press.

Heras-Escribano, M. (2021) Pragmatism, enactivism, and ecological psychology: Towards a unified approach to post-cognitivism. *Synthese* 198 (1), 337–363.

Hohwy, J. (2013) *The predictive mind*. Oxford: Oxford University Press.

Hutto, D. D. (1999) *The presence of mind*. Amsterdam: John Benjamins.
Hutto, D. D. (2003/2006) *Wittgenstein and the end of philosophy: Neither theory nor therapy*. Basingstoke: Palgrave.
Hutto, D. D. (2005) Knowing what? radical versus conservative enactivism. *Phenomenology and the Cognitive Sciences* 4 (4), 389–405.
Hutto, D. D. (2007) Getting clear about perspicuous presentations: Wittgenstein, Baker, and Fodor. In Moyal-Sharrock, D. (ed). *Perspicuous presentations: Essays on Wittgenstein's philosophy of psychology* (pp. 299–322)Basingstoke: Palgrave.
Hutto, D. D. (2008) *Folk psychological narratives: The socio-cultural basis of understanding reasons*. Cambridge: MIT Press.
Hutto, D. D. (2009) Philosophical clarification: Its possibility and point. *Philosophia* 37:4, 629–652.
Hutto, D. D. (2011a) Philosophy of mind's new lease on life: Autopoietic enactivism meets teleosemiotics. *Journal of Consciousness Studies* 18:5–6, 44–64.
Hutto, D. D. (2011b) Elementary mind minding, enactivist-style. In Seemann, A. (ed). *Joint attention: New developments in philosophy, psychology, and neuroscience*. (pp. 307–341). Cambridge: MIT Press.
Hutto, D. D. (2017) REC: Revolution effected by clarification. *Topoi* 36:3, 377–391.
Hutto, D. D. (2022) Remembering without a trace? One step beyond trace minimalism. In Sant'Anna, A., McCarroll, C. J. & Michaelian, K. (eds) *Current controversies in philosophy of memory*. (pp. 61–81). London: Routledge.
Hutto, D. D. (2024) Acts of remembering: From a radically enactive point of view. *Psychoanalytic Inquiry*, 1–19. https://doi.org/10.1080/07351690.2024.2411902
Hutto, D. D. (2025) Enactivism. *The Internet Encyclopedia of Philosophy*, ISSN 2161–0002, https://iep.utm.edu/, 9 March 2025.
Hutto, D. D., Kirchhoff, M. D., & Myin, E. (2014) Extensive enactivism: Why keep it all in? *Frontiers in Human Neuroscience,* 8, 1–11. https://doi.org/10.3389/fnhum.2014.00706
Hutto, D. D., & Myin, E. (2013) *Radicalizing enactivism: Basic minds without content*. Cambridge: MIT Press.
Hutto, D. D., & Myin, E. (2017) *Evolving enactivism: Basic minds meet content*. Cambridge: MIT Press.
Hutto, D. D., & Peeters, A (2018) The roots of remembering: Extensively enactive RECollection. In Michaelian, K. Debus, D., & Perrin, D. (eds). *New directions in the philosophy of memory* (pp. 97–118) London: Routledge.
Hutto, D. D., Myin, E., Peeters, A, & Zahnoun, F. (2019) The cognitive basis of computation: putting computation in its place. In Colombo, M., & Sprevak, M. (eds). *The Routledge handbook of the computational mind* (pp. 272–282). London: Routledge.
Hutto, D. D., & Satne, G. (2015) The natural origins of content. *Philosophia* 43 (3), 521–536.
Hutto, D. D., & Satne, G. (2018a) Wittgenstein's inspiring view of nature: On connecting philosophy and science aright. *Philosophical Investigations* 41:2, 141–160.
Hutto, D. D., & Satne, G. (2018b) Naturalism in the Goldilocks Zone: Wittgenstein's delicate balancing act. In Raleigh, T., & Cahill, K. (eds). *Wittgenstein and Naturalism* (pp. 56–76). London: Routledge.
Kirchhoff, M. (2025) *The idealized mind: From model-based science to cognitive science*. Cambridge: MIT Press.

Kirchhoff, M. D., Kiverstein, J., & Robertson, I. (2025) The literalist fallacy and the free energy principle: Model-building, scientific Realism, and instrumentalism. *The British Journal for the Philosophy of Science*. 76 (3), 519–804.

Kirchhoff, M. D., & Robertson, I. G. (2018) Enactivism and predictive processing: A non-representational view. *Philosophical Explorations* 21 (2), 264–281.

Matthen, M. (2014) Debunking enactivism: A critical notice of Hutto and Myin's *Radicalizing enactivism*. *Canadian Journal of Philosophy* 44 (1), 118–128.

Miłkowski, M. (2015) The hard problem of content: solved (long ago). *Studies in Logic, Grammar and Rhetoric* 41 (1), 73–88.

Mölder, B. (2022) A synthetic, Frankenstein philosopher using analytic tools to test possibilities: An interview with Daniel D. Hutto. *Studia Philosophica Estonica* 15, 1–16.

Noë, A. (2004) *Action in perception*. Cambridge: MIT Press.

Noë, A. (2023) *The entanglement: How art and philosophy make us what we are*. Princeton, NJ: Princeton University Press.

O'Regan, J. K., & Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24, 939–1031.

Plato. *Apology*, in J. M. Cooper (ed.) *Plato: Complete works*, trans. G.M.A. Grube, rev. J. Cooper, Indianapolis: Hackett Publishing Company, pp. 17–36.

Ramsey, W. (2023) The hard problem of content is neither. *Review of Philosophy and Psychology* 16, 153–174. https://doi.org/10.1007/s13164-023-00714-9

Rescorla, M. (2024) *Bayesian models of the mind*. Cambridge: Cambridge University Press.

Richmond, A. (2025) What is a theory of neural representation for? *Synthese* 205, 14. https://doi.org/10.1007/s11229-024-04816-4

Robertson, I. G. (2022) In defence of radically enactive imagination. *Thought: A Journal of Philosophy*. 11 (4), 184–191. https://doi.org/10.5840/tht202421525

Robertson, I. G. (2023) The unbearable rightness of seeing? Conceptualism, enactivism, and skilled engagement. *Synthese* 202, 173. https://doi.org/10.1007/s11229-023-04385-y

Robins, S. K. (2020) Stable engrams and neural dynamics. *Philosophy of Science* 87 (5), 1130–1139.

Robins, S. (2023) The 21st century engram. *Wiley Interdisciplinary Reviews: Cognitive Science* 14 (5), e1653. https://doi.org/10.1002/wcs.1653

Segundo-Ortin, M. (2020) Agency from a radical embodied standpoint: An ecological-enactive proposal. *Frontiers in Psychology* 11, 1–13. https://doi.org/10.3389/fpsyg.2020.01319

Schulte, P. (2023) *Mental content*. Cambridge: Cambridge University Press.

Sganzerla, L., Hutto, D. D., & Kirchhoff, M. D. (2025) Beyond the extended mind: new arguments for extensive enactivism. *Synthese* 205, 118. https://doi.org/10.1007/s11229-025-04931-w

Shapiro, L. A. (2014) Radicalizing enactivism. [Book review] *Mind*, 123 (489), 213–220.

Sprevak, M. (2013) Fictionalism about neural representations. *The Monist* 96, 539–560.

Taylor, S. D., & Williamson, J. (2024) Evidence and cognition. *Erkenntnis* 89, 1927–1948. https://doi.org/10.1007/s10670-022-00611-0

Thompson, E. (2007) *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge: Harvard University Press.

Thompson, E. (2011) Reply to commentaries. *Journal of Consciousness Studies* 18 (5–6), 19–22.

Thompson, E. (2018) Evolving enactivism [Book review], *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/reviews/evolving-enactivism-basic-minds-meet-content/

Thompson, E. (2020) *Why I am not a Buddhist*. New Haven, CT: Yale University Press.

Varela, F. J., Thompson, E., & Rosch, E. (1991) *The embodied mind: Cognitive science and human experience*. Cambridge: MIT Press.

Wheeler, M. (2017) The revolution will not be optimised: Radical enactivism, extended functionalism, and the extensive mind. *Topoi* 36, 457–472.

Wittgenstein, L. (1953) *Philosophical investigations*. Oxford: Blackwell.

Zahnoun, F. (2021a) Some inaccuracies about accuracy conditions. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-021-09771-w

Zahnoun, F. (2021b) On representation hungry cognition (and why we should stop feeding it). *Synthese* 198 (Suppl 1), 267–284.

Zahnoun, F. (2025) Information in the brain: From metaphor to truth. *Theoria*. https://doi.org/10.1111/Theo.70021

Zahidi, K., & Myin, E. (2016) Radically enactive numerical cognition. In Etzelmüller, G., & Tewes, C. (eds). *Embodiment in evolution and culture* (pp. 57–72). Tübingen, Germany: Mohr Siebeck.

Zahidi, K. (2021) Radicalizing numerical cognition. *Synthese* 198 (Suppl 1): S529–S545.

7 Neo-Pragmatism and the natural origins of content

Daniel Martínez Moreno and Manuel Heras-Escribano

7.1 Introduction

Philosophy has long grappled with the nature of thought, objectivity, and human rationality; yet the origins of these capacities seem to remain profoundly puzzling. How did creatures capable of judgment and objective understanding emerge from the natural world? Hutto and Satne (2015) propose a decisive shift in how we approach this question. They argue that "analytic" philosophy's traditional pursuit of naturalizing content—attempting to reduce meaning to purely natural processes—has reached a theoretical impasse. Instead, they advocate for investigating the natural origins of content, a project that acknowledges the distinctive character of rational capacities and objective thought while still seeking to explain its development.

Central to their view is the recognition that objectivity and contentful thought represent *sui generis* achievements—irreducible to the cognitive structures found in other animals, yet nevertheless arising from natural foundations. In other words, they acknowledge (siding with left-wing Sellarsians like Brandom and McDowell) the autonomous structure of the space of reasons. Where reductive naturalists seek to collapse normative phenomena into natural processes, Hutto and Satne aim to understand how such phenomena gradually emerged within nature while retaining their distinctive character.

Hutto and Satne believe that admitting the autonomy of the space of reasons is by no means embracing some kind of spiritualism or non-naturalism.[1] Here, they side with the Aristotelian naturalism of McDowell: Rational capacities such as objective thought express our way of being animals. We are, in fact, rational animals. There is nothing unnatural in us, and this is not in conflict with the autonomy of the space of reasons as long as we do not take a reductive stance toward it.

1 Brandom has claimed on several occasions that his philosophy is non-naturalistic (1994, *passim*). But by this he means that he does not follow any kind of reductive naturalism. He does not embrace any kind of spiritualistic or anti-scientific position.

DOI: 10.4324/9781003649359-9
This chapter has been made available under a CC BY 4.0 license.

However, they also think that it is not enough to take this non-reductive attitude regarding the space of reasons common to left-wing Sellarsians. Contentful thought must not be just *presupposed* but also *explained*. We need an account of the emergence of content in a natural world. And this duty, they argue, is the business of neo-Pragmatism.

In this chapter, we will be concerned with neo-Pragmatism and question of the emergence of content. First, we will make clear Hutto and Satne's understanding of the view. Next, we will discuss the way the emergence of content figures in the philosophical projects of two important advocates of neo-Pragmatism: Robert Brandom and John McDowell. Finally, we will try to extract different ways to look at the relation between neo-Pragmatism, conceptual content, and communal being. To do so, we will make explicit the presuppositions of every perspective within the debate, and we will conclude that it needs to be rethought taking into account the transcendental and empirical aspects at stake.

7.2 Approaching neo-Pragmatism

Neo-Pragmatism is not a univocal approach. Satne (2016) lists thinkers like Hegel, Brandom, Davidson, Wittgenstein, Vygotsky, Piaget, and Tomasello as neo-Pragmatists, while Haugeland (1990) adds Heidegger, Sellars, and McDowell to the list. But when we consider how different these philosophers and scientists are—Heidegger and Tomasello, for example—it is hard to see what makes them all neo-Pragmatists. This raises a basic question: what exactly is neo-Pragmatism?

This question, too, can have many answers. Labels are labels precisely because they do not describe at all. To keep the discussion focused, we will discuss how Hutto and Satne understand neo-Pragmatism. Hutto and Satne (2015) make use of Clapin (2002) for their definition of neo-Pragmatism. According to it, neo-Pragmatists hold that "mental properties are derived from social properties and not vice versa" (Clapin 2002: 17). They go on to assert that, according to neo-Pragmatists, "contentful states of mind develop through processes of engaging in established socio-cultural pratices", and that "we can only make sense of contentful thinking in the context of shared ways of life in which social norm compliance is developed, maintained and stabilized through practices" (Hutto and Satne 2015: 526–527). In a latter paper, Satne has equated this view with a kind of "social constructivism" (Satne 2016: 106).

At this level of analysis, neo-Pragmatism posits a fundamental connection between contentful thought, reason, and objectivity, on the one hand, and communal social practices, on the other. Of course, the nature of this connection can assume various forms, each giving rise to distinct philosophical projects. At this juncture, we just want to emphasize neo-Pragmatism's central claim: The rational capacities essential to our form of life arise only within the context of communal practices, which provide the necessary framework for their development and exercise.

That said, both Hutto and Satne—alongside numerous others—believe that neo-Pragmatism harbors an essential tension. Neo-Pragmatism asserts that contentful thought arises only within social practices. Yet this very possibility presupposes a more primitive foundation—both phylogenetically and ontogenetically—that enables such practices to emerge in the first place. Crucially, participation in these practices demands a form of intelligent thinking that cannot itself be reduced to the product of those very practices. But if neo-Pragmatists confine intelligence to the rational capacities already presupposed by the articulation of the space of reasons, they render themselves unable to account for these antecedent modes of cognition, whether in evolutionary or developmental terms. In their own words:

> The puzzle is this: if all intentionality is of a piece and only derives from social practices, how is it possible that the sort of intelligent, recognitional capacities needed to explain participation in those social practices could be in place prior to their mastery? Unless intentional content is presumed to be already in place this seems impossible.
>
> (Hutto and Satne 2015: 529)

Hutto and Satne propose to resolve these difficulties by distinguishing two forms of intentionality: directness or Ur-intentionality on the one side, and full-blown intentionality or aboutness on the other. By positing a more primitive Ur-intentionality (a basic directedness toward environmental objects without semantic content or aboutness, shared by prelinguistic infants and non-human animals), they aim to provide neo-Pragmatism with a tool for explaining the emergence of contentful intentionality that avoids the tension of how content emerges.

This approach shifts the explanatory burden: rather than circularly justifying normative content-embedded practices, the task becomes tracing how they arise from more basic, causal interactions. As Hutto and Satne put it, the neo-Pragmatist's job is to "give a naturalistic explanation of how normative content-involving practices evolved without trying to justify them in terms of their proposed explanation" (2015: 534). The missing link, then, is an account of how social practices transform Ur-intentionality into full-fledged contentful thought.

7.3 Brandom's three questions

We consider that the previous account of the neo-Pragmatist program is too abstract: different authors can understand the relation between conceptual content and communal beings in different ways. Thus, we need to make our understanding of neo-Pragmatism more determinate. To do so, we will start discussing the philosophical project of Robert Brandom; in particular, we will discuss the way in which the question of the emergence of content figures in his approach.

As a preliminary diagnosis, we contend that Hutto and Satne's critique misfires. Their claim that neo-Pragmatists presuppose a univocal account of intentionality is not the case. For instance, Brandom has explicitly distinguished between discursive and non-discursive forms of intentionality:

> Practical [non-discursive] intentionality[2] is the kind of attunement to their environment that intelligent nonlinguistic animals display—the way they can practically take or treat things as prey or predator, food, sexual partner or rival and cope with them accordingly. Discursive intentionality is using concepts in judgment and intentional action, being able explicitly to take things to be principles.
>
> (Brandom 2011: 10)

For Brandom (2011: 10), the core project of pragmatism is to understand discursive intentionality as a species of practical intentionality. While he acknowledges that practical intentionality can indeed take discursive forms (e.g., mastery of linguistic know-how), he insists this is compatible with an account of its developmental origins. As he clarifies, discursive intentionality may be understood as "having developed out of nondiscursive practical intentionality, while still maintaining that it is a wholly distinctive variety" (Brandom 2011: 10).

The question of content's natural origins is indeed explicitly addressed within this framework (Brandom 2011: 26). Brandom identifies our acquisition of language as the pivotal development in our natural history[3]—an achievement spanning both phylogenetic evolution and ontogenetic development. For him, comprehending this transition demands discussing three fundamental and interconnected issues: demarcation, emergence, and leverage. First, the demarcation problem involves explaining what makes discursive practices essentially different from non-discursive ones. Next, the emergence question requires explaining the possibility of the natural transition from non-discursive abilities—exhibited both by non-linguistic or prelinguistic creatures—to discursive ones. Finally, the leverage problem addresses how to characterize and explain the vast qualitative difference between the capacities of linguistic and non-linguistic creatures.

2 Brandom (2011) characterizes this basic form of intentionality as "practical intentionality." We explicitly qualify it as *non-discursive* practical intentionality to distinguish it from Brandom's own account of discursive intentionality –which itself remains fundamentally practical in nature (being grounded in *discursive know-how*). This distinction preserves the crucial difference between (1) prelinguistic, embodied capacities and (2) the norm-governed know-how specific to linguistic practices.

3 It is at this point that communal being and *sui generis* social normative practices come into play. Moreover, we need to keep in mind that language only matters here as the vehicle of *conceptual content*.

Admittedly, Brandom has occupied himself mainly with the demarcation and leverage questions. *Making it explicit* (*MIE*) begins taking up this very issue:

> What is it we do that is so special? The answer to be explored here—a traditional one, to be sure—is that we are distinguished by capacities that are broadly cognitive. Our transactions with other things, and with each other, in a special and characteristic sense *mean* something to us, they have a *conceptual content* for us, we *understand* them in one way rather than another. It is this demarcational strategy that underlies the classical identification of us as *reasonable* beings. Reason is as nothing to the beasts of the field. We are the ones on whom reasons are binding, who are subject to the peculiar force of the better reason.
>
> (Brandom 1994: 4–5)

The remainder of *MIE* demonstrates how these rational capacities generate entirely new abilities—ones that only become intelligible against the background of reason's own normative structure–, thus answering the leverage challenge.

It is true, however, that Brandom has devoted much less attention to the problem of the emergence of content. There are various, non-accidental reasons that explain this situation, but it will be more informative if we discuss what he has written about this issue.

7.4 Brandom and the question of the emergence of content

Brandom (2011) believes that classical pragmatists such as Dewey made the emergence question too easy at the price of making the leverage one too hard. Those aligned with Hutto and Satne's critique would reverse this charge: neo-Pragmatists successfully articulate how normative practices transform cognition, but at the cost of making their natural origins mysterious. This mirrors Dewey's own diagnosis of a persistent philosophical divide:

> Upon the whole, professed transcendentalists have been more aware than have professed empiricists of the fact that language makes the difference between brute and man. The trouble is that they have lacked a naturalistic conception of its origin and status.
>
> (Dewey 1929: 168)

Dewey sought to reconcile the emergence-leverage tension by emphasizing both the biological continuity across organic life and the qualitative leap represented by language and culture. While insisting on nature's unbroken developmental thread, he nevertheless recognized that linguistic practices institute a new order of complexity. However, his account of this new level was incrementalist and assimilationist in nature, falling short of Brandom's

standards for what constitutes an adequate explanation of the discontinuity between sentience and sapience. Brandom believes that we need an account that roots conceptual capacities in their own *internal norms* without treating their emergence as inexplicable or mysterious (Frápolli 2022).

His treatment of the emergence problem occupies only a few pages within his extensive body of work, making it easy to overlook, but it can be found in *Between saying and doing* (*BSD*). The central focus of *BSD* is on the algorithmic elaboration of discursive practices. While we routinely employ diverse vocabularies—modal, normative, logical, etc.—we are unaware of the abilities that enable such deployment. Brandom's key contribution is to demonstrate how certain vocabularies can be algorithmically elaborated from others. For instance, he shows that the abilities sufficient for wielding modal vocabulary can be algorithmically constructed from those necessary for using ground-level concepts. This illustrates how certain discursive practices and abilities can be algorithmically elaborated from other existing discursive practices and abilities. At this point, an important question prompts: can discursive practices and abilities be algorithmically elaborated from non-discursive ones?

Brandom's (2008: 83) response is negative: discursive practices are not algorithmically decomposable into non-discursive ones. One may think that this makes the emergence question intractable. However, Brandom believes that discursive and non-discursive practices can be related to each other in other ways besides algorithmic elaboration:

> Sometimes those who can engage in one set of practices can learn or be trained to engage in another—not because the target practices can be *algorithmically* elaborated from the original ones, or from some further set into which they can be decomposed, but just because, as a matter of contingent empirical fact concerning creatures of that particular kind, anyone who has the one set of capacities can be brought to have the other as well.
>
> (Brandom 2008: 84)

Brandom defines this second type of practice-ability relationship as "practical elaboration by training". Unlike algorithmic elaboration, this mode operates contingently. The abilities that suffice for tasks like, say, alphabet memorization or face-drawing cannot be determined a priori; they must be identified empirically, being dependent on variable contextual factors. The search for this kind of practical decomposition "is very general and abstract, but also both empirical and important. It is a very general *structural* question about the ability in question" (Brandom 2008: 77).

Brandom himself has not provided the empirical account needed to settle which set of non-discursive practices and abilities suffice, both in phylogeny and ontogeny, for the emergence of discursive practices and, therefore, content. But this is not his job at all. He has distinguished between discursive and non-discursive intentionality, he has provided an account of the necessary and sufficient conditions for something to be a genuine discursive practice

(demarcation), and he has shown how these very conditions are in fact responsible for the kind of rational life that we exhibit (leverage). Finally, and more importantly for our discussion, he has rendered intelligible the possibility of the transition from non-discursive intentionality to propositional thought (emergence). As we can see, this includes already more than what Hutto and Satne grant to the neo-Pragmatist approach.

As shown above, in his discussion of practical elaboration by training, Brandom explicitly states that determining which practices and abilities are sufficient for the emergence of discursive intentionality is an *empirical* affair. This task belongs to scientists from any discipline concerned with life and mind. Researchers like Tomasello and Deacon are examples of scholars addressing this very issue. Philosophy can benefit from these empirical investigations, but it is not forced to *provide* them. In this context, it is enough to ensure that it does not make the natural phenomena studied by science unintelligible. And Brandom does not do so, for he provides an account of the more basic forms of cognition *from which* contentful thought emerges, therefore avoiding any supposedly "essential tension".

Moreover, it is evident that in Brandom's neo-Pragmatist framework, the concept of a discursive (social) practice necessarily operates within the conceptual space defined by both the demarcation question (what distinguishes discursive rationality from more basic cognitive capacities) and the leverage question (how these acquired rational capacities transform the possibilities of thought and action). It would be nonsense to assert that reason emerges from reason through causal dependence. There was a time when rational animals, discursive practices, and communal being did not exist. On the contrary, the point to be made by neo-Pragmatism here is that, once we have access to the concept of a rational being, we find the connection between reason and communal being to be a *necessary* and *internal* one. No one would deny that our species' mode of being had a temporal starting point, but this is not the issue at stake. To understand ourselves as rational animals capable of contentful thought requires thinking of communal being as already at play in the first place. This is not an empirical issue, but a transcendental or constitutive one. We will say more about this in a moment. For now, it is enough to have shown Brandom's account of the question of the emergence of content, thus answering Hutto and Satne's view on the alleged problem of neo-Pragmatism.

7.5 McDowell's exorcism

McDowell, like Brandom, sees our rational capacities as fundamentally rooted in communal forms of life.[4] Moreover, he shares Brandom's view that

4 Though it is important to make clear that McDowell does not believe that rationality is a communally conferred status as it is, for example, being entitled to vote (see McDowell 2013, cap. 9).

responsiveness to reasons both demarcates and elevates our very mode of being:

> We should not be frightened away from holding that initiation into the right sort of communal practice makes a *metaphysical* difference. In this respect achieving free agenthood is quite unlike reaching voting age. Responsiveness to reasons, the very idea of which is inseparable from the idea of communal practices, marks out a fully-fledged human individual as no longer a biological particular, but a being of a metaphysical new kind
>
> (McDowell 2013: 172)

Advocates of the natural origins program would probably accept this with little opposition. Yet they would press: how did responsiveness to reasons emerge in a natural world? This is, by McDowell's lights, a "How possible?" question. In fact, Satne's formulation of the issue explicitly takes this form:

> How is it possible for something such as contentful states of mind to exist in a natural world? How can thoughts be objective, i.e. how can they refer to things beyond themselves? How did the capacity to think objectively emerge in natural history?
>
> (Satne 2016: 105)

McDowell's relationship to this question is complex to articulate. For instance, in *Mind and World* (*M&W*), referring to the issue of how animals possessing the spontaneity of understanding came into being, he states: "That is a perfectly good question. There was a time when there were no rational animals" (McDowell 1996: 123). But immediately after he asserts: "It is true, however, that the good questions we can raise in the evolutionary context come as close as good questions can to the philosophical questions I want to exorcize". (McDowell 1996: 124). How shall we understand this?

M&W constitutes an exercise in philosophical exorcism. Sometimes, in philosophy, we feel the urge to answer questions such as "how is it possible that our thoughts are directed towards the world?", "how is it possible that our words have meanings?", "how is empirical content possible?", etc. McDowell holds that the philosophical anxieties implicit in this these questions should not be contested, but dissolved. If we attempt to answer, we engage in *constructive* philosophy, and this is just what McDowell seeks to avoid. Would attempting to explain the natural origins of content constitute an exercise in constructive philosophy?

The answer to this question depends on our overall theoretical project:

> If we do speculate about how animals might have evolved into a way of living that includes initiating their young into a culture, we must

> be clear that that is what we are doing. It would be one thing to give an *evolutionary account* of the fact that normal human maturation includes the acquisition of a second nature, which involves responsiveness to meaning; it would be quite another thing to give a *constitutive account* of what responsiveness to meaning is. I have been granting that it is reasonable to look for an evolutionary story. This is not a concession to the sort of constructive philosophical account of meaning that I discussed in my last lecture.
>
> (McDowell 1996: 124, our emphasis)

The crucial distinction here lies between evolutionary and constitutive accounts of the space of reasons. Evolutionary accounts are valid here provided we do not mistake them for constructive philosophy. The space of reasons, according to McDowell, is in no need of an external foundation. Contentful thought needs no justification external to the practice of thinking itself. To concede to constructive philosophy here would require philosophers to address content's origins in justificatory terms, effectively demanding they secure the space of reasons through external validation. The natural origins program requires clarification: is its aim to provide *empirical* accounts (evolutionary and developmental) of how our cognitive capacities arose, or to offer *philosophical* explanations for how thought can exist within nature at all?

This distinction is of great importance, for it is clear that McDowell believes that philosophy can only offer a *constitutive* account of the space of reasons. The image here, according to McDowell, is Neurath's, in which a sailor reconstructs his boat while it is still afloat: once we are initiated in the practice of thought, we can only take a stance toward it from the inside. The modern difficulty of situating reason within nature fuels our anxious search for natural foundations, yet we need neither share this anxiety nor erect philosophical constructs to secure reason's place in nature. Here, Aristotle's naturalism shows a valuable lesson: we are, at root, rational animals, creatures whose capacity for second nature opens the normative realm of virtue and reason. This fact holds no intrinsic mystery unless we artificially impose one. As such, the evolutionary and developmental explanations of the fact that we are endowed with a second nature are available. Science's own conditions of possibility impose that this is so, for nothing comes out of nothing. But these explanations must not aim at grounding the space of reasons in facts external to it. Being responsive to reasons means inhabiting a practice whose authority needs no grounding beyond itself. Once we participate in this practice, questions about its chronological origins become meaningless, for we find no matter of fact about its starting point. Like Neurath's sailor at sea, we work with what floats, not with shipbuilding records. The boat's history fades; only its current seaworthiness matters. Both the sailor and philosophy itself are bounded to constitutive questions, not factual-chronological ones.

McDowell's attitude toward the natural origins question has been clarified up to this point: the empirical account of the evolution of our cognitive repertoire must be distinguished from philosophical inquiries into the internal norms of the space of reasons. In any case, empirical investigations shall not be taken as answers to philosophical anxieties. It is evident that there was a time when no rational animals and no thought existed. We know that we have come to being through evolution, so science must render intelligible the empirical facts that lie behind this transition. But these facts, though prior in time, do not constitute the *foundations* of the space of reasons in any recognizable sense. McDowell's claim that thought becomes intelligible solely through its internal norms serves as a deliberate counter to those craving external foundations. Ultimately, advocates of the natural origins program must reflect on the status of their claims.

7.6 Neo-Pragmatism and conceptual content: a clarification

Thus far, we have considered the way in which the question of the natural origins of content figures in the philosophical projects of Robert Brandom and John McDowell. Brandom has argued that any pragmatist must consider three interrelated issues: demarcation, emergence and leverage. He has devoted himself almost all his efforts to the questions of demarcation and leverage. However, he has provided the philosophical resources necessary to make intelligible the emergence of content in a natural world. To begin, he has offered an account of non-discursive practical intentionality in terms of feedback loops between organisms and their environments; further, he has argued that the transition to full-fledged propositional intentionality arises through practical elaboration by training on the basis of these environmental transactions, thereby overcoming any causal tension. What is still required—and this falls to the sciences of life and mind, not to philosophers—is a complete empirical account of the set of non-discursive practices and abilities that suffice, both evolutionarily and developmentally, for the emergence of discursive ones. For his part, McDowell has been careful to distinguish between (i) evolutionary accounts of the fact that our normal maturation includes the possession of a second nature and (ii) constructive accounts that feel the urge to look for the foundations of the space of reasons in facts external to itself. In his view, philosophy's business shall not be the *empirical explanation* of the emergence of the space of reasons. Philosophy must exorcise the dilemmas that make it appear as if we must secure reason's place in nature. That we are rational animals is a fact that we encounter, not something to be reconstructed from the materials of the realm of natural law to relieve our philosophical anxieties.

We must now analyze the neo-Pragmatism's thesis that communal engagement is a necessary condition for conceptual content and objective thought. Recall that advocates of the natural origins program believe that here lies

neo-Pragmatism's essential tension: If social practices must be necessarily in play for there to be conceptual content, how are we to understand the cognitive capacities needed to engage in those very social practices? Hutto and Satne's solution is quite flat: Intentionality is not of a piece. The postulation of contentless Ur-intentionality constitutes a necessary condition for both breaking the causal circle and doing justice to prelinguistic forms of cognition. But we have seen that another concept of intentionality which does not presuppose content, non-discursive practical intentionality, is available for neo-Pragmatists. So Hutto and Satne's critique does not stand up. The neo-Pragmatist acknowledgement of non-discursive intentionality is, in fact, a negative move: it aims to secure scientific intelligibility. Philosophy cannot make intractable perfectly valid scientific questions about the transactions of organisms with their environments. Those very questions *presuppose* that the organism's activities are *directed* toward places, objects, and events. Proponents of the natural origins program maintain, however, that neo-Pragmatism cannot rest here: what is required is a robust account of the *factual determinants* enabling the transition from non-discursive to discursive intentionality. But both Brandom and McDowell have argued that this empirical story is not philosophy's business. Moreover, it would be naïve to assume that merely establishing a relation between Ur-intentionality and contentful intentionality could yield the required answers. The empirical framework requires much more than this.

All the above does not mean that neo-Pragmatism does not have a positive story to tell. It does. But this story necessarily presupposes talk of discursive practices, communal being, and second nature. We now want to delve into this issue in order to clarify the remaining points of the debate of the natural origins of content.

After discussing McDowell's ideas on the constitutive approach to the space of reasons, it is easy to see that the relation between communal being and conceptual content is not a matter of mere chronological causation. Brandom's insistence on demarcation and leverage for understanding the nature of conceptual content presupposes this same point: that the relation between communal being and conceptual content is an internal one. Both Brandom and McDowell are concerned with the internal logic of the space of reasons, and this logic shows us that we cannot make sense of conceptual content without communal practices being *already* in play. Brandom's image[5]

5 A lot more could be said about the way in which communal practices figure both in Brandom and McDowell, but this would take us too far outside the argument of this chapter. The crucial point is that, in one way or another, communal practices are always *presupposed*.

here is that, in applying and instituting conceptual norms, we always find conceptual contents to be already determinate[6]:

> [...] we must understand linguistic practices as *both* instituting conceptual norms *and* applying them. It is precisely by applying concepts in judging and acting that conceptual content is both made more determinate, going forward, and shows up as always already determinate (in the only sense in which conceptual contents are determinate), looking back.
>
> (Brandom 2011: 26)

But this is not to say that contentful intentionality is created by linguistic practices, for it is obvious that linguistic practices are already determined in this same sense. If we approach this issue looking for chronological relations, we will not get it right. Contentful intentionality is not created by anything. It would be, in fact, a category error to assert this. When we describe linguistic practices as both applying and instituting conceptual norms, we are articulating the internal logic of the space of reasons, not mapping its causal structure. Crucially, the very institution or application of a conceptual norm is unintelligible without normative statuses like responsibility and authority—*social* statuses that depend on subjects adopting the practical attitudes of holding one another accountable and recognizing each other as authoritative (see Brandom 2019 for more details on this point). But, again, the determination of conceptual content through normative attitudes and statuses resists temporal framing. To ask which comes first—statuses or attitudes, practices or content, and conceptual capacities or second nature—is already to misapprehend the constitutive logic at work. For Brandom and McDowell, these are not merely difficult questions but conceptual misfires. We must protest their status as *intelligible questions*.

The lesson to be taken here is that rational life comes with its own structure, one that makes the connection between communal being and conceptual content an internal one. Communal practices do not *produce* conceptual content any more than the categories of understanding *create* experience. Rather, experience's own internal norms, once made explicit, reveal that there could not be determinate experiences without the categories. Those who follow Hutto and Satne's logic would feel the urge to avoid the circularity,[7] but

6 Accordingly, a fictitious Ur-member of the space of reasons would find conceptual content as already determinate. But this shows that these are not the questions we need to make ourselves. And, in the case that we make them, we must be aware that their role is negative: to show the limits of what can be intelligibly asked.

7 The history of philosophy after Kant is full of attempts to avoid the circularity by postulating the categories as innate mental structures. These moves show an absolute misunderstanding of the issue at stake.

it is unavoidable. Constitutive explanations work like this, and there is nothing mysterious about them.

Thus, neo-Pragmatism's fundamental insight is clear: what links communal practices to conceptual content is its constitutive structure, not its temporal sequence. Naturally, neo-Pragmatists may diverge in how they articulate this constitutive structure. Brandom and McDowell themselves disagree, for instance, on the origins of conceptual content's normative force. But these disagreements concern the internal norms of the space of reasons, not its external determination. This is why both Brandom and McDowell cannot answer *positively* the question of the natural origins of content. Yet, once we have laid out their philosophical commitments, it becomes clear that this was never their aim. The anxieties that Hutto and Satne want us to feel if we side with neo-Pragmatists are optional.

7.7 Conclusion

In this chapter, we have tried to make explicit the presuppositions within the question of the natural origins of content. We have centered our discussions in Brandom and McDowell for two reasons: First, because Hutto and Satne have not analyzed rigorously their attitude toward this crucial question; second, because this attitude reveals that Hutto and Satne's anxieties stem from a misunderstanding.

First, we have shown that neo-Pragmatism neither requires nor possesses a univocal conception of intentionality: Non-discursive—therefore non-contentful—forms of intentionality are available. At this juncture, it is legitimate to ask what additional value the concept of Ur-intentionality actually provides. We believe that it is nothing more than a reformulation of the concept of non-discursive intentionality, which has been at hand all the time. Thus, Hutto and Satne's first bullet has not reached its target. Next, we have briefly reconstructed Brandom's and McDowell's answers to the question of the emergence of content. The crucial point has been to show that their neo-Pragmatism is not concerned with the chronological determination of conceptual content, but with its internal, immanent norms. Brandom has made clear that the empirical question of the emergence of discursive practices and abilities from non-discursive ones is an important one; and McDowell, while acknowledging the empirical validity of evolutionary accounts, has warned us about their philosophical misuse—therefore his exercise. But both have made clear that philosophy's business is not to answer these positive, factual questions. With this in mind, we have argued that neo-Pragmatism's insistence on the link between conceptual content and communal practices is not vulnerable to the charge of causal circularity: This thought arises from a misunderstanding between chronological and constitutive accounts of the space of reasons. With this, we believe that all the criticisms of the advocates of the natural origins program have been contested and shown inadequate. We hope to have clarified the philosophical credentials of the question of the

natural origins of content, as well as the relationship between communal practices and conceptual content in neo-Pragmatism. But, above all, we hope to have alleviated certain philosophical anxieties that stem from the urge to answer inadequate questions.

References

Brandom, R. (1994) *Making it explicit: reasoning, representing, and discursive commitment*. Harvard University Press.

Brandom, R. (2008) *Between saying and doing: towards an analytic pragmatism*. Oxford University Press.

Brandom, R. (2011) *Perspectives on pragmatism: classical, recent & contemporary*. Harvard University Press.

Brandom, R. (2019) *A spirit of trust: a reading of Hegel 's Phenomenology of spirit*. Harvard University Press.

Clapin, H. (2002) Introduction, in H. Clapin (ed.) *Philosophy of mental representation*. Oxford University Press.

Dewey, J. (1929). *Experience and nature*. W. W. Norton & Company. https://doi.org/10.1037/13377-000

Frápolli, María J. (2022) Tracking the world down, *Philosophical Topics* 50(1), pp. 83–107.

Hutto, D. & Satne, G. (2015) The natural origins of content, *Philosophia* 43, pp. 521–536.

McDowell, J. (1996) *Mind and world*. Harvard University Press.

McDowell, J. (2013) *Having the world in view: essays in Kant, Hegel, and Sellars*. Harvard University Press.

Satne, G. (2016) A two-step theory of the evolution of human thinking, *Journal of Social Ontology* 2(1), pp. 105–116.

8 Eliminativism is at the heart of E-cognition

Luis H. Favela

8.1 Introduction

What exactly defined 20th-century analytic philosophy is controversial (Glock, 2008; Preston, 2007). Still, there were trends, such as emphasizing formal logic (e.g., logical atomism) and the nature of language (e.g., linguistic turn). One prominent facet of analytic philosophy in the 20th century was the attempt to treat philosophy as continuous with science (e.g., Quine, 1960/2013). This movement was due in no small part to the influence of the logical positivists (or logical empiricists; Stadler, 2008). While there was no single adhered to doctrine observed by all members (e.g., Rudolph Carnap and Moritz Schlick), one core aim of early logical positivism was to offer a way to determine if claims are meaningful (Creath, 2023). To that end, be it a "philosophical" or "scientific" claim, in order to determine if it is meaningful, the claim must either be true by way of analyticity (or, true by definition; e.g., a "bachelorette" is defined as "an unmarried woman") or empirically verifiable in practice or in principle (e.g., "there are mountains on the farther side of the moon", could be verified by a rocket being sent either actually or in principle; Ayer, 1952). It was hoped that by taking on this "scientific conception of the world" (Neurath, 1973), a method would be offered to turn away from or outright reject metaphysical theorizing. According to the logical positivists, metaphysical claims—such as those purporting to concern ethics and theology—were not susceptible to the methods of analyticity or empirical verification and, thus, were *meaningless*. Consequently, such claims should be *eliminated* from discourses that attempt to make progress on our understanding of what the world is really like.

The idea of eliminating meaningless concepts in these ways was especially impactful on the history and philosophy of science and philosophy of mind. Regarding the former, elimination was understood as playing crucial roles in scientific progress, for example, during intertheoretic reduction (e.g., phlogiston; Churchland & Churchland, 1998). It is arguable that the attempt to eliminate concepts played its largest role in the philosophy of mind, where it was commonplace to utilize concepts with shaky definitions and to study phenomena that resisted empirical verification, both in practice and potentially

DOI: 10.4324/9781003649359-10
This chapter has been made available under a CC BY 4.0 license.

in principle (e.g., qualia; Dennett, 1993). It is fair to say that the majority of philosophers who employed eliminative strategies with regard to mind and related phenomena were also proponents of brain-centric explanations of said phenomena (e.g., Bickle, 1998; Churchland, 1994; Rorty, 1965; Stich, 1983). Specifically, philosophers of mind who were interested in what the sciences have to contribute typically defended forms of identity theory—e.g., mental state 1 *just is* brain state X—or reductive theories—e.g., mental state 1 is *explained by* brain state X—that treated minds as in some essential way brain-centric phenomena.[1]

It is uncontroversial to state that strategies that can reasonably be called "eliminative" were common among 20th-century analytic philosophers who defended various ways of understanding minds as brain-centric phenomena. However, what is controversial is the main claim defended in the current work: *Eliminativism is at the heart of e-cognition*. In short, "e-cognition" refers to a range of approaches to studying cognition that, at a minimum, place more emphasis on the role of the body than typical brain-centric approaches do (e.g., Jovanov, Clifton, Mazalek, Nitsche, & Welsh, 2015) and, at a maximum, understand cognitive phenomena as necessarily spanning brain-body-world systems (e.g., Chemero, 2009).[2] The primary reason that this claim is controversial is due to the fact that proponents of eliminative approaches—such as those mentioned above (e.g., Bickle, Churchland, and Dennett)—are typically viewed as endorsing views about cognition (e.g., computationalism and representationalism) that are radically contrary to those endorsed by some proponents of e-cognition (e.g., direct perception and dynamicism). As a result, it seems quite strange—at least at first blush—to associate a strategy so commonly attributed to one group (i.e., eliminativism by brain-centric analytic philosophers) as central to another group that tends to hold diametrically opposed views. In order to motivate the thesis that eliminativism is at the heart of e-cognition, the next two sections provide brief overviews of eliminativism and e-cognition. After, examples of explicit and implicit eliminativist moves made in the e-cognition literature are presented. Last, it is argued that proponents of e-cognition ought to be eliminativists and should continue to be so explicitly, especially regarding contemporary work in artificial intelligence (AI) and neuroscience.

1 On the other hand, there were analytic philosophers who championed non-brain-centric approaches to mind. These ranged from proponents of multiple realization—e.g., functionalism (Putnam, 1975)—to nonreductive views—e.g., emergence (Kim, 1993)—that did not treat minds as necessarily being located in or produced by brains. Such approaches are not central to the current work as they tend not to employ eliminative strategies (e.g., Chalmers, 1996).

2 Here, "cognition" is intended to be a catchall term that includes related or synonymous phenomena such as goal-directed behavior, mental states, mind, and the like.

8.2 Eliminativism

This section expands on the brief discussion of eliminativism above, with an emphasis on its application to mind (for a more thorough overview of eliminativism, see Ramsey, 2024). As mentioned in the previous section, with regard to 20th-century analytic philosophy, eliminativism can be viewed as originating with the logical positivists and their strong opposition to metaphysics (Carnap, 1932/1959). Claims can be understood as meaningful if they meet one of two criteria: either they are analytic (i.e., true by definition) or they are empirical (i.e., they can be verified by empirical data either in practice or in principle). Metaphysical claims—inclusive of aesthetic, ethical, and theological statements—are considered *meaningless* because they resist both criteria. If a claim is meaningless, then it should be eliminated from use in work aimed at understanding how the world really is.

There is no doubt that logical positivism had an immeasurable impact on 20th-century analytic philosophy. Still, the movement was quickly met with considerable challenges, not least of which included the following: First, one of their foundational works was seriously undermined by its author. Specifically, while the logical positivists were greatly influenced by the *Tractatus Logico-Philosophicus* (Wittgenstein, 1921/2001), the author, Ludwig Wittgenstein, would critique many, if not all, of the main claims in his later work (e.g., Wittgenstein, 1958). Second, numerous challenges were made to verificationism, such as the idea that it is self-undercutting. Specifically, the claim that "meaningful statements are either analytic or empirically verifiable" is not itself a claim that is analytic or empirically verifiable (Creath, 2023).

With all that said, the general idea of eliminativism persisted in two main forms. One is via work in the history and philosophy of science. Here, progress in the history of science is understood as sometimes occurring by way of elimination. For example, Patricia Churchland and Paul Churchland (1998) describe a number of forms of scientific progress: some occur when one theory is displaced by way of being reduced to a more encompassing theory—such as Newton's laws of motion being intertheoretically reduced to Einstein's special theory of relativity—while others occur when a theory is displaced by way of being eliminated—such as phlogiston being totally replaced by Lavoisier's oxygen theory of combustion.[3] The other form in which eliminativism persisted is via the philosophy of mind. Here, progress on theorizing about minds is understood as sometimes occurring by way of elimination. For example, Paul Churchland (2005) describes what Elizabeth Irvine and Mark Sprevak (2020) call "discourse eliminativism" by way of explaining subjective color qualia in terms of neuronal coding vectors. In this case, there

3 It is worth acknowledging that the Churchlands' story about phlogiston is not accepted by all (e.g., Ludwig, 2014).

is no attempt to eliminate the experience of subjective color qualia, but its explanation is offered in terms of neuronal activity. Daniel Dennett's (1993) critique of the concept of "qualia" offers a rich example of what Irvine and Sprevak call "entity eliminativism" (2020). In this case, there is an attempt to eliminate the notion of qualia from philosophical theorizing. The idea, in short, is that because each of the criteria for qualia are fundamentally flawed, then what those criteria are purported to refer to does not exist.[4]

It is safe to say that eliminativism peaked by the late 20th century. Even so, eliminativism continues to be present in contemporary philosophy of mind, such as the recently offered illusionism, which claims that phenomenal consciousness is an introspective illusion (Frankish, 2016). Illusionism is eliminative in both the entity sense (e.g., denies that "phenomenal properties" exist) and the discourse sense (e.g., there are forms of consciousness that are not properly characterized as having phenomenal properties; Frankish, 2016). So, while eliminativism may not have the controversial pop it once had (e.g., Churchland, 1981; Hardcastle, 1999), it remains, minimally, a position worth acknowledging in histories of analytic philosophy or, maximally, an approach that has taken on new forms (e.g., illusionism). Perhaps surprisingly, an underappreciated area of research where eliminativism is thriving—or so the current work claims—is in the e-cognition literature. The next section offers an overview of e-cognition and is followed by examples of eliminativism in such work.

8.3 E-cognition

There is an increasingly long list of what counts as an "e" in "e-cognition". For that reason, this section does not aim to provide anything close to a comprehensive overview.[5] Instead, it has the more modest aim of providing a flavor of "e-cognition" for uninitiated readers. To start, e-cognition can be understood as first appearing in the form of "4E cognition", where "E" refers to embodied, embedded, enacted, and extended cognition (Rowlands, 2010). *Embodied cognition* refers to a range of views. At one end, cognition remains a brain-centric phenomena but recruits brain regions commonly understood as being involved in motor control and perception during "higher-order" cognitive tasks, such as abstract reasoning. At the other end, cognition is not brain-centric, but instead spans nonneural physiology, such as representing numbers with one's fingers during counting. *Embedded cognition*, sometimes synonymous with situated cognition, refers to the idea that cognition is caused and/or constituted by a body and environments, which is inclusive

4 It is important to make clear that Dennett's (1993) eliminativism about qualia does not mean that he is an (entity) eliminativist about minds and what can be called "mental experiences".

5 Helpful resources for understanding what "e-cognition" refers to include Anderson (2003); Calvo and Gomila (2008); Chemero (2009); Coelllo and Fischer (2016); Favela (2024); Menary (2010); Newen, de Bruin, and Gallagher (2018); Rowlands (2010); and Varela, Thompson, and Rosch (1991).

of everything from ambient light to social interactions. *Enacted cognition*, at its most basic, typically claims that cognition emerges from sensorimotor activity. The longer story is that "enacted cognition" is commonly synonymous with enactivism (e.g., Varela, Thompson, & Rosch, 1991), which also observes the crucial role of sensorimotor activity in cognition, but incorporates other commitments, such as the idea that life and mind are continuous (Thompson, 2007). *Extended cognition* is the idea that cognition is caused and/or constituted by features external to the commonly accepted boundaries of organisms (e.g., feathers, scales, and skin). As such, nonbiological tools can be part of cognitive systems, including artifacts like canes and smartphones.

These four "e's"—five, if you include "enactivism" as distinct from "enacted", not to mention the additional "s" provided by "situated"—can be challenging for the newly initiated to get their head around. Unfortunately (or fortunately), there are more "e's" and other letters as well. Other letters include "d" for *distributed cognition*, which—much like embedded cognition—understands cognition as being spread across brains, bodies, and environments, such as the crew (e.g., navigators) and equipment (e.g., engine controls) on the bridge of a ship while it traverses bodies of water. Other "e's" include *ecological*, which is primarily associated with ecological psychology (Gibson, 1986/2015) and refers to an approach to perception that treats it as continuous with action and involves the organism-environment system (Segundo-Ortin & Raja, 2024). *Emotional cognition*, *affective cognition*, and others have also been thrown into the mix, not to mention *dynamical* approaches (Favela, 2020).

For the remainder of the current work, when "e-cognition" is referred to, it is not to invoke any particular combination of the above-mentioned forms of cognition (e.g., embodied, emotional, and distributed). Instead, it refers to two of the more noteworthy lessons to be extracted from the relevant literature. First, whatever goes before "-cognition" plays a rhetorical role in drawing attention to often underappreciated or ignored phenomena that are causally and/or constitutively relevant to cognition. To invoke embodied cognition, for example, is to draw attention to the fact that even if one is a neuroreductionist about cognition, brains are *always* in bodies and those bodies are important (e.g., they enable brain activity). Invoking emotional cognition is, among other reasons, to push back on the hard line drawn between reason and emotion. Thinking about cognition as extended motivates expanding the purview of what is relevant to investigating and understanding particular cognitive phenomena. Taken together, e-cognition motivates going outside of brains to take a more encompassing view on what is causally and constitutively relevant to cognitive phenomena.

Second, and although not applicable to all forms of e-cognition (e.g., brain-centric versions of embodied cognition), is the lesson that *cognitivism*, the mainstream way to understand cognition, is not the only game in town. That is to say, cognition does not have to be understood as consisting of computations and/or representations, nor as being exclusively located in brains. Undoubtedly,

cognitivism is the mainstream understanding of cognition observed across the cognitive, neural, and psychological sciences, as well as the philosophy of mind. It is essentially a kind of information processing, where cognition—broadly construed to include everything from the control of bodily movement to solving complicated mathematical problems and imagining a story—involves computations and representations (e.g., Koch, 1999; Neisser, 1967/2014; Thagard, 2023; Von Eckardt, 1995). Computations are the processes of cognition and representations are the objects that are computed. E-cognition calls all of these assumptions into question: "Must cognition happen in brains"; "is cognition a computational process"; and "does cognition necessarily involve representations?" The answer to all three is minimally "not necessarily" and maximally "no". Embodied cognition (of the radical flavor) and ecological psychology offer powerful arguments for undermining the need to appeal to representations for all cognitive phenomena—that is, they are antirepresentational. Distributed and embedded cognition offer compelling reasons to dethrone brains as the locus of cognition—that is, they are non-brain-centric. Enacted/enactivist and dynamical approaches offer plausible alternatives to understanding cognition as a form of computation—that is, they are noncomputational.

This section had two aims. The first was to offer a brief overview of what is meant by "e-cognition", such as embodied cognition, ecological psychology, and dynamicism. The second aim was to offer a sample of the more noteworthy lessons to be taken from the e-cognition literature, particularly that cognition can be fruitfully understood by way of non-brain-centric, antirepresentational, and noncomputational commitments. With the groundwork laid by the previous two sections, the following section motivates a defense of the main thesis of this work, that eliminativism is at the heart of e-cognition.

8.4 The eliminativist heart of E-cognition

In a previous section (Section 8.2), it was stated that a surprising and underappreciated area of research where eliminativism is thriving is in the e-cognition literature. It is surprising because history demonstrates that eliminativist strategies have been predominantly applied in the service of brain-centric understandings of cognition, particularly in philosophy of mind in 20th-century analytic philosophy. It is underappreciated because, as will now be shown, proponents of e-cognition regularly make eliminativist moves in their work, even if they have not referred to them as such. To that end, three examples of eliminativist moves in the e-cognition literature are offered.

8.4.1 Ecological psychology and perception-action

Gibsonian ecological psychology can be boiled down to four primary principles (Favela, 2024): First, *perception is direct*, is the idea that an organism's perceptual capacities can make noninferential contact with its environment

in order to detect ecological information, which are patterns of energy that uniquely specify properties of the world and are temporal in nature. Second, *perception and action are continuous*, is the idea that an organism's perceptual abilities were evolutionarily selected to guide action; conversely, action abilities were selected to enable perception. Third, as a consequence of the previous two, detected ecological information can specify meaningful opportunities for action, or *affordances*. Fourth, the *organism-environment system* is the proper spatiotemporal scale for investigating and understanding the previous three.

If any part of ecological psychology is to be viewed as eliminativist, it would likely be the first principle that perception is direct because of its anti-representational flavor. While that is reasonable, an underappreciated application of eliminativism is to be found in the second and fourth principles. Across his two most popular books, Gibson (1966; 1986/2015) dedicated significant amounts of space to attempting to dissolve commonly accepted dichotomies, or "dualisms", as he put it (1986/2015, p. 108). Examples include mental sensations and physical movements, observer and environment, and subjective and objective, to name a few. In this fashion, Gibson applies both discourse and entity eliminative strategies, with the latter following from the first. As Irvine and Sprevak put it, the "discourse eliminativist seeks to rid science of certain ways of talking, thinking, and acting" (2020, p. 349). Put that way, Gibson clearly aims to rid science—at least perceptual psychology—of ways of talking and thinking that assume dualisms like minds as distinct from bodies. Entity eliminativism is a consequence of Gibson's discourse eliminativism in that rejecting those dualisms eliminates kinds of ontologies (i.e., perception *and* action, organism *and* environment) in favor of terminology that highlights continuities (i.e., perception-action, organism-environment). Taken together, three of the four primary principles are readily viewed as eliminativist strategies.

8.4.2 Radical embodiment and representations

Radical embodiment is summarized by Andy Clark as centering on the following thesis:

> Structured, Symbolic, Representational and Computational views of cognition are mistaken. Embodied cognition is best studied using non-computational and non-representational ideas and explanatory schemes involving e.g. the tools of Dynamical Systems Theory.
>
> (Clark, 1997, p. 461)

Anthony Chemero (2009) further developed the thesis into a research program, radical embodied cognitive science, which integrates much of the theory of Gibsonian ecological psychology with the methods of nonlinear dynamical systems theory (DST). Collectively, principles of ecological

psychology (e.g., affordances) and dynamical methods offer a rich framework for investigating and understanding perception(-action) without need to appeal to representations. Not appealing to representations to account for perception(-action) has led some to claim that "Chemero explicitly endorses eliminativism by rejecting the traditional assumption that agents solve problems and navigate through the world by consulting mental representations" (Ramsey, 2024, Section 3.2.3). It is not clear that Chemero is an eliminativist in any form.[6] However, it is clear that practicing radical embodied cognitive science minimally permits discourse eliminativism. In particular, radical embodied cognitive science offers a theoretically valid and empirically supported framework to investigate and understand perception(-action) without needing to appeal to talk of mental representations.

With that said, while Chemero and other proponents of radical embodied cognitive science could be comfortable with understanding their approach as being motivated to some degree by discourse eliminativism, there is room to see entity eliminativism in the framework as well. It is not just that representations can be eliminated from *talk* of perception(-action), the fact is that the relevant experimental work makes no use of representations. For example, there is a rich empirical literature on affordances that makes no appeal to representations in neither the process of experimental design nor discussions and interpretations of their findings (for review, see Blau & Wagman, 2023; Lobo, Heras-Escribano, & Travieso, 2018; Segundo-Ortin & Raja, 2024). That most certainly seems to be an approach that at least implicitly eliminates representational entities.

8.4.3 Dynamicism and computationalism

Many e-cognitive approaches have embraced DST to some degree, such as enactivism (Varela, Thompson, & Rosch, 1991) and ecological psychology (Turvey, Shaw, Reed, & Mace, 1981) and, of course, radical embodied cognitive science (Chemero, 2009). All three of these examples have utilized DST as a methodology in support of another theory, such as integrating DST in affordance experiments. Nevertheless, DST has been appealed to by other proponents of e-cognition as offering theoretical commitments as well. The dynamical hypothesis offers both a methodology (i.e., DST) and a theoretical understanding of cognition. It claims that cognitive organisms are and can be understood as dynamical systems (Favela, 2020; van Gelder, 1995). At the

6 In recent work, Chemero (2026) neither explicitly nor implicitly endorses any form of eliminativism, especially about minds. He describes his view as being in line with Dennett's, in particular, a view that follows from observing the intentional stance, "that having a mind is a matter of being the sort of thing whose behavior can be explained by the attribution of mental states."

core of that claim is the view that natural cognitive systems are temporal and continuous in nature.

Such an approach is straightforwardly contrary to cognitivism, namely, the position that cognition is an information processing phenomenon based on computations operating over representations. Consequently, the dynamical hypothesis is as close to being explicitly eliminativist about computational approaches to cognition as one can be without saying as much. DST offers a methodology to facilitate discourse eliminativism. Experiments and results can be designed and interpreted without necessitating computational vocabulary. Instead, such vocabulary is eschewed in favor of a dynamical one. Moreover, and in line with the dynamical hypothesis, DST offers a theory to facilitate entity eliminativism, where cognitive phenomena are viewed not as constituted by computational processes but by dynamic ones.[7]

8.5 Conclusion

The previous section offered three examples of eliminativist moves in the e-cognition areas of ecological psychology, radical embodiment, and dynamicism. All three exhibited discourse eliminativism by way of critiquing one way of talking about topics of research interest that are more fruitfully approached by different vocabulary, such as ecological psychology eschewing dichotomous talk in favor of terminology that stresses continuity. All three also exhibited entity eliminativism by way of doing away with terminology that does not refer to the nature of the actual phenomena, such as dynamicism rejecting computationalism about natural cognitive systems in favor of DST. In none of this literature is the work referred to as "eliminativist". As claimed above (Section 8.4), this is likely due to the history of usage of eliminativist strategies in the service of brain-centric understandings of cognition. Consequently, and as argued in this work, it is surprising that non-brain-centric frameworks have provided the rich soil for eliminativism to continue to flourish.

Up until this point, the current work has been largely descriptive, appealing to e-cognition research that can readily be described as employing eliminativist strategies. In closing, a prescriptive claim is offered: Proponents of e-cognition ought to be eliminativists and should continue to be so explicitly. The *ought* of explicit eliminativism stems from the *is* of the place of e-cognition in the cognitive, neural, and psychological sciences, as well as the philosophy of mind. The fact is that non-e-cognition research programs dominate scientific and philosophical understandings of cognition. We are well into the 21st century and the science literature is largely dominated by both

7 The dynamical hypothesis has faced challenges. For example, it has been argued that DST is not a replacement for computationalism, but merely supports it (e.g., Eliasmith & Anderson, 2003). Others have argued that DST is a form of computationalism itself (e.g., Sussillo, 2014).

metaphorical and analogical understandings of brains as computers (e.g., Marcus, 2015; Sprevak & Colombo, 2019). The recent attention garnered by "AI" is proliferating this view.[8] The media abounds with claims of generative AI exhibiting behaviors formerly attributed only to natural cognitive systems, such as creativity and hallucinations (e.g., O'Brien, 2023). Commitments to brain-centrism and cognitivism will continue to strengthen alongside increased research in "NeuroAI" (Laird, Lebiere, & Rosenbloom, 2017; Zador et al., 2023). As a result, e-cognition remains the minority approach to cognition. One strategy is to build bridges by showing how e-cognition and brain-centric/cognitivist approaches can be integrated. Another strategy is to show how e-cognition and brain-centric/cognitivist approaches are radically different, and that the former is better suited to the investigation and understanding of cognition. Given its minority status, proponents of e-cognition ought to take the stronger position: highlight the radical differences and eliminate misdirected talk (i.e., discourse eliminativism) and adherence to nonexistent entities (i.e., entity eliminativism). To proponents of e-cognition, eliminativism has always been at the heart of what you do. Embrace it.

References

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence*, *149*(1), 91–130. https://doi.org/10.1016/S0004-3702(03)00054-7

Ayer, A. J. (1952). *Language, truth and logic*. New York: Dover.

Bickle, J. (1998). *Psychoneural reduction: The new wave*. Cambridge: The MIT Press.

Blau, J. J. C., & Wagman, J. B. (2023). *Introduction to ecological psychology: A lawful approach to perceiving, acting, and cognizing*. New York: Routledge.

Calvo, P., & Gomila, A. (Eds.). (2008). *Handbook of cognitive science: An embodied approach*. San Diego, CA: Elsevier.

Carnap, R. (1932/1959). The elimination of metaphysics through logical analysis of language. In A. J. Ayer (Ed.). *Logical positivism* (pp. 60–81). New York: The Free Press.

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

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge: The MIT Press.

Chemero, A. (2026). *Intertwined creatures: The embodied cognitive science of self and other*. New York: Columbia University Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, *78*(2), 67–90. https://doi.org/10.2307/2025900

Churchland, P. M. (2005). Chimerical colors: Some phenomenological predictions from cognitive neuroscience. *Philosophical Psychology*, *18*(5), 527–560. https://doi.org/10.1080/09515080500264115

8 "Artificial intelligence" is put in quotes here because it is unclear that generative AI, such as large language models (e.g., ChatGPT), is "intelligent" in any way resembling the foundational goals of computer science in the mid-1900s. Such forms of "AI" seem more like fancy ways to do regression to the mean than intelligence. But that is an argument for another day.

Churchland, P. M., & Churchland, P. S. (1998). *On the contrary: Critical essays, 1987–1997*. Cambridge: The MIT Press.

Churchland, P. S. (1994). Can neurobiology teach us anything about consciousness? *Proceedings and Addresses of the American Philosophical Association*, *67*(4), 23–40. https://doi.org/10.2307/3130741

Clark, A. (1997). The dynamical challenge. *Cognitive Science*, *21*(4), 461–481. https://doi.org/10.1016/S0364-0213(99)80030-5

Coelllo, Y., & Fischer, M. H. (Eds.). (2016). *Foundations of embodied cognition: Volume 1: Perceptual and emotional embodiment*. New York: Routledge.

Creath, R. (2023). Logical empiricism. In E. N. Zalta, & U. Nodelman (Eds.). *The Stanford encyclopedia of philosophy* (winter 2023 ed.). Stanford, CA: Stanford University. Retrieved from https://plato.stanford.edu/archives/win2023/entries/logical-empiricism/

Dennett, D. C. (1993). Quining qualia. In A. I. Goldman (Ed.). *Readings in philosophy and cognitive science* (pp. 381–414). Cambridge: The MIT Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge: The MIT Press.

Favela, L. H. (2020). Dynamical systems theory in cognitive science and neuroscience. *Philosophy Compass*, *15*(8), e12695, 1–16. https://doi.org/10.1111/phc3.12695

Favela, L. H. (2024). *The ecological brain: Unifying the sciences of brain, body, and environment*. New York: Routledge. https://doi.org/10.4324/9781003009955

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11–12), 11–39.

Gibson, J. J. (1966). *The senses considered as perceptual system*s. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1986/2015). *The ecological approach to visual perception* (classic ed.). New York: Psychology Press.

Glock, H.-J. (2008). *What is analytic philosophy?* Cambridge: Cambridge University Press.

Hardcastle, V. G. (1999). *The myth of pain*. Cambridge: The MIT Press.

Irvine, E., & Sprevak, M. (2020). Eliminativism about consciousness. In U. Kriegel (Ed.). *The Oxford handbook of the philosophy of consciousness* (pp. 348–370). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198749677.013.16

Jovanov, K., Clifton, P., Mazalek, A., Nitsche, M., & Welsh, T. N. (2015). The limb-specific embodiment of a tool following experience. *Experimental Brain Research*, *233*, 2685–2694. https://doi.org/10.1093/oxfordhb/9780198749677.013.16

Kim, J. (1993). *Supervenience and mind: Selected philosophical essays*. New York: Cambridge University Press.

Koch, C. (1999). *Biophysics of computation: Information processing in single neurons*. New York: Oxford University Press.

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*(4), 13–26. https://doi.org/10.1609/aimag.v38i4.2744

Lobo, L., Heras-Escribano, M., & Travieso, D. (2018). The history and philosophy of ecological psychology. *Frontiers in Psychology: Cognitive Science*, *9*(2228), 1–15. https://doi.org/10.3389/fpsyg.2018.02228

Ludwig, D. (2014). Hysteria, race, and phlogiston. A model of ontological elimination in the human sciences. *Studies in History and Philosophy of Science Part*

C: Studies in History and Philosophy of Biological and Biomedical Sciences, *45*, 68–77. https://doi.org/10.1016/j.shpsc.2013.10.007

Marcus, G. (2015). Face it, your brain is a computer. *The New York Times*, June 28, 2015, p. SR12. Retrieved from https://www.nytimes.com/2015/06/28/opinion/sunday/face-it-your-brain-is-a-computer.html

Menary, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, *9*, 459–463. https://doi.org/10.1007/s11097-010-9187-6

Neisser, U. (1967/2014). *Cognitive psychology* (classic ed.). New York: Psychology Press.

Neurath, O. (1973). *Empiricism and sociology* (M. Neurath, & R. S. Cohen, eds.). Boston, MA: D. Reidel Publishing Company.

Newen, A., de Bruin, L., & Gallagher, S. (Eds.). (2018). *The Oxford handbook of 4E cognition*. Oxford: Oxford University Press.

O'Brien, M. (2023). Chatbots sometimes make things up. Is AI's hallucination problem fixable? *AP News*. Retrieved from https://apnews.com/article/artificial-intelligence-hallucination-chatbots-chatgpt-falsehoods-ac4672c5b06e6f91050aa46ee731bcf4

Preston, A. (2007). *Analytic philosophy: The history of an illusion*. New York: Continuum.

Putnam, H. (1975). The nature of mental states. In *Mind, language, and reality: Philosophical papers* (vol. 2, pp. 429–440). Cambridge: Cambridge University Press.

Quine, W. V. O. (1960/2013). *Word and object* (new ed.). Cambridge: The MIT Press.

Ramsey, W. (2024). Eliminative materialism. In E. N. Zalta, & U. Nodelman (Eds.). *The Stanford encyclopedia of philosophy* (winter 2024 ed.). Stanford, CA: Stanford University. Retrieved from https://plato.stanford.edu/archives/win2024/entries/materialism-eliminative/

Rorty, R. (1965). Mind-body identity, privacy, and categories. *The Review of Metaphysics*, *19*(1), 24–54. https://www.jstor.org/stable/20124096

Rowlands, M. (2010). *The new science of the mind: From extended mind to embodied phenomenology*. Cambridge: The MIT Press.

Segundo-Ortin, M., & Raja, V. (2024). *Ecological psychology*. New York: Cambridge University Press. https://doi.org/10.1017/9781009451413

Sprevak, M., & Colombo, M. (Eds.). (2019). *The Routledge handbook of the computational mind*. New York: Routledge.

Stadler, F. (2008). The Vienna Circle: Context, profile, and development. In A. Richardson, & T. Uebel (Eds.). *The Cambridge companion to logical empiricism* (pp. 13–40). New York: Cambridge University Press.

Stich, S. (1983). *From folk psychology to cognitive science: The case against belief*. Cambridge: The MIT Press.

Sussillo, D. (2014). Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology*, *25*, 156–163. https://doi.org/10.1016/j.conb.2014.01.008

Thagard, P. (2023). Cognitive science. In E. N. Zalta, & U. Nodelman (Eds.). *The Stanford encyclopedia of philosophy* (winter 2023 ed.). Stanford, CA: Stanford University. Retrieved from https://plato.stanford.edu/archives/win2023/entries/cognitive-science/

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of the mind*. Cambridge, MA: Belknap Press.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition*, *9*, 237–304. https://doi.org/10.1016/0010-0277(81)90002-0

van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, *92*(7), 345–381. https://doi.org/10.2307/2941061

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge: The MIT Press.

Von Eckardt, B. (1995). *What is cognitive science?* Cambridge: The MIT Press.

Wittgenstein, L. (1921/2001). *Tractatus logico-philosophicus* (D. F. Pears & B. F. McGuinness trans.). New York: Routledge.

Wittgenstein, L. (1958). *Philosophical investigations* (G. E. M. Anscombe trans.). Oxford: Basil Blackwell.

Zador, A., Escola, S., Richards, B., Ölveczky, B., Bengio, Y., Boahen, K., … & Tsao, D. (2023). Catalyzing next-generation artificial intelligence through NeuroAI. *Nature Communications*, *14*(1597), 1–7. https://doi.org/10.1038/s41467-023-37180-x

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Part 2

From skills and habits to language and communication

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

9 Skillful coping in the metaverse

On the challenges of immersion

Marta Pérez-Verdugo

9.1 Introduction

This chapter revisits the well-known debate between Hubert Dreyfus and John McDowell, which constituted a fruitful point of contact and blurring between post-analyitic and post-phenomenological[1] traditions in philosophy and centers on whether all human action is –or can be- conceptually articulated, or whether some forms of experience, such as expert performance, are fundamentally non-conceptual. For Dreyfus, the mindless flow of absorbed coping exemplifies skillful action. McDowell, however, warns that such a view risk collapsing agency into automaticity, threatening our ability to distinguish autonomous responsiveness from mere causally determined behavior. Within e-cognition, this tension finds echoes and resolutions in the braiding of analytic and phenomenological inspirations within its different families of theorizing, many of which adopt Dreyfus' anti-representational commitments.

I will begin in Section 9.1 by reviewing the debate in the light of the agential worry around Dreyfus' account of absorbed coping. In Section 9.2, I show how, while sharing Deyfus' aims, some positions within e-cognition—notably ecological psychology and enactivism—offer more nuanced accounts of skill and agency that resist both intellectualism and automatism. In Section 9.3, I turn to the concept of immersion in digital technologies to explore a contemporary example of "mindless" behavior that sparks moral worry. We will look at the similarities between these kinds of behavior and Dreyfusian desrciptions of expertise, thus arguing that –as McDowell and other critics feared- an over-reliance on a principle of automaticity or flow cannot, on its own, distinguish between skillful absorption and mindless entrainment. Finally, in Section 9.5, we then turn to how one specific approach within e-cognition (enactivism) can offer a

1 We use the lable "post-phenomenological" to underline the phenomenological foundations of Dreyfus' philosophy, mainly through his drawing from Heidegger, that are however mixed with other styles of philosophy in his work. This use of the term should not be mistaken with the specific school of "postphenomenology" within philosophy of technology as developed by Don Ihde and others.

DOI: 10.4324/9781003649359-12
This chapter has been made available under a CC BY 4.0 license.

more fruitful account of skillful coping and asymmetrical immersion, the latter referring to cases where, contrary to what happens in true expertise, environmental design plays a dominant role in stabilizing behavior, bypassing the agent's regulatory capacities.

9.2 Absorbed coping in the Dreyfus-McDowell debate and the agential problem

The Dreyfus-McDowell debate (see Schear, 2013) has been a rich soil for philosophy of mind, and a particularly interesting point of contact between analytic philosophy and the phenomenological orientation embraced by many e-cognition theorists. In broad terms, the debate concerns the role of conceptuality in everyday action and experience. Can we make sense of a form of engagement with the world that is not conceptually articulated, as Dreyfus claims, or must we, like McDowell, insist that all intentional action occurs within a potentially conceptual space?

The debate roughly starts as follows: Dreyfus (2005) argues that McDowell is falling prey to a common myth of Kantian inspiration in analytic philosophy (and classic AI research); the Myth of the Mental. Dreyfus charges these philosophers with mistakenly believing that since much of our experiences are characteristically mediated by conceptual articulation, *all of them* are. And McDowell in particular, even if he claims to want to find a rather situated view of mind and action, is championing this idea through his insistence (laid out in his 1996 book, Mind and World) on placing human behaviour and experience in the logical space of reasons. Dreyfus' strategy to unmask this myth is to find at least some form of human experience that is *incompatible* with rational or conceptual articulation. And he finds his example in the skillful, absorbed coping of experts and the phenomenological characterization of flow (a term later developed by Csikszentmihalyi, 2013):

> in total absorption, sometimes called flow, one is so fully absorbed in one's activity that one is not even marginally thinking about what one is doing.
>
> (Dreyfus, 2013, p. 28)

His claim, to be clear, and the force of his argument, is not that these kinds of experiences just happen to occur without conceptual articulation, but that they *cannot* co-occur with it. To argue for this, he draws on the Heideggerian intuition of *breakdowns* and their disruptive character: when an expert's action is going smoothly, fluently, it is characterized by this experience of flow, and it is only when something disrupts the action, when there is a breakdown, that she stops and thinks. But this breakdown is a disruption of the experience. If a sports player starts to think about her movements, if she conceptually articulates the position of her arms when throwing or her legs when running, or even just what exactly she is doing or why, her action

will be disturbed and she will no longer be in the state of flow that characterized her skillful coping. In his own words, "the perceptual/social field is in principle unthinkable. In so far as I'm absorbed in the field of forces I can't think them, and in so far as I distance myself in order to think them they vanish" (Dreyfus, 2013, p. 27). It is not that we just generally do not rationally articulate our skillful actions, that we generally do not think about what we are doing, it is rather that the kind of experience that absorbed coping is, is *non-articulable*. The moment we rationally think about it, it turns into a different experience.

Here lies the bedrock of Dreyfus' position in the debate. If he manages to convince us that absorbed coping is a relevant, intentional, human experience (and not just any experience, but the experience that characterizes expertise!), and that it is incompatible with articulation, his point will be made that the pervasiveness of the mental (understood as a conceptual articulation of experience) is a myth. Conceptual articulation will have been proven to not pervade all aspects of experience. Facing this, however, McDowell's replies still try to find a way to retain a *potential* conceptual articulation for all intentional human experience. He argues that Dreyfus is misconstrues his point as if he were claiming that all experience is always conceptually articulated, when he is simply claiming that it can *potentially* be. He grants that Dreyfus is right in his description of the phenomenology of absorbed coping, and in the disrupting character of someone stopping to articulate their action. But he claims that this does not mean that said experience is non-articulable, just that it is not articulated during flow. For McDowell, what disrupts the flow when we articulate what we are doing is the fact that we now have to perform a different action simultaneously; that, for instance, of answering a question about the reasons for acting, reasons which we nevertheless already had when in flow.

> I have granted that if [the expert chess player] *does* say such things [what he is doing], he is no longer acting in flow. But if he says such things, he gives expression to knowledge he already had when he was acting in flow.
>
> (McDowell, 2013, p. 46)

To actually weigh the balance in favor of one or another on this debate, we would need to get deeper into the many nuances and contributions in each position (for instance, Dreyfus' reliance on fields of forces to strengthen the non-articulability of skilled action, or McDowell's identification of the Myth of the Mind as Detached in Dreyfus' account). However, what I am interested in here is exploring one of the reasons why McDowell is so keen on defending his position. One of the main motivations for him is *moral*, and inherited from his analytic tradition: he wants to be able to endow human behavior with freedom. Within McDowell's Sellersian backdrop, to withdraw reasons and conceptual articulation from any kind of human behavior would amount

to saying that that behavior lacks real agency; that it is determined by the logical space of causes. As Rietveld (2010) states,

> he seems to fear that without a role for conceptuality we would not be able to place unreflective action in the category called space of reasons, being rather forced to see it as determined by disenchanted causal interaction. This would block an account of its freedom, rationality and normativity.
>
> (p.185)

McDowell wants to secure a place for freedom, an autonomy from determined causality, in skillful coping (and in the rest of our mental experiences). He is in agreement with Dreyfus in a description of the phenomenology of flow, but he is concerned that granting Dreyfus the kind of mindlessness that he seems to be arguing for would devoid human skillful action of some of its most important qualities.

And this worry is not coming out of nowhere. In the description that Dreyfus provides of absorbed coping, he does place special importance on the idea that movements seem to be "drawn out of agents" by the environment. For Dreyfus,

> unlike deliberate action, skillful coping turns out to have a world-to-mind direction of causation. We do not experience our intentions as causing our bodily movements; rather, in skillful coping we experience the situation as drawing the movements out of us.
>
> (Dreyfus, 2002, p. 380)

This description does make it seem like skillful coping is some sort of automatism; the environment or the situation causally evoking specific behaviors. It is important to remember here that Dreyfus' description is not of some minor aspects of our experience, but precisely of *expert* behavior, which is found in our most meaningful everyday coping (including ethical behavior, see his discussions around Aristotelian *phronesis*, a key theme of the debate, in e.g. Dreyfus, 2005). There is then a real worry that he is describing such behaviors as somewhat lacking freedom, as being overdetermined by the environment. Dreyfus seems to be aware of this possibility, and he aims to reinstate agency in his account when he is cautious in saying that "I am in control of my movements in the sense that I can stop doing what I'm doing if I will to do so" (Dreyfus, 2002, p.380). It is unclear, however, that being able to stop what one is doing is enough to grant it the relevant moral freedom, and one could hope for a notion of agential control to require more.

This worry that lies at the back of McDowell's stakes in the debate is not only brought up by him, but has been the object of much conversation. A particularly clear and strong challenge to the Dreyfusian view is raised

by Barbara Montero in her 2013 essay *A dancer reflects*. Here, the author identifies in Dreyfus' account (and others') another Myth, although in this case she terms it a principle; *the principle of automaticity*. In her words, this would be the commonly accepted idea that "when all is going well, expert performance significantly involves neither self-reflective thinking, nor planning, nor predicting, nor deliberation, nor mental effort" (Montero, 2013, p. 304). Drawing from her own experience as a dancer, she argues that following this principle in defining expert action leads us to ignore crucial aspects of what makes it expert, such as its creativity and flexibility, as well as the effort that goes into it. Although she focuses specifically on the idea of expertise, her remarks echo the aforementioned worry about these phenomenological descriptions of absorbed coping not leaving sufficient room for agential control in general. Sutton et al. (2011) also offered an analysis of certain views of expert experience, that they identify with Dreyfus, as a refinement of descriptions of reflex-like responses that cannot cover the richness of expert performance. They focus on how Dreyfus is keen to reject any sort of *mentality*[2] in skillful coping and criticize this anti-cognitivism as not being able to make sense of the extraordinary adaptability of expert behavior. Shaun Gallagher and Somogy Varga (2020) similarly (although in a slightly less cognitivist vein) accuse Dreyfus of ending up depicting experts as zero-intelligent agents in his account of skillful coping (p. 3). They borrow this term from economics to refer to "an agent who, to perform a task, acts in a purely automatic way and whose performance would involve no cognitive contribution" (p. 3). Within these critiques, we can see laid out, in a different manner, the core of McDowell's concern; that we risk turning agents into automata. Or, in other words, that under these descriptions, our behavior can be seen as fully determined by the environment.

9.3 E-cognition on the face of the debate

How has e-cognition positioned itself in relation to this debate? E-cognition as a family of traditions (embodied cognition, ecological psychology, extended and embedded cognition, enactivism) generally shares the philosophical commitments to embodiment and anti-representationalism that motivate Dreyfus in his account. Many of them also share his inspiration from phenomenological traditions, particularly Merleau-Ponty's embodied phenomenology. As

2 As we have seen, Dreyfus does explicitly talk about absorbed coping as mindless, and as mind being disruptive for flow, but it is important to understand that Dreyfus is using the term "mindless" in opposition to McDowell's specific picture of mind as conceptual. This is, nevertheless, surprising and potentially problematic, as he seems to be too quick to leave the notion of mind to be defined by the other side of the debate. This theme is also picked up by Segundo-Ortín and Heras-Escribano (2021), who we discuss later. Similar comments around this were brought up by María José Frapolli when presenting a previous version of this work.

such, in some cases, e-cognition clearly defends that the most basic mode of relating to the world is similar to Dreyfus'description of absorbed coping in that it is non-conceptual. The influence is not one-sided, and Dreyfus himself draws from J.J. Gibson and ecological psychology's theory of affordances to articulate his view. The focus on *direct, unmediated* perception within ecological psychology is certainly in alignment with Dreyfus' defense of skillful coping as escaping conceptual articulation.

But does e-cognition share the problematic suppositions that McDowell and others accuse Dreyfusian views of? That will depend on what E of the e-cognition family we are talking about. One that does share the principle of automaticity in its descriptions is *extended cognition* (developed after Clark & Chalmers, 1998).[3] Extended cognition, concerned with exploring whether and how we can consider elements of the environment as realizing cognitive processes (together with the brain and body), draws from similar Heideggerian phenomenological insights as Dreyfus in respect to tool use. To identify cases of successful integration of an artifact in a system of extended cognition, one of the key dimensions to take into account is that of transparency. The more phenomenologically transparent (seamless, requiring no reflective awareness) our use of an artifact is, the more integrated it is in a system of extended cognition (Heersmink, 2015). I have elsewhere termed this idea of transparency within extended cognition as transparency-as-automaticity (Pérez-Verdugo, 2022) precisely to underline its connections with the principle of automaticity as derived from a Dreyfus-like view of experience. The assumption here is that in non-extended cognition—i.e. the kind of cognition that we would do with just our brain and bodies—this phenomenological transparency or automaticity is also at play (at least with regards to our own bodies), and thus our extended use of artifacts should mimic it (these assumptions are debated by Andrada, 2020). But this notion of transparency has been shown to lead precisely to ethical problems related to agency (Clowes, 2020; see also Pérez-Verdugo, 2022; Pérez-Verdugo & Barandiaran, 2023), particularly in relation to digital technologies. With this sort of transparency, the possibilities of these artifacts manipulating our behavior in ways that annul our agency seem particularly threatening. The agential worry thus prevails in the transparency-as-automaticity focus of extended cognition, and materializes itself in our relationship with digital technologies. The move to include the environment into the mind, at least in the functionalist, analytic-inspired version of the extended mind, does not preclude that the environment can determine our behaviors.

3 Which is somewhat ironic since it is, arguably, the branch of e-cognition that is less concerned with its anti-representationalist commitments, given its original functionalist focus on the possibly multiple realizers of cognition.

But other developments within other Es of the e-cognition family, that retain a strongly situated character, have explicitly questioned some of these problematic underlying assumptions of the Dreyfusian picture of skillful coping. Within ecological psychology, for instance, Segundo-Ortín and Heras-Escribano (2021) aim to resituate intelligence and the mind in skillful action without having to abandon the principles of radically embodied cognitive science. That is, without needing to resort to representationalist cognitive processes, but rather relying on the theoretical framework of ecological psychology and on Dewey's theory of habits. They focus, as Dreyfus also does (although him more implicitly), on Gibson's idea of the education of attention; that is, of learning through practice to better perceive and discriminate the affordances that are more specific and relevant to one's goals. The difference is that Dreyfus left the mind out of this engagement with affordances:

> there must be some detectable invariant features in what J.J Gibson calls the ambient optic array and that human beings and animals can learn to respond to them. *These features, although available to the perceptual system, needn't be available to the mind.*
>
> (Dreyfus, 2005, p. 54, italics from the original)

Reading this, one is left wondering where the mind is if not in dealing with affordances in increasingly efficient ways (see footnote 2). Segundo-Ortin and Heras-Escribano stress that these processes are rightfully minded and, crucially, not passive; the agent is able to control this learning to increase the effectiveness of their performance by seeking out more relevant affordances. They defend that "it is because the agent is not acting on autopilot but constantly searching for new information for affordances that she can control her behavior, adapting previously learned habits to the new, particular circumstances" (Segundo-Ortin & Heras-Escribano, 2021, p. 10128).

Another position within e-cognition that challenges Dreyfus can be found within enactivism. Here, it has been Alva Noë (2012, 2023) in particular who challenges part of the Heideggerian assumptions behind Dreyfus' story through his analysis of the fragility of presence and perception. For Noë, the mere establishing of perceptual relationships with our environment is a deployment of skilled understanding, even if that kind of understanding is not necessarily conceptual—but rather a sensorimotor kind of understanding. But the crucial thing is that these relationships are fragile, they need to constantly be accomplished, they do not come for free (Noë, 2025). The constant achieving of presence, of maintaining a perceptual relation with the environment, contrasts with the principle of automaticity that claims that when everything is going well, no effort is required from the skilled coper. In that view, it is only in the specific moments of breakdown that we need to work on our agentive understanding of the situation. Alva Noë explicitly counters this idea, and as such grants the perceiver a more constant display

of agency and control, as she is constantly making the skilled effort to sustain her relationship with her environment. The environment, in this view, can be present in different ways depending on our mode of skillful engagement with it, and the kind of withdrawal of the environment that is experienced in absorbed coping does not mean that it is completely absent (Noë, 2012, p. 9). The environment, even if a strong structuring force, is not in skillful coping merely a trigger of automated responses from the agent; the skillful coper is instead always renegotiating its relationship to it.

We can see then that not all of e-cognition share the problematic presuppositions of Dreyfusian views of absorbed coping, despite their shared motivations in opposing classical cognitivist and representationalist views that are many times derived from analytic philosophy and its comitment to a specific view of human rationality. The kind of concerns that motivated McDowell in the debate can be accounted for by other developments within e-cognition, such as enactivism. To better show this, we will now turn to analyze a particular example where the notion of absorbed coping becomes central. We will explore the ethical shortcomings of adopting a Dreyfusian view of skillful by showing its inability to differentiate these cases of automatic, not fully agential coping from "true" expert coping..

9.4 Immersion in digital technologies

The concept of immersion plays a central role in the design and study of interactive digital media, especially in videogames, digital environments, and extended reality technologies. It is widely considered a key marker of user engagement—an indicator that the experience has successfully drawn the user into a world of perceptual, cognitive, or affective involvement (Salen & Zimmerman, 2003). The idea of immersion in digital environments was first articulated within the context of robotics and remote-control technologies. Marvin Minsky's (1980) notion of *telepresence* was meant to capture the phenomenological effect of being "present" in a different location through technological mediation. Though initially concerned with haptic control and visual feedback in teleoperated systems, the notion of telepresence quickly migrated into the field of virtual reality, where it was reformulated as the sense of "being there" in a digitally constructed space (Lombard & Ditton, 1997). This idea is a crucial design and research concern in videogame design and Virtual Reality (VR), particularly focused on achieving this immersion in specific fictional or constructed worlds. Interestingly enough, Dreyfus was an early critic of the possibility of achieving true presence in virtual, "disembodied" environments (Dreyfus, 2001).[4]

4 I believe that here as well, a much more fruitful and nuanced account of the specific kinds of presence that can be experienced within digital environments can be attained if we follow the previously discussed view of Alva Noë (2012). The specific modes of skilled access that

But this sense of immersion as transportation to another place—as Calleja (2011) notes—is not the only kind of immersion relevant in digital media. There is a second, and arguably more pervasive, sense of immersion in digital technologies: immersion as absorption. Unlike transportation, which is often tied to narrative or representational realism, absorption is tied to a feeling of deep involvement characterized by a state of flow. Think of the experience of immersion in playing Tetris (see Calleja, 2011, pp. 26–27 for a discussion); it is the kind of immersion one feels not because they are imaginatively elsewhere, but because they are fully engaged in the activity at hand—losing track of time, tuning out distractions, and becoming absorbed in the unfolding task. This form of immersion is phenomenologically marked by transparency: the interface recedes, the player's actions feel smooth and unmediated, and awareness of one's physical surroundings often diminishes; users "lose themselves in the game" (Jennett et al., 2008). The phenomenological description of this kind of immersion is almost identical to the description of absorbed coping by Dreyfus. Dreyfus would then have to grant, as per his characterization, that this sense of immersion in virtual environments is a case of absorbed coping.

But this phenomenology of absorption within digital environments is not exclusive to gaming. It also appears critically in the increasingly common experiences of streamlined, continuous engagement on algorithmically driven platforms. Digital platforms, such as social media or short video applications, where interaction is made to be fluid and undisputed, evoke these kinds of immersion. One particular instance of this interaction is in fact commonly referred to as "mindless" scrolling, a term that carries both experiential and normative weight. From a phenomenological standpoint, the term "mindless" signals the diminished role of reflective awareness in the interaction. One is not aware, for instance, of the specific movements one's finger is making (or how often it is making them), nor, characteristically, of the time spent scrolling. The user is immersed not in a representational world but in a behavioral loop that proceeds with minimal interruption or conscious guidance. Many of our digital technologies are designed precisely to have this effect. As Cox et al. (2016) report, "points of difficulty encountered during user's interaction with technology [are] removed from technology in order to reduce the risk of user disengagement. (…) [resulting] in mindless forms of interaction that can have negative consequences". (p.1391). The phenomenology of these "mindless forms of interaction" is amenable to the kind of absorbed coping described by Dreyfus: uninterrupted flow, the withdrawal of reflective articulation, a feeling of being drawn into the situation by the environment (Dreyfus, 2002, 2013). But there is an important difference to

digital environments allow create new forms of presence, with their own qualities. The specific character of these virtual realities and how they are experienced, either as real or as imaginary environments, has already been discussed within e-cognition (Baggs et al., 2024; Rolla et al., 2022), particularly after Chalmers' (2022) realist account of virtual environments.

be found in the negative connotations that the term also carries. Unlike in the virtuosity examples used by Dreyfus of experts displaying their skillful coping, the behavior involved in mindlessly scrolling is often compulsive rather than expressive, entrained rather than skillful. After a session of mindless scroll, one is in many cases left with the feeling that it was not an action they wanted to be engaged in at all, or at least not for that long.

Within these hyperdesigned[5] digital technologies, we seem to actually be facing the problematic situation that was foreshadowed in McDowell's and others' worries around agency in absorbed coping. Our behavior is here, in a sense, overdetermined by a designed environment. We do behave like zero-intelligence agents. It would seem weird to say that we are behaving like experts when we get caught up in mindless scrolling and we would prefer to be doing something else. Nevertheless, as per Dreyfus' phenomenological description of skillful coping, or other views that would follow the principle of automaticity, we would be having the same kind of experience as experts. The case of mindless scroll, then, embodies the worries around how much room these accounts leave to agency (and freedom or autonomy). A successful description of skillful coping should be able to differentiate when we are absorbed in the flow as a result of skillful coping, and when our actions are fluid because of an overdetermination of the environment. In the next section, I will return to e-cognition to show the theoretical tools that enactivism provides to face these agential worries.

9.5 The enactive picture of (asymmetrical) immersion

As explored in Pérez-Verdugo and Barandiaran (2023), a nuanced framework for understanding the difference between cases of truly skillful coping and cases of environment-induced absorption can be found in the enactive approach to cognition, and in particular in the operational model of sensorimotor agency developed by Di Paolo et al. (2017). A central concept in this tradition is the notion (adapted from Piaget) of sensorimotor schemes: coordinated, plastic patterns of perception and action that agents enact across time. These schemes or habits are not static, but precarious, adaptive structures that require continuous regulation in order to be maintained and deployed effectively (Di Paolo et al., 2017). As such, this proposal shares Segundo-Ortin and Heras-Escribano's (2021) focus on the usefulness of the notion of habit, even if in each case it is articulated differently.

5 The notion of hyperdesign, introduced in Pérez-Verdugo and Barandiaran (2023), aims to highlight the fact that the scale and detail of design in digital environments are unprecedented, as well as the dynamical possibilities that these environments provide to continuously modify their design. As such, digital environments are a different kind of environment than encountered before, and they are being effectively designed to induce states of absorption in a manner that no other previous technology could achieve.

Sensorimotor schemes are constituted by the dynamic coupling between agent and environment. This includes not only the agent's neural and bodily structures but also environmental structures that support the patterns of sensorimotor coordination. For example, walking through a familiar city may involve a dense network of schemes stabilized by the layout of the streets, the affordances of sidewalks and habitual routes, apart from the many musculoskeletal and neural structures in the brain and body. These schemes are fluid, yet stable; plastic, yet robust. And their enactment is both structured by the environment and regulated by the agent according to the norms of her sensorimotor identity. The possibility to adjust behavior through the agentive, asymmetrical equilibration of such sensorimotor schemes is precisely what Di Paolo et al. (2017) consider characteristic of the kind of intentional everyday action, explicitly placing their account beyond Dreyfus and McDowell:

> (…) a description of skillful intentional action that is characterized neither by the total absence of mindedness suggested by Dreyfus, nor by the always rational mindfulness proposed by McDowell. (…) On this account, we can act intentionally (…) because we are non-conceptually involved in the process of shaping the dynamics that lead to the engagement and control of particular sensorimotor schemes.
>
> (pp. 182–183)

Here again, taking the expression by Segundo-Ortin and Heras-Escribano (2021), habits are neither mindful nor mindless, but minded.

In this context, immersion arises as the phenomenological feel of "coherent, long-range relations between integrated sensorimotor schemes" (Di Paolo et al., 2017, p. 156). When a network of sensorimotor schemes becomes sufficiently stabilized and coherent, such that the interaction unfolds with minimal disruption and maximum fluency, we experience a feeling of flow. But—and this is the crucial point—the enactive framework makes it clear that there are two sides to the stabilization of these networks, reflecting the two different kinds of support structures that constitute it; agent and environment. In genuine expertise, as the cases described by Dreyfus, immersion is the result of agentive equilibration—the plastic, context-sensitive modulation of sensorimotor schemes by the agent. But in asymmetrical immersion (Pérez-Verdugo & Barandiaran, 2023), by contrast, the environment plays a disproportionate role in driving the stabilization process. In cases like mindless scrolling, then, immersion emerges not from the agent's skillful regulation, but from environmental structures engineered to equilibrate specific sensorimotor schemes. It becomes possible to analyze, then, the practical difference between both cases by exploring what support structures carry the weight of the stabilization of schemes, as well as by making a more temporally extended analysis of the extent to which the agent is able to further regulate this stabilization. In both cases, we do have a phenomenology of immersion as seen in a robust and coherent network of schemes. But true

experts exhibit a regulative flexibility upon this network that has led them to be able to retain control of its stabilization. The sensorimotor schemes here are stable but remain open to plastic changes driven by the agent. In mindless scrolling, by contrast, the network is composed of rigidly stabilized schemes that are shaped by the platform's design. The user's sense of fluency is not accompanied by adaptability; the phenomenological transparency of the interface conceals a form of interaction that is driven from the outside.

9.6 Conclusion

As we have seen, the debate between Dreyfus and McDowell on absorbed coping reveals a tension, later taken up in e-cognition, between preserving the embodied phenomenology of skillful action and securing a robust notion of agency. While Dreyfus' view provides a compelling description of unreflective, embodied fluency, it risks, as critics have pointed out, collapsing expert performance into a form of automatism. This highlights a fundamental shortcoming in his model: it lacks the tools to distinguish between genuinely skillful, autonomous coping, and environmentally orchestrated behavioral entrainment. This can lead to obscuring specific problematic forms of relating with the environment that do actually pose a threat to our agency and autonomy, as is the case with mindless scrolling. E-cognition approaches, as surveyed here, show that a meaningful notion of agency can be preserved without having to accept a pervasiveness of conceptual articulation, by shifting the focus from reasons to regulation. The difference between mindless and mindful immersion lies not in the presence or absence of conceptuality, but in the distribution of control over the dynamics of stabilization. In sum, the enactive approach as part of the e-cognition family gives us the vocabulary and the conceptual tools to disentangle absorption from automatism as it was problematized in the Dreyfus-McDowell debate.

References

Andrada, G. (2020). Transparency and the phenomenology of extended cognition. *LÍMITE Revista Interdisciplinaria de Filosofía y Psicología*, *15*(20), 1–17. Retrieved from: https://revistalimite.uta.cl/index.php/limite/article/view/232/209

Baggs, E., Grabarczyk, P., & Rucińska, Z. (2024). The visual information available in virtual reality. *Ecological Psychology*, *36*(1), 24–38. https://doi.org/10.1080/10407413.2024.2322992

Calleja, G. (2011). In-Game from Immersion to Incoporation. The MIT Press.

Chalmers, D. J. (2022). *Reality+: Virtual Worlds and the Problems of Philosophy*. W. W. Norton & Company.

Clark, A., & Chalmers, D. (1998). La mente extendida (traducción). The extended mind. *Analysis*, *58*(1), 7.

Clowes, R. (2020). The internet extended person: Exoself or doppelganger? *LÍMITE Revista Interdisciplinaria de Filosofía y Psicología*, *15*(22), 1–23. Retrieved from: https://revistalimite.uta.cl/index.php/limite/article/view/234/211

Cox, A. L., Gould, S. J. J., Cecchinato, M. E., Iacovides, I., & Renfree, I. (2016). Design frictions for mindful interactions: The case for microboundaries. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1389–1397. https://doi.org/10.1145/2851581.2892410

Csikszentmihalyi, M. (2013). *Flow: The Psychology of Happiness*. Random House.

Di Paolo, E. A., Buhrmann, T., & Barandiaran, X. (2017). *Sensorimotor Life: An Enactive Proposal*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198786849.001.0001

Dreyfus, H. L. (2001). *On the Internet*. Routledge.

Dreyfus, H. L. (2002). Intelligence without representation – Merleau-Ponty's critique of mental representation. *Phenomenology and the Cognitive Sciences*, *1*, 367–383.

Dreyfus, H. L. (2005). Overcoming the myth of the mental: How Philosophers can Profit from the Phenomenology of Everyday Expertise. *Proceedings and Addresses of the American Philosophical Association, 79*(2), 47–65.

Dreyfus, H. L. (2013). The myth of the pervasiveness of the mental. In J. K. Schear, *Mind, Reason and Being-In-The-World: The McDowell-Dreyfus Debate* (pp.15–40). Routledge.

Gallagher, S., & Varga, S. (2020). Meshed architecture of performance as a model of situated cognition. *Frontiers in Psychology*, *11*, 2140. https://doi.org/10.3389/fpsyg.2020.02140

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences*, *14*(3), 577–598. https://doi.org/10.1007/s11097-014-9355-1

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence | | Oxford Academic. *Journal of Computer-Mediated Communication*, *3*(1). Retrieved from: https://academic.oup.com/jcmc/article/3/2/JCMC321/4080403?login=false

McDowell, J. (1996). *Mind and World*. Harvard University Press.

McDowell, J. (2013). The myth of the mind as detached. In J. K. Schear, *Mind, Reason and Being-In-The-World: The McDowell-Dreyfus Debate* (pp. 41–58). Routledge.

Minsky, M. (1980). Telepresence. *OMNI Magazine*.

Montero, B. G. (2013). A dancer reflects. In J. K. Schear, *Mind, Reason, and Being-in-the-World* (pp. 303-319). Routledge.

Noë, A. (2012). *Varieties of Presence*. Harvard University Press.

Noë, A. (2023). *The Entanglement: How Art and Philosophy Make Us What We Are*. Princeton University Press.

Noë, A. (2025). Perception as a relationship. In O. Beck & F. Masrour, *The Relational View of Perception: New Philosophical Essays* (pp. 114–129). Routledge.

Pérez-Verdugo, M. (2022). Situating transparency: An extended cognition approach. *Teorema: Revista Internacional de Filosofía*, *41*(3), 7–24.

Pérez-Verdugo, M., & Barandiaran, X. (2023). Personal autonomy and (digital) technology: An enactive sensorimotor framework. *Philosophy & Technology*, *36*(4), 1–28. https://doi.org/10.1007/s13347-023-00683-y

Rietveld, E. (2010). McDowell and Dreyfus on unreflective action. *Inquiry*, *53*(2), 183–207. https://doi.org/10.1080/00201741003612203

Rolla, G., Vasconcelos, G., & Figueiredo, N. M. (2022). Virtual reality, embodiment, and allusion: An ecological-enactive approach. *Philosophy & Technology*, *35*(4), 95. https://doi.org/10.1007/s13347-022-00589-1

Salen, K., & Zimmerman, E. (2003). *Rules of Play*. MIT Press.

Schear, J. K. (2013). *Mind, Reason, and Being-in-the-World: The McDowell-Dreyfus Debate*. Routledge.

Segundo-Ortin, M., & Heras-Escribano, M. (2021). Neither mindful nor mindless, but minded: Habits, ecological psychology, and skilled performance. *Synthese*, *199*(3), 10109–10133. https://doi.org/10.1007/s11229-021-03238-w

Sutton, J., McIlwain, D., Christensen, W., & Geeves, A. (2011). Applying intelligence to the reflexes: Embodied skills and habits between Dreyfus and descartes. *Journal of the British Society for Phenomenology*, *42*(1), 78–103. https://doi.org/10.1080/00071773.2011.11006732

10 The organismic turn. Teleosemantics after 4E

Xabier E. Barandiaran and Tiago Rama

> The third, to lead my thoughts in order, beginning by the most simple objects, and the easiest to be known; to rise by little and little, as by steps, even to the knowledge of the most mixt; and *even supposing an Order among those which naturally does not precede one the other*.
>
> (Emphasis added)
>
> Descartes

10.1 Introduction: to descend little by little

The philosophical quest to provide a naturalistic account of the mind, one that situates mental phenomena within the causal realm of a scientifically addressable world, has been a cornerstone of analytic philosophy. This pursuit gained particular momentum following the "linguistic turn" that characterized much of early to mid-20th-century philosophy (Rorty, 1967). During this period, many philosophical problems, including those concerning the mind, were often rephrased in linguistic terms. The focus was on the logical structure of language, with the hope that understanding linguistic representation would solve classical epistemological issues, particularly concerning scientific knowledge, and illuminate (or fade away) the psychological, inner, or first-person dimension of meaning (Carnap, Wittgenstein). However, by the mid-20th century, the rise of cognitive science made increasingly apparent that a purely linguistic approach was insufficient for addressing old epistemological questions and explaining the nature of meaning (Quine, 1969). This realization, advanced by the likes of Chomsky, Tolman, Turing, and Miller, coupled with the rise of new scientific disciplines like cybernetics, cognitive psychology, generative linguistics, information theory, and early artificial intelligence, paved the way for the "cognitive turn" (Gardner, 1985). This shift refocused attention directly onto mental processes as internal cognitive states, seeking to understand them as informational token processes by the brain in analogy with the way in which digital tokens are processed in a computer; with the weight of the linguistic turn still pressing the explanation of (representational) meaning (Fodor, 1980; Putnam, 1975).

DOI: 10.4324/9781003649359-13
This chapter has been made available under a CC BY 4.0 license.

Central to this renewed focus on the mind was the persistent puzzle of intentionality, the mind's capacity to be "about" or "directed at" objects, properties, and states of affairs. This feature had been famously highlighted by Franz Brentano (1995) in the 19th century as a distinctive, if not defining, mark of the mental, distinguishing it from the merely physical. Brentano argued that every mental phenomenon is characterized by "intentional inexistence" or the "reference to a content, direction toward an object". The challenge for naturalistic philosophy, then and now, has been to explain this "aboutness" in terms compatible with a scientific, non-mysterious view of reality, without recourse to irreducible mental substances or properties, and beyond the purely descriptive frame of behaviorism.

Alongside intentionality, the problem of error or misrepresentation has posed a profound difficulty: how can a purely physical system be *wrong* about the world? If mental states are just physical states, what makes one such state a correct representation and another an incorrect one? More generally, these problems are entangled with the *problem of normativity*: namely, the task of providing a normative standard by which correctness can be assessed—a standard that is not readily available if one relies solely on physical descriptions or the descriptive, non-prescriptive character of scientific explanation.

In this historical context, classical teleosemantics developed in the 1980s (Millikan, 1984; Dretske, 1988; Papineau, 1987), offering a powerful research program—for recent advances see (Neander, 2017; Shea, 2018). The central ambition of this approach was to demonstrate that intentional states—such as beliefs and desires—can be accommodated within a physicalist ontology, insofar as they can be explained by naturalistic theories of biological function. Classical teleosemantics is a form of etiological theory (Wright, 1976). Among the various etiological frameworks, classical teleosemantics relies most heavily on natural selection. Within this paradigm, the Selected-Effect theory of function (henceforth: SE) was developed (Ayala, 1970; Millikan, 1989; Neander, 1991; Ruse, 1971), according to which the proper function of a biological trait is defined by the beneficial causal role it historically played in the reproductive success of the organism. As such, the teleosemantic solution proposed that the "proper function" of a cognitive mechanism (e.g., a belief-forming system or a perceptual state) is determined by the effects for which that type of mechanism was selected during its evolutionary history. A mental state, therefore, correctly represents the world if it is fulfilling its evolutionarily endowed function and errs if it malfunctions or is triggered in ancestrally atypical conditions.

However, in recent decades, the landscape of cognitive science has been significantly reshaped by 4E Cognition (Embodied, Enactive, Extended, and Ecological approaches). These perspectives challenge the traditional cognitivist assumptions that underpin many classical teleosemantic accounts, particularly the disembodied, computational view of mind and the sharp separation between agent and environment. While 4E approaches offer compelling alternatives for understanding cognition as a dynamic, world-involving

activity, many strands within this broad movement have struggled to articulate a robust, intrinsic source of *normativity*. Some embodied or dynamical systems approaches, by focusing purely on descriptive dynamics, even appear to sideline or deny the need for normative concepts, thereby precluding a direct path to naturalizing meaning or solving the problem of error within an embodied agency framework (Hutto & Myin, 2012; Villalobos & Palacios, 2021). Others directly reject the possibility of naturalizing norms below the social domain (Heras-Escribano et al., 2014; for a recent critical assessment see Prokop & Barandiaran, under review).

Various, deeply interconnected trends have attempted a systematic naturalizing path to normativity within 4E approaches, with a strong influence of autopoietic theory, complex systems theory, and theoretical biology. The organizational approach to biology as an autonomous grounding of cognitive capacities (Barandiaran, 2002; Bickhard, 2000; Christensen & Hooker, 2000; Moreno & Mossio, 2015) and *enactivism*, as a specific and increasingly prominent branch of 4E cognition (Di Paolo et al., 2017; Thompson, 2010; Varela et al., 1991),[1] stands out by explicitly addressing the issue of normativity (Barrett, 2017). All seek to ground norms not in evolutionary history (alone), but in the ongoing, self-maintaining organization of living autonomous systems. This chapter proposes to build on this foundation to articulate an organismic path to teleosemantics. This path aims to retain the teleosemantic ambition of naturalizing meaning and purpose but shifts the primary locus of normativity from distant evolutionary selection to the present organizational embodied dynamics of sensorimotor life. We call it the *organismic turn*, implying also a gradual descent of philosophy to the materiality of (biological) organization, its embodiment, and embeddedness.

10.2 The organismic turn: shifting foundations for teleosemantics

10.2.1 Limitations of evolutionary grounding: the need for a new perspective

The opening path for an organismic teleosemantics emerges from the convergence of two significant developments within the philosophy of the life sciences. On one hand, the biological—Neo-Darwinian—underpinnings of classical teleosemantics have come under sustained critique in recent decades. On the other hand, this critique has led to the rise of an "organism-centered perspective" in the life sciences (Laland et al., 2015) in which rich theoretical and empirical advances have been made. We have both reasons and resources to move beyond classical teleosemantics. Focusing on the criticisms directed at classical teleosemantics, we can discern two principal lines of argument.

1 For a detailed account of how enactivism relates to the ecological E of 4E cognition, see Heras-Escribano (2021).

(i) First, some scholars contend that SE is insufficient to support the explanatory goals of teleosemantics. (ii) Second, others argue that SE is inadequate in light of recent developments and debates in evolutionary theory.

Regarding point (i), some critics argue that an evolutionary account of function fails to adequately capture the kinds of phenomena that teleosemantics seeks to explain. This line of critique—hereafter referred to as the *intrinsic-grounding problem* of classical teleosemantics—asserts that if the goal is to account for the specific properties that render a system genuinely intentional (or, conversely, to identify the properties absent in non-intentional systems), then the historical and population-level framework of traditional teleosemantics remains largely silent on these matters (Rama, 2022). A canonical illustration of this issue is provided by the Swampman thought experiment (Davidson, 1987). Swampman is a replica of a human being that comes into existence through a sudden, fortuitous process (e.g., a random collision of atoms). The absence of a phylogenetic history appears to generate counterintuitive implications, since, even if it is materially identical to a human, classical teleosemantics is committed to claiming that Swampman does not possess intentionality—that it does not make behavioral errors, perform successful actions, or have true or false beliefs. In light of such a scenario, Bickhard (2000) characterizes teleosemantics as epiphenomenal at the individual level: it fails to illuminate the internal causal processes that ground intentionality within the organism (Mossio et al., 2009). Moreover, normative explanations in the biomedical sciences are typically grounded in analyses of causal processes, their systemic organization, and their embeddedness in environmental contexts, rather than in purely historical accounts (Gerrans, 2021). Rama (2023) extends this critique by emphasizing that the inherently statistical nature of natural selection precludes classical teleosemantics from offering a robust causal foundation for its naturalistic ambitions—rendering SE causally epiphenomenal even at the population level. Due to the *intrinsic-grounding problem*, evolutionary-selectionist models appear insufficient to support the explanatory aspirations of teleosemantics.

Regarding point (ii), a body of critics converges in their rejection of several foundational biological assumptions underpinning Neo-Darwinism, upon which the concept of SE is built. Ongoing debates concerning the necessity of extending or revising the Modern Evolutionary Synthesis (Laland et al., 2015; Lewens, 2019) have highlighted a range of biological phenomena that directly challenge two central theses of SE theory (see Rama, 2025 for a detailed exposition).

The first thesis holds that the function of a trait explains its existence. However, research programs aligned by structuralist thinking—particularly evolutionary developmental biology (evo-devo)—have increasingly emphasized that certain traits may arise due to structural or developmental constraints rather than adaptive selection alone (Griffiths, 2006; Wagner, 2014). This line of thought was articulated in Gould and Lewontin's (1979) celebrated *Spandrels* paper, which argued that some features of organisms may be evolutionary by-products rather than direct adaptations. Further investigations into

developmental constraints (Amundson & Lauder, 1994) and self-organization (Newman, 2023) support the view that, much like spandrels are architectural necessities in church construction, many biological traits may be structurally indispensable for organismal development. By taking the developmental dimension of evolution seriously, it becomes plausible that some traits persist in nature not because they confer a selective advantage, but because they are necessary conditions for viable development (Balari & Lorenzo, 2012; Fodor & Piattelli-Palmarini, 2011). These traits may thus be relatively stable across taxa—such as the conserved tetrapod limb plan—and largely invisible to selection. Consequently, the assumption that trait function always explains trait existence is undermined, posing a significant challenge to the explanatory power of SE within teleosemantic frameworks.

Second, the SE framework maintains that it is natural selection that confers functions upon traits. As articulated by Garson (2016, p. 51; 2019, p. 28), there can be no function without selection. This reflects a strong Neo-Darwinian commitment: variation is assumed to be random—adaptively undirected—and functional explanations are decoupled from the origin of traits. Under this view, function only emerges *post hoc*, once selection has acted upon blind, stochastic variations. From a contemporary perspective, however, this principle has faced growing criticism, particularly when applied to the evolution of intentionality (Christie et al., 2022). Multiple lines of research—including ecological developmental biology (eco-devo), niche construction theory, molecular epigenetics, and developmental psychobiology —have converged on a developmentalist, adaptively oriented framework for understanding trait variation and novelty. At the core of this alternative paradigm is a decisive move away from gene-centrism and from the notion that phenotypic development is merely a passive unfolding of inherited genetic programs. Instead, contemporary biology emphasizes the systemic embeddedness of gene expression within cellular, multicellular, and ecological contexts. Within this framework, the regulation of trait formation can be responsive to functional demands—suggesting that development itself is an adaptive, context-sensitive process (Bouchard, 2013). As a result, functional explanations in evolutionary biology need not be contingent upon prior selection. Rather, development can generate traits for functional reasons independently of selection. From this vantage point, the relationship is reversed: "evolution is adaptive because development is adaptive" (Walsh, 2015, p. 236). Consequently, since the explanation of the very origins of life (Moreno & Ruiz-Mirazo, 2009) to the origins of evolutionary change (Reid, 2007), it may be more accurate to invert the SE principle and propose instead that there is "no selection without function" (García-Valdecasas & Deacon, 2024; Rama, 2025).[2]

2 Note that this problem is not solved by adopting a pluralist view of function in the manner of Shea (2018). It is one thing to argue that there are ontogenetic functions in addition to evolutionary SE functions (e.g., through learning mechanisms). However, it is another thing

10.2.2 The organismic alternative: normativity from organizational self-maintenance

An alternative approach to naturalizing normativity emerges from the study of autonomous systems, manifested in traditions like interactionism (Bickhard, 2000; Bickhard & Terveen, 1995),[3] organizational biology (Bich, 2024; Moreno & Mossio, 2015), and enactivism (Di Paolo et al., 2017; Varela et al., 1991). This perspective defines an autonomous system by its *self-maintaining organizational closure*. Autonomous systems are understood (in the Kantian tradition of his third critique[4]) as emerging from a set of interdependent, mutually enabling processes, so that each process depends (at least) on another one on the network and enables at least another one; with the effect that the network actively sustains itself and distinguishes itself from its environment (Maturana & Varela, 1980). *Normativity is intrinsic* to this process organization; the "proper" functioning of a component or process is determined by its contribution to the continued, far-from-equilibrium or precarious functioning and existence of the whole system. Functions arise from the *dynamic presuppositions* among the system's components on its contribution to self-maintenance (Christensen et al., 2002; Mossio et al., 2009).[5] An analysis of how different processes contribute to self-maintenance at specific rates and coordination delivers a *normative field* for adaptive processes under varying internal and environmental conditions (Barandiaran, 2025).

The best illustration of this principle is given by (proto)cellular models (Barandiaran & Egbert, 2014; Piedrafita et al., 2012; Ruiz-Mirazo & Moreno, 2004; Varela et al., 1974). A cell is basically a network of metabolic reactions that produces itself. None of the reactions would take place (at the same consistent rate) out of the autocatalytic network that the (proto)cell makes possible; and every reaction is dependent upon and contributes to at least another reaction of the network. Altogether, the system creates some of its boundary conditions, like its membrane encapsulating the reaction network and retaining far-from-equilibrium concentrations (Ruiz-Mirazo &

to say, as we do here, that ontogenetic functions influence evolutionary functions because ontogenetic processes are involved in adaptive evolutionary change.

3 The naturalist project advanced by Bickhard throughout his career has stimulated many of the ideas developed here, particularly his insights on emergent normativity. The path toward a non-standard, autonomy-based teleosemantics was originally initiated by him, even if it differs from our proposal in other respects (e.g., we do not adopt his interactionist approach to representations).

4 For a proper historical trace of how Kantian organicism has influenced recent philosophy of biology and cognitive science, see Varela and Weber (2002), Gambarotto and Nahas (2022) and, for a more complexified reading, see Cuciniello (2025).

5 There have been notable attempts to undermine the organizational approach. Some have even claimed that organizational and SE theories are ultimately the same (Artiga & Martínez, 2016), which could be tempting to use against our narrative in this chapter, but this criticism has also been contested (Mossio & Saborido, 2016), and by putting the emphasis on "cross-generational traits" bear little impact on the research program we defend here.

Moreno, 2004); while actively distinguishing itself from its environment. Embodiment matters. Materiality manifests in multiple aspects, energetic and thermodynamic considerations are crucial, autonomous systems channel energy to produce constraints, that in turn make possible the production of more work, including that of moving and ensuring the supply of energy and matter to keep the organization going (Kauffman, 2000; Moreno et al., 1994; Moreno & Etxeberria, 2005). What the system is and what it does (as the enactment of physical work) is deeply intertwined; thereby grounding normativity. There are certain things the system *must do* to become itself. Physiological functions in multicellular organisms (respiration, digestion, motility, etc.) respond to a similar logic of self-maintenance. The function of the heart is not to pump blood because it has been selected for it, but because pumping blood is dynamically and materially presupposed by the rest of the body parts for their existence and operation (and, in a circular fashion, for the maintenance of the heart itself), thus providing an organizational embodiment for functional normativity.

Autonomous *monists* defend a single source of normativity, the biological one, whose closure is fundamentally metabolic. The cognitive domain is then established as a type of function (e.g., representational) that ultimately subserves this normativity (Bickhard, 2000; Christensen & Hooker, 2000). Others have defended an autonomous *pluralism* by which recursively self-maintaining (or organizationally closed) precarious systems can emerge with different domains, embedded-in (and ultimately dependent-on) but still distinct-from the basic biological domain. Some of these domains include the immune system (Varela & Coutinho, 1991), neural and sensorimotor domains (Barandiaran, 2017; Barandiaran & Moreno, 2006; Di Paolo et al., 2017; Smithers, 1997), or the linguistic (Di Paolo et al., 2018) or social (Lewis-Martin, 2022; Luhmann, 1995), generating a variety of normative sources (García & Barandiaran, 2025; Prokop & Barandiaran, under review).

10.2.3 Sensorimotor organization: the 4E locus of cognitive normativity

The principles of autonomy, organizational closure, and normative functionality extend beyond basic biological metabolism to the *sensorimotor domain*. This is a space where neural electrochemical dynamics and their internal and environmental sensory and effector surfaces make possible the emergence of a behavioral domain that gives rise to cognitive normativity. Piaget (1969) provided early insights, framing knowledge as an organizational function, not subordinated to biological self-maintenance, but to the very organization of behavior. Autonomous robotics, dynamical cognitive science, and progress in neuroscience made it possible to formulate the claim more explicitly:

> the specificity of cognitive dynamics […] is given by a particular kind of dynamic organization within the NS [nervous system] and between the NS and the internal and external environment, i.e. the adaptive

> preservation of a web of dynamic sensorimotor structures sustained by continuous interactions with the environment and the body.
>
> (Barandiaran & Moreno, 2006, p. 180).

Building on this, Barandiaran describes *mental life* (2007, 2008) and Di Paolo et al. (2017) *sensorimotor life* as constituted by networks of habits or sensorimotor schemes. Figure 10.1 depicts a minimal approximation to sensorimotor organization. A sensorimotor scheme (Figure 10.1 left) is a structured set of sensorimotor coordinations composed of agent-side (N) and environment-side (N') support structures and mutual dependencies (dashed lines). These sensorimotor schemes or habits are typically *precarious*, meaning they require recurrent enactment to be maintained and strengthened, to avoid decay or disintegration (Egbert & Barandiaran, 2014). Moreover, they embody a basic or elementary sense of mindedness (Segundo-Ortin & Heras-Escribano, 2021). An adaptive sequence of sensorimotor schemes constitutes a strategy, marked by an inherent normative character. A network of sensorimotor schemes and a set of strategies are organized into an *activity* (e.g., cooking, dancing, writing, and building.). The emerging web of sensorimotor activities constitutes the *identity* of a sensorimotor agent (often composed of regional or role identities, like being a parent or a teacher).

The viability of the entire sensorimotor web depends on the appropriate functioning and coherent coordination of its constituent schemes and networks. As Di Paolo et al. (2017, p. 154) argue, "every enacted scheme in the network can have positive or negative consequences for the viability of the whole. The set of structural and functional dependencies between schemes defines the viability conditions for the ongoing maintenance of the sensorimotor network…". This rich organizational framework provides a

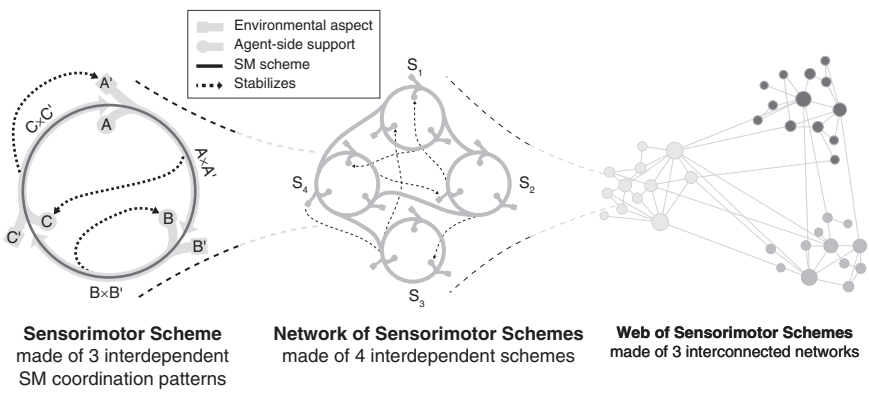

Figure 10.1 Illustration of sensorimotor organization, from a single sensorimotor scheme to the identity of a sensorimotor agent expressed as a web of sensorimotor schemes (see text for details) [Adapted, with permission, from Di Paolo et al. 2017 with a CC-by-sa licence].

foundation for understanding *sensorimotor teleology* and addressing the *problem of normativity*—that is, the *explananda* that concerns teleosemantics. This analysis was carried out in detail in Barandiaran and Rama (2025), including the nature of goals as dynamic attractors, the heterarchical organization of actions, and a detailed analysis of the minimal necessary requirements, for genuinely teleological behavior. As we proposed, "[t]he intrinsic normativity and teleology displayed by autonomous sensorimotor agents, that stems from the dynamic presuppositions within the sensorimotor organization, manifests as a complex set of tensions established by the goal heterarchy" (Barandiaran & Rama, 2025, p. 28). This alternative teleosemantic approach employs conceptual tools from autonomous organizational theory to explain how behavior becomes goal-directed, how errors are constituted as disruptions to this organization, and how purposeful activity emerges.

The crucial point for this chapter is that the fundamental normativity underpinning such teleology is sourced from the current, precarious, and self-maintaining organization of sensorimotor life itself, embedded-within yet distinct-from mere (multi)cellular, metabolic, or physiological organization.

10.3 An organismic twist to semantics in analytic philosophy

The shift proposed by the organismic turn is not merely a matter of revising the biological foundations of teleosemantic theory; it also carries significant implications for long-standing debates within the analytic tradition. Notably, organismic teleosemantics offer a framework through which several internalist insights—previously marginalized by classical teleosemantics—can be rehabilitated. The internalist-externalist debate about the mind has been repeatedly formulated in semantic terms—see Schulte (2023) and Braddon-Mitchell and Jackson (2006) for overviews of the debate.[6] In this context, as Millikan (2006, p. 1) noted, "naturalistic teleological theories are 'externalist' theories of mental content", in which the content of a mental representation is dependent on its referent, a position advanced most prominently by figures such as Bertrand Russell, Saul Kripke, Hilary Putnam, and Jerry Fodor. In contrast, internalist approaches contend that intentional states are defined by the internal relations among their constituent components—an intellectual lineage that traces back to Gottlob Frege. Classical teleosemantics emerged, in part, as a response to perceived shortcomings in internalist theories, particularly their tendency to explain intentionality in terms of

6 Disputes between externalist and internalist approaches to semantics—whether in linguistic terms (primarily during the first half of the 20th century) or mental terms (predominantly in the second half)—have been framed in various ways: as debates between content externalism and content internalism, referentialism and intensionalism, narrow and broad content, or Fregean and non-Fregean perspectives.

other intentional concepts, thereby risking circularity. A central *desideratum* of naturalistic theories of intentionality is that intentional phenomena—whatever their ontological status—must be explicable in non-intentional terms. Thus, *original* intentionality (Haugeland, 1981) must be grounded in causal relationships between representations and their referents. However, this commitment to causalist (and information-theoretic) models has given rise to a persistent problem of normativity: the difficulty of explaining how content can be correct or incorrect in naturalistic terms. As previously discussed, classical teleosemantics attempts to resolve this issue by appealing to evolutionary functions.

The organismic approach challenges Millikan's assertion that teleological theories of intentionality must necessarily be externalist. It argues that alternative teleological frameworks are indeed possible. Our proposal contributes two key elements to a naturalistic theory of intentionality. First, from an organismic standpoint, internalism need not imply cognitive closure or solipsism. Rather, internalism should be interpreted through the lens of autonomy: a cognitive system is constituted by the dynamic *organizational closure* of sensorimotor schemes, its identity. This closure is *extended* into the environment (since sensorimotor schemes are both dependent on agent- and environment-side support structures), yet it remains crucially tied within the agent. It is within the agent's brain-body where most part of the meaning-producing sensorimotor integration takes place. This enables an *asymmetrical interaction*, agency, between the system and its environment. Second, the theory avoids the threat of vicious circularity by drawing on the ontological foundations of autonomy: *emergence*. The central tenet of the internalist perspective is that the normativity of intentional items is determined by their interrelations within a network of intentional states. Against machine-like metaphors of the mind that disclose the mind-brain as a software-hardware relationship, our organicism foundations rely on levels of organization and emergent properties. This ontological approach has its roots in the cybernetic and systems-theoretic view that demonstrates that a holistic conception does not necessarily entail circularity in a problematic sense—aligned thus with Quine's epistemological holism. What an organizational perspective adds is a naturalist framework for understanding normativity. As such, organismic teleosemantics can be productively integrated with various internalist frameworks within the analytic tradition. Among the many internalist theories, two deserve to be specifically mentioned: Conceptual Role Semantics (CRS) and Semantic Networks (SN)—other theories are based on definitional structures, prototypes, or mental imagery.

CRS broadly maintains that "the meaning of a representation is the role of that representation in the cognitive life of the agent" (Block, 1998). Within our organismic framework, this principle might be reformulated as *Sensorimotor Role Semantics*, wherein "the meaning or function of a sensorimotor scheme is defined by its role in the cognitive—and crucially, sensorimotor—life of the agent". This role is not merely inferential, as is often emphasized in standard

CRS accounts, but is instead articulated through the contribution that a sensorimotor scheme makes to the agent's self-maintenance and adaptive organization. Meaning, in this context, is grounded in the dynamically enacted, world-involving, and organizationally constrained patterns of sensorimotor interaction that constitute the agent's viability and mode of existence. The "use" that determines meaning is thus recast as the embodied function of a scheme within the agent's holistic, life-sustaining sensorimotor organization.

Similarly, promising intersections can be anticipated between organismic teleosemantics and SN. Various network-based approaches to semantics—including associative analysis, connectionism, artificial neural networks, and topological models—share deep cybernetic foundations, much like the organizational perspective advanced here. SN approaches typically attribute semantic value to an item based on its topological relations to other items, reflecting a dynamic presupposition inherent in network modeling and relying heavily on a holistic epistemology. A *Sensorimotor Network Semantics* would build upon this insight: the functional meaning of a sensorimotor pattern is determined by its topological relationship with other patterns in the enactment of goal-directed behavior. This perspective aligns closely with the enactive and organizational paradigms, emphasizing meaning as an emergent property of systemic interactions rather than as a fixed or atomistic attribute.

Our central claim is that fruitful cross-pollination may occur between these traditions. Specifically, the conceptual and modeling frameworks developed within enactive and organizational theories could enter into productive dialogue with amended internalist approaches from the analytic tradition—particularly those emphasizing structural or network-based semantics, provided that the nodes of such a network are not encapsulated representational units detached from perception and action, but (re)enactable grounded sensorimotor schemes.

As geneticist Eva Jablonka (2004, p. 366) once said, "it is not sufficient to point out problems with a concept. It is as important to find an alternative that will be free of these difficulties and that will offer at least as fruitful a research program as the old perspective". The organismic turn and the cross-pollination approach advanced here cannot ignore long-standing issues in analytic philosophy (of language and mind). Among these, traditional difficulties in teleosemantics come to the fore, such as *disjunctionitis* (Neander, 2017, p. 149), content indeterminacy (Fodor, 1990), and the problem of novel contents (Garson & Papineau, 2019). It is therefore a collaborative task for future research to delve into the nightmares of analytic philosophy and reconcile them with the sweet dreams of 4E.

10.4 Conclusion: supposing a natural order

As articulated in this chapter, the organismic approach to teleosemantics seeks to re-establish the naturalization of meaning and purpose—a central objective

in analytic philosophy. It does so by relocating the source of normativity from the distal, evolutionary past characteristic of classical teleosemantics to the immediate, dynamic, and self-sustaining organization of autonomous sensorimotor agents. This conceptual shift aims to address long-standing difficulties inherent in evolutionary accounts of normativity while capitalizing on contemporary developments in 4E cognition. It offers a synchronic foundation for understanding how behavior acquires goal-directedness, how errors manifest as disruptions within this organizational coherence, and how purposive activity emerges from the autonomous regulation of action.

The proposal advanced here offers several advantages derived from its organismic foundations. Most notably, it addresses the *intrinsic-grounding problem* outlined in Section 10.2.1, providing a robust framework that accommodates cases such as the Swampman scenario, avoids epiphenomenalism, and sidesteps the limitations of a purely statistical account of normativity. What are the properties that render a system intentional? Conversely, what properties are absent in non-intentional systems? These questions were answered in Section 10.2, where it is argued that the self-organization of sensorimotor schemes—underpinning the maintenance of mental autonomy—is central to intentionality. Furthermore, the organismic grounding of this approach establishes a deep conceptual affinity with organism-centered perspectives in evolutionary theory, such as niche construction theory, developmental systems theory, and ecological developmental biology. As such, this framework is not only philosophically significant but also holds promise for broader applicability and operationalization across the life sciences. Moreover, we can now complement the descent of analytic philosophy from language to cognition to biological materiality with a synthetic philosophy that makes use of complex simulation techniques to deliver explanations of how normative and semantic properties emerge in nature.

The history of 4E cognition could be summarized as the last attempt to claim a "definitive" victory over Cartesian dualism, after its computationalist revival. But less attention has been paid to the methodological Cartesianism. In fact, early analytic philosophy embraced Descartes' third principle with which we opened this chapter:

> "to lead my thoughts in order, beginning by the most simple objects, and the easiest to be known; to rise by little and little, as by steps, even to the knowledge of the most mixt; and *even supposing an Order among those which naturally does not precede one the other*.

Logical atomism and its many descendants turned the analyst's ladder into an ontological blueprint: reality *is* a set of atomic facts and meanings that combine by the rules of logic. The astonishing fruitfulness of that stance was won at a price. Problems of holism, context-sensitivity, and normativity—especially the puzzles of error and misrepresentation—grew in the cracks

left by an architecture too linear for lived cognition. Classical teleosemantics tried to patch those cracks by appealing to the *historical* order of natural selection; but the more deeply it tied content to remote evolutionary antecedents, the less grip it had on the synchronic, organism-bound norms that emerge from the natural order of organic beings. It is the turn of organicism to try to explain how meaning emerges in the physical universe. Time for an organismic turn in a synthetic becoming of analytic philosophy.

Acknowledgments

XEB and TR acknowledge IAS Research group funding IT1668-22 from the Basque Government grants PID2019–104576GB-I00 for the project Outonomy, and PID2023–147251NB-I00 for the project Outagencies funded by MCIU/AEI/10.13039/501100011033. TR also acknowledges the National Agency of Investigation and Innovation (Uruguay), grant PD_NAC_2023_1_176930. We are deeply grateful to Manuel Heras-Escribano for organizing the workshop in Granada and for careful revision and edition of this chapter and the whole book.

References

Amundson, R., & Lauder, G. V. (1994). Function without purpose: The uses of causal role function in evolutionary biology. *Biology & Philosophy*, *9*(4), 443–469. https://doi.org/10.1007/BF00850375

Artiga, M., & Martínez, M. (2016). The organizational account of function is an etiological account of function. *Acta Biotheoretica*, *64*(2), 105–117. https://doi.org/10.1007/s10441-015-9256-x

Ayala, F. J. (1970). Teleological explanations in evolutionary biology. *Philosophy of Science*, *37*(1), 1–15. https://doi.org/10.1086/288276

Balari, S., & Lorenzo, G. (2012). *Computational phenotypes: Towards an evolutionary developmental biolinguistics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199665464.001.0001

Barandiaran, X. E. (2002). *Adaptive Behaviour, Autonomy and Value systems. Normative function in dynamical adaptive systems* [MSc Thesis, University of Sussex]. https://xabierbarandiaran.files.wordpress.com/2022/06/barandiaran_-_2002_adaptive_behavior_autonomy_and_value_systems_-_msc_thesis.pdf

Barandiaran, X. E. (2007). Mental Life: Conceptual models and synthetic methodologies for a post-cognitivist psychology. In B. Wallace, A. Ross, J. Davies, & T. Anderson (Eds.), *The World, the Mind and the Body: Psychology after cognitivism* (pp. 49–90). Imprint Academic.

Barandiaran, X. E. (2008). *Mental Life: A naturalized approach to the autonomy of cognitive agents*. [PhD Thesis, University of the Basque Country (UPV-EHU)]. https://xabier.barandiaran.net/phdthesis/

Barandiaran, X. E. (2017). Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi*, *36*(3), 409–430. https://doi.org/10.1007/s11245-016-9365-4

Barandiaran, X. E. (2025). Organizational accounts of malfunction: The dual-order approach and the normative field alternative. *Biological Theory*. https://doi.org/10.1007/s13752-025-00500-z

Barandiaran, X. E., & Egbert, M. D. (2014). Norm-establishing and norm-following in autonomous agency. *Artificial Life*, *20*(1), 5–28. https://doi.org/10.1162/ARTL_a_00094

Barandiaran, X. E., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior*, *14*(2), 171–185. https://doi.org/10.1177/105971230601400208

Barandiaran, X. E., & Rama, T. (2025). *Sensorimotor teleology and goal-directedness. An organismic framework for normative behaviour*. https://philsci-archive.pitt.edu/id/eprint/25369

Barrett, N. F. (2017). The normative turn in enactive theory: An examination of its roots and implications. *Topoi*, *36*(3), 431–443. https://doi.org/10.1007/s11245-015-9355-y

Bich, L. (2024). *Biological organization* (1st edition). Cambridge University Press. https://doi.org/10.1017/9781009393959

Bickhard, M. H. (2000). Autonomy, function, and representation. *Communication and Cognition—Artificial Intelligence*, *17*(3–4), 111–131.

Bickhard, M. H., & Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*. Elsevier.

Block, N. (1998). Conceptual role semantics. In *Routledge Encyclopedia of Philosophy* (pp. 242–256). Routledge.

Bouchard, F. (2013). How ecosystem evolution strengthens the case for functional pluralism. In P. Huneman (Ed.), *Functions: Selection and Mechanisms* (Vol. 363, pp. 83–95). Springer Netherlands. https://doi.org/10.1007/978-94-007-5304-4_5

Braddon-Mitchell, D., & Jackson, F. (2006). *Philosophy of Mind and Cognition: An Introduction*. Wiley.

Brentano, F. (1995). *Psychology from an Empirical Standpoint.(trans. AC Rancurello, DB Terrell, & LL McAlister).(Original work published 1874)*. Routledge.

Christensen, W. D., Bickhard, M. H., & The Hegeler Institute. (2002). The process dynamics of normative function. *Monist*, *85*(1), 3–28. https://doi.org/10.5840/monist20028516

Christensen, W. D., & Hooker, C. A. (2000). Autonomy and the emergence of intelligence: Organised interactive construction. *Communication and Cognition*, *17*(3–4), 133–157.

Christie, J. R., Brusse, C., Bourrat, P., Takacs, P., & Griffiths, P. E. (2022). Are biological traits explained by their 'selected effect' functions? *Australasian Philosophical Review*, *6*(4), 335–359. https://doi.org/10.1080/24740500.2024.2370630

Cuciniello, R. (2025). *The Organisational Approach to Naturalised Biological Teleology: With Kant, Beyond Kant*. https://philarchive.org/rec/CUCWKB

Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association*, *60*(3), 441. https://doi.org/10.2307/3131782

Di Paolo, E. A., Buhrmann, T., & Barandiaran, X. E. (2017). *Sensorimotor Life: An Enactive Proposal*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198786849.001.0001

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic Bodies: The Continuity between Life and Language*. MIT press.

Dretske, F. I. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

Fodor, J. A. (1980). *The Language of Thought* (1st edition). Harvard University Press.

Fodor, J. A., & Piattelli-Palmarini, M. (2011). *What Darwin Got Wrong* (First Picador edition). Farrar, Straus and Giroux.

Gambarotto, A., & Nahas, A. (2022). Teleology and the organism: Kant's controversial legacy for contemporary biology. *Studies in History and Philosophy of Science*, *93*, 47–56. https://doi.org/10.1016/j.shpsa.2022.02.005

García, E., & Barandiaran, X. E. (2025). Varieties of normativity and mental health: An enactive approach. *Synthese*, *205*(2), 1–29. https://doi.org/10.1007/s11229-024-04854-y

García-Valdecasas, M., & Deacon, T. W. (2024). Biological functions are causes, not effects: A critique of selected effects theories. *Studies in History and Philosophy of Science*, *103*, 20–28. https://doi.org/10.1016/j.shpsa.2023.11.002

Gardner, H. (1985). *The Mind's New Science: A History of the Cognitive Revolution* (First edition). Basic Books.

Garson, J. (2016). *A Critical Overview of Biological Functions* (1st edition 2016). Springer.

Garson, J. (2019). *What Biological Functions Are and Why They Matter* (1st edition). Cambridge University Press. https://doi.org/10.1017/9781108560764

Gerrans, P. (2021). Harmful dysfunction and the science of salience: Adaptations and adaptationism. In L. Faucher & D. Forest (Eds.), *Defining Mental Disorder* (pp. 397–414). The MIT Press. https://doi.org/10.7551/mitpress/9949.003.0024

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *205*(1161), 581–598.

Griffiths, P. E. (2006). Function, homology, and character individuation. *Philosophy of Science*, *73*(1), 1–25. https://doi.org/10.1086/510172

Haugeland, J. (1981). Semantic engines: An introduction to mind design. In *Mind Design: Philosophy, Psychology, Artificial Intelligence* (pp. 1–34). The MIT Press.

Heras-Escribano, M. (2021). Pragmatism, enactivism, and ecological psychology: Towards a unified approach to post-cognitivism. *Synthese*, *198*(1), 337–363.

Heras-Escribano, M., Noble, J., & de Pinedo, M. (2014). Enactivism, action and normativity: A Wittgensteinian analysis. *Adaptive Behavior, 23*(1). https://doi.org/10.1177/1059712314557364

Hutto, D. D., & Myin, E. (2012). *Radicalizing Enactivism: Basic Minds Without Content*. MIT Press.

Kauffman, S. A. (2000). *Investigations*. Oxford University Press US.

Kraemer, D. M. (2014). Revisiting recent etiological theories of functions. *Biology & Philosophy*, *29*(5), 747–759. https://doi.org/10.1007/s10539-014-9430-6

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., & Odling-Smee, J. (2015). The extended evolutionary synthesis: Its structure, assumptions and predictions. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1813), 20151019. https://doi.org/10.1098/rspb.2015.1019

Lewens, T. (2019). The Extended Evolutionary Synthesis: What is the debate about, and what might success for the extenders look like? *Biological Journal of the Linnean Society*, *127*(4), 707–721. https://doi.org/10.1093/biolinnean/blz064

Lewis-Martin, J. (2022). What kinds of groups are group agents? *Synthese*, *200*(4), 283. https://doi.org/10.1007/s11229-022-03766-z

Luhmann, N. (1995). *Social Systems*. Stanford University Press.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition*. D. Reidel Publishing Company.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.

Millikan, R. G. (1989). In Defense of proper functions. *Philosophy of Science*, *56*(2), 288–302.

Millikan, R. G. (2006). Mental content, teleological theories of. In *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470018860.s00128

Moreno, A., & Etxeberria, A. (2005). Agency in natural and artificial systems. *Artificial Life*, *11*(1–2), 161–175.

Moreno, A., Etxeberria, A., & Umerez, J. (1994). *Universality without Matter?* 406–410.

Moreno, A., & Mossio, M. (2015). *Biological Autonomy: A Philosophical and Theoretical Enquiry*. Springer.

Moreno, A., & Ruiz-Mirazo, K. (2009). The problem of the emergence of functional diversity in prebiotic evolution. *Biology & Philosophy*, *24*(5), 585–605. https://doi.org/10.1007/s10539-009-9178-6

Mossio, M., & Saborido, C. (2016). Functions, organization and etiology: A reply to Artiga and Martinez. *Acta Biotheoretica*, *64*(3), 263–275. https://doi.org/10.1007/s10441-016-9283-2

Mossio, M., Saborido, C., & Moreno, A. (2009). An organizational account of biological functions. *The British Journal for the Philosophy of Science*, *60*(4), 813–841. https://doi.org/10.1093/bjps/axp036

Neander, K. (1991). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science*, *58*(2), 168–184. https://doi.org/10.1086/289610

Neander, K. (2017). *A Mark of the Mental: In Defense of Informational Teleosemantics*. MIT Press.

Newman, S. A. (2023). Inherency and agency in the origin and evolution of biological functions. *Biological Journal of the Linnean Society*, *139*(4), 487–502. https://doi.org/10.1093/biolinnean/blac109

Papineau, D. (1987). *Reality and Representation*. B. Blackwell.

Piaget, J. (1969). *Biology and Knowledge*. University of Chicago Press.

Piedrafita, G., Ruiz-Mirazo, K., Monnard, P.-A., Cornish-Bowden, A., & Montero, F. (2012). Viability conditions for a compartmentalized protometabolic system: A semi-empirical approach. *PLOS ONE*, *7*(6), e39480. https://doi.org/10.1371/journal.pone.0039480

Prokop, M., & Barandiaran, X. E. (Under review). Sensorimotor norms and social norms: A pluralistic proposal. *Submitted to Review of Philosophy and Psychology*.

Putnam, H. (1975). *The Meaning of "Meaning."* https://hdl.handle.net/11299/185225

Quine, W. (1969). Epistemology naturalized. *Ontological Relativity and Other Essays*, *13*(3), 69–90.

Rama, T. (2023). Evolutionary causation and teleosemantics. In J. M. Viejo & M. Sanjuán (Eds.), *Life and Mind* (Vol. 8, pp. 301–329). Springer International Publishing. https://doi.org/10.1007/978-3-031-30304-3_14

Rama, T. (2022). *Agential Teleosemantics. PhD Dissertation*. Autonomous University of Barcelona.

Rama, T. (2025). Function and selection beyond externalism: Addressing the impact of explanatory internalism in the selected-effect theory. *Journal for General Philosophy of Science*. In press.

Reid, R. G. B. (2007). *Biological Emergences: Evolution by Natural Experiment*. MIT Press.

Rorty, R. M. (Ed.). (1967). *Linguistic Turn: Recent Essays in Philosophical Method* (1st edition). University of Chicago Press.

Ruiz-Mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life*, *10*(3), 235–259. https://doi.org/10.1162/1064546041255584

Ruse, M. E. (1971). Functional statements in biology. *Philosophy of Science*, *38*(1), 87–95. https://doi.org/10.1086/288342

Schulte, P. (2023). Mental content. *Elements in Philosophy of Mind*. https://doi.org/10.1017/9781009217286

Segundo-Ortin, M., & Heras-Escribano, M. (2021). Neither mindful nor mindless, but minded: Habits, ecological psychology, and skilled performance. *Synthese*, *199*(3), 10109–10133. https://doi.org/10.1007/s11229-021-03238-w

Shea, N. (2018). *Representation in Cognitive Science* (1st edition). Oxford University Press.

Smithers, T. (1997). Autonomy in robots and other agents. *Brain and Cognition*, *34*(1), 88–106.

Thompson, E. (2010). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Belknap.

Varela, F. J., & Coutinho, A. (1991). Second generation immune networks. *Immunology Today*, *12*(5), 159–166. https://doi.org/10.1016/S0167-5699(05)80046-5

Varela, F. J., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, *5*(4), 187.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Villalobos, M., & Palacios, S. (2021). Autopoietic theory, enactivism, and their incommensurable marks of the cognitive. *Synthese*, *198*(1), 71–87. https://doi.org/10.1007/s11229-019-02376-6

Wagner, G. P. (2014). *Homology, Genes, and Evolutionary Innovation*. Princeton University Press.

Walsh, D. M. (2015). *Organisms, Agency, and Evolution*. Cambridge University Press. https://doi.org/10.1017/CBO9781316402719

Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, *1*(2), 97–125.

Wright, L. (1976). *Teleological Explanations: An Etiological Analysis of Goals and Functions*. University of California Press.

11 Communication beyond Inferentialism and Individualism

Glenda Satne

There are many varieties of joint action. While some of these require little communication and exchange between participants, communication can make joint action smoother and help avoid misunderstandings. But the links between communication and joint action run deeper. Communication itself can be seen as a collaborative activity.

The conception of communication as a shared intentional activity was first advanced by Grice (1957, 1969), and later elaborated by Sperber and Wilson (1986), Clark (1996), and Tomasello (2008). These thinkers conceive communication as a complex intentional process. What Tomasello (2008) refers to as the "social infrastructure of human communication". This perspective sees communication not only as a transactional act, but as a dynamic form of collaborative activity, where speakers and listeners are engaged in interaction.

This chapter aims to critically explore this social and argue that the traditional Gricean model, while influential, is not fully equipped to account for the nature and dynamics of this foundational interaction. The discussion will present an alternative view inspired by recent philosophical debates, particularly those concerning the second-person perspective. This alternative view emphasises the embodied, embedded and active character of intersubjective engagements and challenges the traditional third-personal frameworks that have dominated philosophical and scientific theories of communication. These third-person views emphasize observation, inference, and theory, often prioritizing the cognitive and individualistic aspects of communication, which this chapter seeks to critique.

11.1 The Gricean model

At the heart of the Gricean model is a simple framework involving a speaker (U), an utterance (x), and an audience (A).

Grice (1969) famously defined communication in these terms:

DOI: 10.4324/9781003649359-14
This chapter has been made available under a CC BY 4.0 license.

"'(U) meant something by uttering (x)' is true iff, given some audience (A), U uttered (x) intending:

1 A to produce a particular response R;
2 A to think (recognize) that U intends (1);
3 A to fulfill (1) on the basis of his fulfillment of (2)" (Grice, 1969/1991: 92).

In this framework, successful communication requires that the audience (A) recognizes the speaker's (U) intentions and responds appropriately. The process is inherently inferential—the audience must infer the speaker's intentions and derive meaning from the utterance based on contextual information and common knowledge. The model proposes an inferential structure exploited by both speaker and audience, through which individuals transmit their individual mental states to others.

Thus, the Gricean model can be characterized by two key features: *inferentialism* and *individualism*.

The first key feature of the Gricean model is its conception of communication as based on inferential processes. Both the speaker and the audience engage in inferential reasoning to encode and decode meaning. Communication, therefore, is not just a matter of direct transmission of content but involves interpretation and inference about the mental states of an interlocutor. The speaker intends to convey meaning, and the audience infers that meaning by interpreting the speaker's utterance within a specific context.

The second key feature of the Gricean model is *individualism*. In this view, the communicative act is largely understood in terms of the mental states of individuals. "(x)" is in the Gricean model an expression of a mental state that (U) purports to communicate, e.g., in asserting p, U is expressing the belief that p, and the aim of communication, the result of the process if all goes well, is to produce a state of mental recognition of (x) in (A): that p is the case, or more minimally, that U believes that p. Thus, in this model, the inferential process of communication is driven by individual mental states of transmitting individual mental contents and recognizing them by generating, if not matching mental states, at least mental states with the same meaning but a different mental attitude, that is, e.g., for a belief, an attribution of such belief.

This framework presumes that the mental content being communicated is distinct and separable from the social interaction itself, and communication is essentially a process of mental matching: the audience infers and matches the speaker's mental state, whether that's a belief, a desire, or an intention.

Michael Tomasello (2008) presents a critical refinement of the Gricean framework, arguing that the biological roots of human communication lie in the shared cooperative activities that preceded and gave rise to more sophisticated linguistic forms of communication. The target of his analysis is the

understanding of the ontogenetic and phylogenetic roots of human communication. According to Tomasello, "(t)he most fundamental aspects of human communication are seen as biological adaptations for cooperation and social interaction in general" (Tomasello, 2008: 11).

Tomasello sees communication as a fundamental form of collaborative activity itself rooted in shared cooperative activities of a more basic form, e.g., joint action. These activities and the capacities they presuppose together provide the platform for even more sophisticated forms of shared activity to emerge, especially those dependent on linguistic conventions. Thus, in this view, the ability to engage in simple forms of communication—e.g., pointing and pantomiming—can be thought to be prior in development compared to other abilities for shared activity like following social conventions and communicating linguistically.

Tomasello's insight is that communication itself emerges out of the shared intentionality of humans engaging in joint action. The communicative acts of pointing, showing, or mimicking are not just individual acts of signaling but are socially shared, requiring a mutual understanding of intent between participants in face-to-face social interaction. Tomasello emphasizes that these primitive forms of communication develop alongside joint actions and cooperative behaviors, providing the scaffolding for more sophisticated communicative activities grounded in social conventions.

While Tomasello acknowledges that communication in its more complex, linguistic form (e.g., language use) depends on shared conventions, he argues that even these forms of communication have their origins in the cooperative and intentional activities that humans engage in from early developmental stages. Therefore, the Gricean model, with its focus on individualism and mental inferencing, cannot fully account for the social nature of communication, which is always co-constructed by the interaction of individuals.

Tomasello's cooperative model of human communication (Tomasello, 2008: 97ff) posits that communication is an inherently social and cooperative process, involving a communicator and a recipient. The communicative exchange unfolds through a series of stages, shaped by both cognitive and social dimensions of interaction. These stages, as described by Tomasello, are as follows:

1 Individual goals: the first step in communication arises from individual goals that drive social interaction. As Tomasello (2008) explains, each person has goals that lead them to interact with others. For example, I may seek to obtain assistance, share information, or seek emotional support from another individual.
2 Social intention/motives: once goals are established, they give rise to a social intention: "I feel that you can help me on this occasion with one or more of them, by helping me or accepting my offer of information, or sharing attitudes with me" (Tomasello, 2008: 98). This is the realization that the recipient might be able to assist in achieving one's goals.

3 Communicative intention: the social intention is then expressed in the form of a communicative intention: "I decide to make mutually manifest to us (in our current joint attentional frame) a communicative act" (Tomasello, 2008: 98). This marks the intentional act of communicating with the other person, signaling to them that they are the intended recipient of the communicative act.
4 Referential intention: once the communicative act takes place, attention is drawn to an external referent:

> "I draw your attention to some referential situation in the external world—my *referential intention*—which is designed (along with some expression of motive) to lead you to infer my social intention via processes of cooperative reasoning, since you are naturally motivated to find out why I want to communicate with you (based on mutual assumptions or norms of cooperation that are common ground).
> (Tomasello, 2008: 99)

This process comes in stages in phylogeny and ontogeny, and it is made possible because of the creation of a common ground (including social motives) through joint attention.

If we look at the process from the side of the recipient, we can identify the following stages:

1 Identification of the referent: the recipient's first task is to identify the referent or object of attention, typically relying on common ground: "First attempt to identify my referent, typically within the space of our common ground" (Tomasello, 2008: 99).
2 Inference of social intention: following the identification of the referent, the recipient attempts to infer the underlying social intention of the communicator: "From there attempt to infer my underlying social intention, also typically by relating it to our common ground" (Tomasello, 2008: 99).
3 Decision to cooperate: finally, the recipient decides whether to cooperate based on their understanding of the social intention: "Assuming you have comprehended my social intention, you decide whether or not to cooperate as expected" (Tomasello, 2008: 99). This decision to engage in the cooperative act is based on shared assumptions about helpfulness and reciprocity.

In short, from the point of view of the conditions required for the communicative act to operate, that is, for communication to be successful, we can emphasize the following: first, there are individual motivations or social motives of cooperation that motivate communication; second, in the communicator-receiver relationship, joint intentionality operates, as they are involved in the cooperative act of communicating, third, there is shared

knowledge or conceptual common ground (I know you know and you know that I know…that I utter "U" with the intentions that…) that materializes in joint attention; and fourth, the understanding of the communicative act operates based on shared cooperative norms and the ability of cooperative reasoning.

Within this framework, the acquisition of language is based on the development of the ability to capture the communicative intentions of the other.

Approaching the first year of life, infants exercise linguistic abilities that were not possible before and that are built in the flow of social interaction. The child can now access the linguistic conventions required to be competent in the language in as far as s/he engages in social interactions that involve the ability to interpret and understand the Gricean communicative intentions of adults through exploiting the common ground. (Tomasello, 2008: 130ff)

The common ground refers to the shared understanding between the speaker and listener, which is crucial for interpreting the meaning of utterances and actions. Tomasello notes,

> [t]his interpretation is given additional credence by evidence that one-year-old infants understand the basics of the Gricean communicative intention that 'we know together' or it is 'mutually manifest' that I want something from you—based crucially on mutual expectations of helpfulness
>
> (Tomasello, 2008: 130)

Infants around their first birthday clearly produce communicative acts "for another person". This involves behaviors like making sure the other person is paying attention, directing the act to them, and making eye contact. These early communicative gestures indicate that the infant recognizes the ostensive cues (i.e., cues meant to indicate that the act is directed at someone and carries a communicative purpose) that others use toward them (Liszkowski et al., 2008). Also, they seem to recognize such ostensive cues when produced by others as designating acts that are "for" them. According to Csibra (2003) studies on infants' recognition of the communicative/pedagogical intentions of partners in interaction, infants from around their first birthday both produce and recognize these ostensive cues, highlighting their growing awareness of others' communicative and pedagogical intentions (see Tomasello, 2008: 130).

According to Tomasello, the acquisition of language is closely linked to the child's ability to engage in joint attention, mind reading (i.e., understanding the mental states of others), and a form of cultural learning referred to as reversed role imitation (Tomasello, 2003: 21). These skills are necessary for interpreting the communicative intentions of others, which, in turn, are required for meaningful social interaction and language development. By 12–14 months of age, children begin to capture the social intentions of the speaker and locate the object of the referential intention, while also engaging

in cooperative reasoning to infer the communicative intention behind the speaker's actions. This is an unfolding process that will culminate at about three to four years of age.

As children continue to develop these skills, they begin to engage in more recursive thinking and reasoning. Recursivity refers to the ability to understand and process multiple levels of intention within a communicative act.

Tomasello (2008) explains that Gricean communicative intentions are inherently recursive: for example, a speaker may want the listener to know something (e.g., that their friend is approaching), but the speaker's deeper communicative intention is that the listener knows that the speaker wants them to know this. This recursive structure of communication implies a multi-layered understanding of intentions, where intentions are nested inside each other.

As Tomasello states, recursivity is implied in common ground as well as in Gricean communicative intentions:

> First, the creation of common ground and/or joint attention between two persons requires that each of them sees, knows, or attends to things that she knows the other sees, knows, or attends to as well—and knows that the other knows this about her as well, and so on recursively potentially ad infinitum. Also, the Gricean communicative intention is clearly recursive—at least to several levels. Thus, in Sperber and Wilson's (1986) account, in a declarative speech act I want you to know something (e.g., that your friend approaches), but my communicative intention is that you know that I want this. In this analysis, therefore, communicative intentions are either third or fourth order (depending on how one counts): I want1 you to know2 I want3 you to know4 your friend approaches.
>
> (Tomasello, 2008: 94)

Finally, norms of cooperative communication in Tomasello's account are also recursive. The mutual expectation of cooperation, Tomasello argues, is deeply embedded in the norms of communication, where individuals both expect and are expected to engage in cooperative communicative exchanges. Tomasello adds,

> [t]he motivational structure of human communication is also recursive in that we both know together that we both are helpful—so that you are expecting me to expect you (and so on with further embeddings as needed) to be helpful. Most clearly, such recursivity is absolutely required for norms of cooperation in which it is mutually expected by everyone (including oneself) that everyone will be a cooperative communicative partner.
>
> (Tomasello, 2008: 94)

In sum, Tomasello's model of ontogeny stresses the importance of shared intentionality, joint attention, and recursive reasoning in the development of communication.

While Tomasello's cooperative model of human communication provides a compelling framework for understanding the cognitive and social processes involved in communication, it has been subject to significant criticisms, particularly regarding the cognitive demands it places on young children. Several authors (Tollefsen, 2005; Brownell et al., 2006; Michael et al., 2014; Zahavi & Satne, 2015; Satne & Salice, 2020) have raised concerns about Tomasello's model, arguing that it requires too complex a set of cognitive abilities for children to participate in basic forms of shared communicative activity. In the context of understanding root forms of communication, this concern with cognitive demandingness questions the assumption that children need to possess a full understanding of communicative, social, and referential intentions, as well as common knowledge of others' mental states and the norms of cooperation to engage in communication. This requirement for a recursive understanding of nested intentions—where individuals must grasp third- or fourth-order beliefs—appears to be a substantial challenge, especially given the cognitive limitations of young children (Carpenter & Liebal, 2012).

At the core of the criticism is Tomasello's commitment to complex cognitive abilities, which are seen as essential for communication to be possible. In this framework, children must not only recognize intentions (both social and referential) but also possess the ability to comprehend the mental states and motivations of others, all while adhering to norms of cooperation that require recursive reasoning. This leads to the assertion that Theory of Mind (ToM)—the capacity to attribute mental states to oneself and others—is a crucial prerequisite for communication to be possible. However, research in social cognition suggests that children under the age of four show limited abilities in ToM (Low & Perner, 2012). Even though some argue that certain "indirect" tests suggest rudimentary forms of ToM may emerge earlier (Rubio-Fernández & Geurts, 2013), the recursive nature of the beliefs required for common ground and cooperative communication remains a significant challenge for children below this age.

The issue does not solely hinge on when the development of ToM is located, but on the recursive nature of the beliefs necessary for understanding communication. Tomasello's model implies that for effective communication, a child must be capable of understanding and reasoning about complex nested intentions, which goes beyond simple belief attribution. This raises a substantial concern: how can children, especially those under three years old, process such recursive cognitive structures? Carpenter and Liebal (2012) argue that this level of recursive reasoning is beyond the cognitive capacities of young children, making Tomasello's model seem overly demanding from a developmental standpoint.

Tomasello's view shortcomings depend largely on his commitment to a Gricean model of communication. In the next section, we will explore some

of the challenges posed to the Gricean model by revising each of its two key assumptions, namely, *inferentialism* and *individualism*. In so doing, we will be better placed to provide an alternative picture of communication as a shared intentional activity that can account for its root forms in human ontogeny.

11.2 Beyond the Gricean model

11.2.1 Beyond inferentialism

One of the aspects of Grice's model of communication that has given rise to criticisms is his claim that communication involves the interpretation of intentions through inference. Critics like Bar-on (2018) and Medina (2013) challenge this intentional-inferential framework by proposing an alternative conception of communication rooted in expressive behaviors rather than complex inferential reasoning. According to these critics, basic forms of communication do not necessarily require that a communicator consciously intends to convey a message to an audience, nor that the audience engages in complex inferential reasoning about the speaker's mental states. In their view, communication can be understood as a process involving expressive signs that do not require fully formed communicative intentions or deep cognitive representations. According to Bar-On,

> intentional communication need not require the communicator to *intend to communicate some message to her audience* – at least not if by that we mean that she has to have a conception of what her audience thinks, or wants, or intends, etc. *and intentionally to design her communicative behavior so as to accomplish a desired goal*
>
> (Bar-on 2018, emphasis added)

Bar-on (2018) contends that expressive behaviors—such as bodily movements or facial expressions—"are *naturally designed* for the purpose of intersubjective communication, may be sufficient to put communicators on the right path – the behavioral repertoire itself need not be invented or learned" (Bar-on 2018 pp, my emphasis). In contrast to the Gricean model, which sees communication as a conscious, intentional exchange, expressivism holds that bodily movements and expressions can communicate mental states without the need for elaborate mental representations of those states. As Medina 2013 explains,

> [e]xpressive behavior is not self-reflective intentional-inferential communication among rational agents who are representing each other's minds and their contents. The production and uptake of expressive behavior place much weaker representational demands on their producers and responders than self-reflective intentional-inferential communication does.
>
> (Medina, 2013: 326)

Expressivism introduces a radical shift in how we understand mental states in communication. Instead of requiring a reflective, inferential process, it suggests that direct perception of others' emotions and intentions—such as recognizing sadness or desire through facial expressions and body language—is sufficient for communication.

In the same vein, some authors (Zahavi, 2014; Gallagher, 2001, 2015; Thompson, 2001) have criticized inferentialism of the sort defended by Tomasello for offering a "third personal model" of social cognition. They argue that this model denies the possibility of direct knowledge of others' mental states, which they claim is essential for understanding basic forms of communicative interaction. According to this view, emotional states and intentions can be directly perceived in someone's face or bodily movements—there is no need for complex inferential reasoning to understand others' emotional or intentional states. This marks a shift away from the Gricean and Tomasellian reliance on recursive reasoning in communication.

Zahavi and Gallagher argue that a third-person perspective—where communication is viewed as detached and involves inferring another's mental state—fails to capture the immediacy and directness of human interaction. Instead, they propose starting from a second-person perspective, where communication is understood as a face-to-face encounter where mutual understanding is direct and unmediated. In this model, understanding is not inferred from signs or gestures but is based on direct engagement with the other person.

These authors see as problematic that these views, including the Gricean view of communication, assume a third personal, detached perspective as the point of departure of our understanding of others. Having started from the wrong perspective, a misconception of what is required for communication follows. In particular, the need for inferences in these accounts is a consequence or such a wrong point of departure. They propose starting with a second-person perspective, by which they mean face-to-face encounters in which direct and unmediated understanding is possible.

This view challenges the Gricean inferential model by positing that understanding others' intentions, emotions, and other mental states can occur without the need for complex inferences. For example, hand movements are seen as solicitations for particular objects or face expressions as emotional states of sadness or grief. That these acts can be expressive and communicate something to others means, in this view, that bodily movements, for example, hand movements, or face expressions can be "seen as" an expression of mentality.

One might think that the emphasis on expression and direct perception goes some way down the right path in accounting for basic forms communication, one that does not require complex mentalizing, reasoning, and mental inferences. But might these views still be too cognitively demanding for modeling the cooperative roots of communicative practices?

Second-personal accounts so understood remain committed to a central idea that is implicit in the Gricean model, namely that our knowledge of another individual's mental life is based on observation. The difference is that, in this case, the emphasis is on perception rather than inferential reasoning.

Their focus on (perceptual) observation of other people's mental states commits these views to either of two problematic ideas of the development and evolution of social understanding and communication, depending on whether they are robust or modest forms of expressivism.

Robust views (Zahavi, 2014; Gallagher, 2011, 2015; Thompson, 2001; Bar-on, 2018; Medina, 2013) think of social cognition in terms that already require the agents to have some understanding of the mental states of others, e.g., desires, intentions, and beliefs, by which they "see (a bodily movement) as" the expression of a mental state: sadness, grief, reaching for an object, etc. The problem with these views is that they presuppose what needs explaining, namely, that a bodily movement communicates the mental content of an agent to another agent. This leaves the question of how this is possible unaddressed, for it is the starting assumption of the view that this is the case. Even if this was seen as an innate capacity (part of parcel of early forms of social cognition, see e.g., Trevarthen, 1978, 1979), the explanation of its emergence would be pushed back to its evolutionary origins.

Modest views don't assume this understanding is already there. Instead, modest expressivists[1] see bodily movements and facial expressions of other animals (conspecifics or not) as indicating features of the environment and thus giving information about it, as, e.g., smoke can indicate the presence of fire. In contrast to robust forms of expressivism, which presume a certain level of understanding about the agent's intentions, modest expressivism views an agent's perception of another's expressive acts as a recognitional mechanism—a way of responding to signals that indicate features of the environment rather than directly understanding others' mental states.

We can understand these recognitional mechanisms as a result of evolutionary pressures. However, even if this can be seen as a step toward communication and a platform for it, such recognition cannot yet count as communication, and this for two reasons. First, because it posits a completely general mechanism independent of any specific aims of the agents, which is what the explanation of human communication is aiming to explain, namely, purposeful communication between agents; second, because it seems not rich enough to justify talking of understanding of mental states of the agents rather than their targets.

1 Some views within ecological psychology accounts of perception (see Gibson, 1979; Reed, 1996) and radical enactivist accounts of directed intentionality, that rely on Telesemiotics (see Hutto & Myin, 2013, 2017), despite their differences, can be seen as modest expressivists views.

While the modest view offers a more evolutionarily plausible explanation of how communication and mental understanding might have emerged, leaving room for positioning social cooperation as the driving force behind the development of the ability to engage with others as mental agents, the problem with these views is that bodily expressions would remain insufficient to communicate an agent's mentality, as they are too thinly specified and pervasive to signal out the expressions of animals that are *seen as* mentality.

The yet unanswered question is how to make sense of root forms' intentional communication, whenever in ontogeny, or phylogeny, they are exactly placed. For the reasons advanced, in the expressivist views, an answer to this question is already presupposed, as in robust expressivism, or completely absent, as in modest expressivist views.

Ultimately, the criticisms of Tomasello's model and the Gricean inferential framework that he relies on the one hand, and the concerns with the alternative, expressivist views, on the other, point to a need for a more relational and engagement-based understanding of communication.

In line with this, I want to suggest that fundamental forms of social engagement are at the basis of basic forms of mental understanding. That is to claim that social cooperation, a central motive of evolutionary studies into human cognition, is the driving force of their evolutionary trajectory. Thus, like Tomasello, my suggestion is that very plausibly, engaging in some shared activities is essential for acquiring the ability to understand other persons as mental agents, i.e., as agents with mental life, with whom we communicate.

It is apparent that shared activities are pervasive both in the phylogenesis of human forms of life, going back to the social lives of Chimps and Bonobos with whom we share our last common ancestor, as well as in the life of infants that are highly dependent on interactions with others who care for them. Can we understand the development of the kind of mental competence that is at issue in early forms of communication in terms of the emergence and development of root forms of social engagement?

To do so, we need to revise the second assumption underpinning Grice's and Tomasello's picture of communication, namely, *individualism*.

11.2.2 Beyond individualism

As stated above, both Grice's model of communication and Tomasello's account of ontogeny and phylogeny based on it are committed to an individualistic understanding of communication. This approach treats communication as a series of discrete acts performed by separate individuals: one act by the speaker (utterer) and another by the listener (audience).

However, authors like Richard Moran (2018), Michael Thompson (Ms. 2012), and Sebastian Rödl (2015) challenge this individualistic view by emphasizing the importance of a second-person interaction in communication, which they argue offers a more accurate and comprehensive account of shared communicative acts, compared to the third-personal account that

results from the assumption that communication is a cooperative endeavor between two separate individuals that observe and interpret one another.

The individualistic picture, particularly in Gricean models, separates the acts of the speaker and the listener, creating a gap between what is said and what is understood. According to Thompson, this gap exists because each party operates as an individual agent who must infer the other's communicative intentions. The listener, for instance, must rely on their own the ability to attribute mental states to others—their own ToM—in order to infer the speaker's intentions, goals, and the meaning of the message. This mentalizing process requires that individuals possess a complex understanding of the situational context, the intentions of the other, and the norms of cooperation. Such a model places substantial cognitive demands on both parties involved in communication, as they must continually infer each other's intentions and mental states.

However, this gap—created by separating the communicative acts of the speaker and listener—can be problematic. It assumes that understanding is always an individual and internal process, relying on mental inferences to bridge the divide between what is said and what is understood. The individualistic approach fails to capture the shared nature of communication, wherein the interaction is not just a series of discrete acts but a cooperative and relational process.

To overcome this individualistic framework, Moran, Thompson, and Rödl propose thinking of communication as a shared act in which both the speaker and the listener are co-authors of the communicative exchange. This view challenges the assumption that communication is primarily about one individual trying to infer the intentions of the other. Instead, it emphasizes that communication is inherently relational and social, with both parties jointly engaged in creating meaning.

As Rödl (2015) and Moran (2018) argue, communication should be understood as more than just a matter of two distinct acts performed by separate individuals. It should be seen as a shared endeavor, akin to other social actions such as marrying, promising, or buying. In these actions, the participants are not merely performing isolated acts in relation to one another; rather, they are both performing one single act. As Thompson puts it, "communicating like marrying, promising, buying, etc. [share some aspects, namely,] 1. that they are relational and social, and 2. that the same cognitive and volitional requirement must hold of the other party" (Thompson, Ms., my additions).

By framing communication within the context of interpersonal social actions, such as promising, marrying, or giving a gift, we can identify several key features that characterize the second-person model of communication (see Satne, 2021[2]).

2 I am following here the model of second personal joint action that I developed in more detailed in Satne 2021.

This account is motivated by a number of examples of joint action, such as marrying and promising, that is, interpersonal interactions which are reciprocal. The activities that have served to illustrate other theories of joint action, such as "painting a house" (cfr. Bratman, 1992), "walking together" (cfr. Gilbert, 1990), and "pulling a pram onto a bus" (cfr. Butterfill, 2012), are activities that individuals might perform alone, in parallel as well as together with others. The key to the debate between the competing theories is to determine the ingredients that are necessary and/or sufficient to make sense of the case in which these activities are done together, rather than alone or in parallel. All of these theories assume an individualistic understanding of the kind of activities at issue, namely, that they can be done alone. In contrast, in order to understand communication, unlike in the individualist picture, I focus on interpersonal activities which cannot be carried out alone or in parallel, these are activities that are necessarily shared with other interactants in reciprocal and symmetrical relations.

The set of activities at issue includes activities such as playing catch, dancing tango, playing tennis, and breastfeeding. In all these cases, agents engage in joint activities that they understand as being necessarily shared with other agents.

There are a number of features that are distinctive of these interpersonal social activities. They are:

i Reciprocal.

These interpersonal actions are necessarily reciprocal, i.e., for any agent's action, there is a corresponding action of the other agent; they are two sides of one coin. This means that for each action of one of the agents, there is a corresponding action (or coordinated reaction) of the other agent. Take the example of playing catch. When player 1 throws the ball, player 2 is supposed to catch it; player 1 has the goal of throwing the ball while player 2 has a corresponding goal of catching it. Each individual's action in playing catch is associated with a corresponding re/action on the part of the other agent. Thus, each individual's contribution to the joint activity is tailored by the expectations she has of the other's corresponding reaction to it.

ii Unintelligible if not shared.

These activities cannot be done alone, for when agents engage in these activities, they engage in activities whose goals they understand as being necessarily shared with the other agent. This is so because the actions each agent performs are only intelligible against the background of shared goals that both agents have in engaging in the activity. Take, for example, playing catch. It is both the throwing and the catching that together constitute one case of the action of playing catch. For an action to be a case of playing catch, more than what player 1 or player 2 alone is respectively doing is required. This is a kind of action that would

be best described as an agent, player 1, throwing a ball, and another agent, player 2, catching it, where the two actions are coordinated, and mutatis mutandis for other throwings and catchings. Their throwings and catchings are only intelligible for each agent as actions within the shared activity of playing catch that both agents are performing. Thus, the actions of the individual agents (a) need to be part of a more complex activity in which both agents are taking part and (b) each agent necessarily understands her action as part of a more complex action that incorporates the action of the other.

iii Their goals or ends are collective.

The individual goals that each individual in the shared activity has is part of, or a means toward, the more complex goal that brings them together. The individual's goals are individuated against the background of the shared activity. The individual sub-goals that each individual in the joint action pursues, that is, the goals by which the shared goal is accomplished, depend on a more complex goal that brings all of them together. Take again the case of playing catch: any individual goal within the game, e.g., this particular throw, depends on the shared goal of playing catch, which is the goal of both individuals. To identify the goal to which each individual is individually directed, reference to the shared goal is required.

This suggests that the social infrastructure of human communication should not be understood as a form of joint action performed by two detached individuals acting independently and interpreting one another. Instead, communication should be framed in collective or plural terms, where communication is not a mere sum of individual actions but rather a shared activity of two subjects acting together that emerges from the bodily and enactive interaction between individuals that are immersed and embedded in activities together. In this view, it is "we", both you and I, together, who communicate, marry, promise, and engage in similar social acts, rather than each of us performing separate actions that are merely coordinated *post facto*.

This is because actions of this kind are reciprocal and thus require two or more subjects to have a shared understanding of what they are doing. Without this shared understanding, the individuals could not have had the goals at issue. Thus, these actions are best explained as activities of a plural agent that the different interacting subjects constitute together and not as activities of collections of individuals that might not have a shared understanding of what they are doing, as the individualistic understanding of communication assumes.[3]

3 This does not mean that sometimes we do not communicate with others in the absence of a shared understanding, rather it presupposes that (i) that shared understanding is the primary case, in the light of which cases of interpretations and inference are to be understood and

11.3 The collective infrastructure of human communication

We can now reassess the ontogenesis of human linguistic communication in terms of this embodied, embedded and enactive second-personal account of communication as a shared intentional activity presented in the previous section and assess how it fares compared to Tomasello's view.

Interpersonal social actions can be thought to be the root form of shared intentionality that precedes and grounds Gricean communication.

In this view, these activities are collective from the outset, having different stages of development (see Satne & Salice, 2020; Satne, 2021).

The progression from one to the other form of collective engagement involves the scaffolding of shared intentional activities by ecological and cultural environments and tools, including social norms, especially natural languages, and the exercise of guided and individual practical reasoning, making it possible for more sophisticated reason-guided activities to emerge. Importantly, throughout shared intentional activities are shaped by mutual responsiveness, in which each individual tailors her participation contingently upon the reactions/responses/of others, in such a way that agents are co-authors of their interactions.

More specifically, children's capacities for engaging in collective activities break in several stages' progressive of development. From minimal collective intentionality, based on bodily coordination and emotional tuning (see Satne, 2014; Satne & Salice, 2020; Satne, 2021a, 2021b), to collective intentionality grounded on social norms, and instruction-following, (Gergely & Csibra, 2009; Tomasello, & Racokzy 2003; Schmidt & Tomasello, 2012), to finally, collective intentionality grounded in practical reasoning (Satne, 2021a, 2021b, 2024) and social folk-psychological narratives (Gallagher & Hutto, 2008; Hutto, 2008), to which Gricean communicative cooperative reasoning belongs.

Thus, in this view, minimal collective intentionality, characteristic of root forms of communication, serves as the foundation for more sophisticated forms of communication to emerge and does not require complex cognitive abilities. Rather, it is grounded on bodily forms of coordination and emotional tuning, through which individual contingently responds to one another tailoring their participations in the activities to the expectations of their partners in interaction.

But can the picture offered here deliver the goods? Can we think of the capacity of mental understanding at stake in other views of communication

(ii) the activity of communication builds shared understanding even in the absence of an established common ground, just by the nature of the interaction that is taking place, where both agents respond contingently to each other contributions to a joint activity, that they understand as shared In the case of communication, this means that interlocutors understand that they engaged in the share activity of communicating, even if they do not always or not completely understand what the other has said.

in sufficiently cognitively minimal pathways by going collective in the way I propose? More specifically, can we understand the development of the kind of mental understanding that is at issue in robust expressivism in terms of the emergence and development of basic forms of engagement, or is that understanding already presupposed in the interpersonal interaction's characteristic of minimal collective intentionality?

I cannot provide a full answer here, which will require to present an integral account of the ontogeny and phylogeny of human capacities for social cognition and communication, instead, in what follows, I argue that minimal collective intentionality as outlined here can account for the emergence of communication in ontogeny without the cognitive demandingness that is characteristic of Tomasello's and Grice's accounts.

Minimal collective intentionality does not require prior knowledge of another agent's mental states but depends on the willingness of acting together, interpersonal emotional tuning and the abilities to coordinate bodily.[4]

Children are engaged in shared activities, and so in root forms of communication, since birth. The first activities at issue include interpersonal feeding practices and early interbodily coordination between them and their careers (see Reddy et al., 2013a). These forms of interaction can be thought to be reciprocal, unintelligible if not shared, and informed by goals that are shared with other agents. Yet, the goals relevant for these early forms of engagement and communication can be thought to be "individuated in the interaction". That is, the shared goal is identified only in reciprocally coordinating one's action with another agent, for plausibly, in this case, the agents do not share any further goals that inform their activities. Thus, arguably, no prior representation of the goals of others is required, but shared directed activity can nonetheless take place.

Reddy and colleagues have identified anticipatory adjustments in body posture in children as young as two-month-old when they are about to be picked up by their carers turning their bodies in very specific ways (Reddy et al., 2013b), as well as emotional and intentional responses in two-month-old babies, and the capacity of complying to directives, including linguistic at six-month (see Reddy, 2015; Satne, 2021 for a review). These studies describe how two-month-old babies, before any evidence of shared attention, respond

4 There is a broad range of studies regarding human tendencies to bodily coordinate their movements with one another. There is widespread evidence of patterns of emergent coordination specifically among humans (cfr. Schmidt & Richardson, 2008; Knoblich et al., 2011, see also De Jaegher & Di Paolo, 2007, in which coordination is seen as "part and parcel of the self-production and self-maintenance of the interaction process" De Jaegher & Di Paolo, 2007: 494). These sorts of emergent coordinated patterns can be thought to be at the basis of the root forms of communication as action coordination described here, showing that tendencies of mutual coordination and bodily and physiological synchronizing are likely innate and universal, thus, themselves constituting a very important platform for understanding basic forms of human social cognition (see Satne, 2021 for an argument in this direction).

to others in emotion-rich interactions (Reddy, 2003). Reddy claims that these early interactions provide infants with a kind of know-how understanding of others being intentionally directed toward them (Reddy, 2008).

It is what happens in these early stages in children's engagements with others intentionally, what might be accounted for by invoking the collective intentionality framework here proposed. In particular, capacities for minimal collective intentionality, where no prior representation of the goals of others is required for shared activity to take place. This contrasts with cognitivist views of social cognition and joint action that require agents to represent the goals of other agents in order to interact with them, whether via simulation, inference, or both. Rather, these early interactions can be thought to be a training platform in which the ability to attribute mental states can be thought to progressively unfold by means of further social learning and practice (see Hobson, 2002; Rochat, 2015). Enactivism has provided rich accounts of this developmental process (see, e.g., Gallagher, 2001; Gallagher & Hutto, 2008) that we cannot rehearse here. It should be noted though that according to Enactivism, it is only when a full range of other capacities over and above basic interactive ones are mastered, crucially including linguistic and narrative ones, that a folk-psychological understanding of others—the understanding of others in terms of concepts of mental states such as belief, desire, hope, and their inferential articulations—is acquired.

At the early stages of human development, particularly from six to nine months of age, joint attention to objects of common interest begins to emerge, as evidenced by behaviors such as teasing (Reddy, 2015), social reach involving infants' expectations of goal completion from other agents (see Ramenzoni & Liszkowski, 2016, and overview in Trevarthen, 1979). These forms of social interaction provide platforms for the learning of social norms and practices in which the first words make their appearance (Trevarthen, 1979; Satne & Salice, 2015; Satne & Salice, 2020). From these forms of coordination more complex and derivative forms of exchange of a Gricean kind can be thought to emerge, through the learning of social norms, folk-psychological narratives and cultural practices, within which the learning of natural languages is intertwined, all of them supporting and enabling the development of the kind of mental understanding and complex reasoning that the Gricean model presupposes. The mastery of these capacities starts to take shape at around two to three years of age, and consolidates at four to five years of age (see e.g., Gallagher, 2001; Gallagher & Hutto, 2008, and discussion in Satne, 2021a).

All these different strands of empirical research on interaction and social cognition support and give flesh to the account proposed here showing how the root forms of interaction discussed in this chapter can be thought to constitute, in a wide range of different aspects, minimal forms of communication between agents.

Engagement in reciprocal interactions of the sort described, interpersonal social activities, explains how one-year-old infants can fulfill the requirements

for communication without resourcing to complex reasoning, recursive inferences, or full-fledged ToM abilities. We may think that the abilities that Tomasello identifies in one-year-old children, namely, their ability to stand in the space of common ground with other agents (that "we know together" or it is "mutually manifest" that I want something from you; see Warneken et al., 2006; Warneken & Tomasello, 2007), are not based on complex reasoning abilities of the Gricean kind. Rather, they are developments of the mutual expectations of helpfulness that are *built in* reciprocal embodied interactions, that are embedded in shared activities, as part of their constitutive dynamics of turn-taking and goal sharing. Communicative acts on the part of infants can be understood as acts "for another person", solely based on their seeking the attention of the other, a potential partner in interaction: directing their actions to them, making eye-contact, and tailoring their bodily movements to the other's movements, and so on and forth, as they do when interacting with adults in early pick-ups and teasing at two months of age, and later on, in reaching and pointing with the expectation of others completing their actions, at eight months of age. In the same vein, in virtue of their early and extensive engagement in reciprocal interactions with caregivers, infants can recognize ostensive cues when produced by others, as designating acts that are "for them" to respond to, in the context of face-to-face interactions—e.g. those described in Csibra's 2003 studies on one-year old infants recognition of communicative/pedagogical intentions—without the need of attributing to others complex communicative intentions.

11.4 Conclusion

I have argued that the Gricean model of communication is problematic in its attempt to account for the social infrastructure of root forms of communication. It assumes *individualism* and *inferentialism*, and thereby places too highly cognitive demands on communicative agents. This makes this model unsuited for explaining the ontogeny and evolution of communicative practices. As an alternative, I have considered expressivist accounts of communication, which contest the need of the inferential structure characteristic of the Gricean account for communication. However, I have argued that these views either presuppose what needs explaining, or invoke too thin notions of expressive behavior to account for the sorts of communication that are the target of the Gricean model. Instead, I have suggested that a more promising approach is to understand human communicative practices as grounded on capacities for minimal collective intentionality, where reciprocal second-personal bodily interactions embedded in shared intentional activities are their foundation. This view rejects both the individualistic and inferential assumptions of the Gricean model, providing a better understanding of the developmental pathways through which communication and mental understanding progressively unfold.

References

Bar-On, D. 2018. Communicative intentions, expressive communication, and origins of meaning. In K. Andrews & J. Beck (Eds.), *The Routledge Handbook of Philosophy of Animal Minds* (pp. 301–312). New York: Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315742250-29

Bratman, M. 1992. Shared cooperative activity. *The Philosophical Review 101*(2), 327–341. https://doi.org/10.2307/2185537

Brownell CA, Ramani GB, Zerwas S. 2006. Becoming a social partner with peers: cooperation and social understanding in one- and two-year-olds. Child Dev. 2006 Jul-Aug;77(4):803-21. doi: 10.1111/j.1467-8624.2006.00904.x. PMID: 16942491; PMCID: PMC3351034.

Butterfill, S. 2012. Joint action and development. *Philosophical Quarterly 62*(246), 23–47.

Carpenter, M., & Liebal, K. 2012. Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint Attention. New Developments in Psychology, Philosophy of Mind and Social Neuroscience* (pp. 159–182). Cambridge: The MIT Press.

Clark, H. 1996. *Uses of Language*. Cambridge: Cambridge University Press.

Csibra, G. 2003. Teleological and referential understanding of action in infancy. *Philosophical Transaction Royal Society Londan B Biological Science* 2003 Mar 29;*358*(1431), 447–458. https://doi.org/10.1098/rstb.2002.1235. PMID: 12689372; PMCID: PMC1693135.

Csibra, G., & Gergely, G. 2009. Natural pedagogy. *Trends in Cognitive Sciences 13*(4):148–153.

De Jaegher, H., & Di Paolo, E. 2007. Participatory sense-making. *Phenomenology and the Cognitive Sciences 6*(4), 485–507.

Gallagher, S. 2001. The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies 8*(5–7), 83–108.

Gallagher, S. 2008. Direct perception in the intersubjective context. *Consciousness and Cognition 17*(2), 535–543.

Gallagher, S. 2015. The problem with 3 year-olds. *Journal of Consciousness Studies 22*(1–2), 2015, 160–182

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

Gibson, J. 1979. *The Ecological Approach to Visual Perception*. New York: Psychology Press.

Gilbert, M. 1990. Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy 15*(1), 1–14.

Grice, H. Paul. 1957. Meaning. *Philosophical Review 66*(3), 377–388.

Grice, H. Paul 1969. Utterer's meaning and intentions. *Philosophical Review 78*(2), 147–177.

Hobson, P. 2002. *The Cradle of Thought*. Oxford: Macmillan.

Hutto, D. 2008. *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Bradford: MIT Press.

Hutto, D., & Myin, E. 2013. *Radicalizing enactivism. Basic minds without content*. Cambridge: MIT Press.

Hutto, Daniel D., & Myin, E. 2017. *Evolving Enactivism: Basic Minds Meet Content*. Cambridge: MIT Press.

Knoblich, G., Butterfill, S., & Sebanz, N. 2011. Psychological research on joint action: Theory and data. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 59–101). Burlington, MA: Academic Press.

Liszkowski, U., Carpenter, M., & Tomasello, M. 2008. Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition* 108, 732–739. https://doi.org/10.1016/j.cognition.2008.06.013

Low, J., & Perner, J. 2012. Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology* 30, 1–13.

Medina, J. 2013. An Enactivist approach to the imagination: Embodied enactments and "fictional emotions". *American Philosophical Quarterly 50*(3), 317.

Michael, J., Christensen, W., & Overgaard, S. 2014. Mindreading as social expertise. *Synthese 191*(5), 1–24.

Moran, R. 2018. *The Exchange of Words: Speech, Testimony, and Intersubjectivity*. Oxford: Oxford UP.

Ramenzoni, V., & Liszkowski, U. 2016. The social reach: 8-month-olds reach for unobtainable objects in the presence of another person. *Psychological Science* 2016 Sep; *27*(9), 1278–1285.

Reddy, V. 2003. On being the object of attention: Implications for self-other consciousness. *Trends in Cognitive Sciences 7*(9), 397–402.

Reddy, V. 2008. *How Infants Know Minds*. Cambridge, MA: Harvard UP.

Reddy, V. 2015. Joining intentions in infancy. *Journal of Consciousness Studies 22*(1–2), 24–44.

Reddy, V., Liebal, K., Hicks, K., Jonnalagadda, S., & Chintalapuri, B. 2013a. The emergent practice of infant compliance: An exploration in two cultures. *Developmental Psychology 49*(9), 1754–1762.

Reddy, V., Markova, G., & Wallot, S. 2013b. Anticipatory adjustments to being picked up in infancy. *PLoS One 8*(6), 65289.

Reed, E. 1996. *Encountering the World: Toward an Ecological Psychology*. Oxford: Oxford University Press.

Rochat, Ph. 2015. Self-conscious roots of human normativity. *Phenomenology and the Cognitive Sciences 14*(4), 741–753.

Rödl, S. 2015. Joint action and recursive consciousness of consciousness. *Phenomenology and the Cognitive Sciences,14*(4), 769–779.

Rubio-Fernández, P., & Geurts, B. 2013. How to pass the false-belief task before your fourth birthday. *Psychological Science* 24–1, 27–33.

Salice, A. and Satne, G. 2020, Helping others in interaction. *J Soc Philos*, 51: 608-627. https://doi.org/10.1111/josp.12377

Satne, G. 2014. Interaction and Self-Correction. Frontiers in Psychology, Volume 5 - 2014 | https://doi.org/10.3389/fpsyg.2014.00798

Satne, G. 2021a. Understanding others by doing things together: An enactive account. *Synthese* 198, 507–528.

Satne, G. 2021b. "Collective intentionality, inferentialism and the capacity for claim-making". In L. Koreň, H. B. Schmid, P. Stovall, & L. Townsend (Eds.), *Groups, Norms and Practices: Essays on Inferentialism and Collective Intentionality* (pp. 99–118). Switzerland: Springer Nature.

Satne, G. 2024. Practical knowledge and shared agency: Pluralizing the anscombean view. *Inquiry 67*(4), 1018–1045.

Satne, G., & Salice, A. 2015. Helping behaviour and joint action in young children. *Phenomenology and Mind*, n. 9–2015, 98–106.

Satne, G., & Salice, A. 2020. Shared intentionality and the cooperative evolutionary hypothesis. In A. Fiebich (Ed.), *Minimal Cooperation and Shared Agency* (pp. 71–92) New York: Springer.

Schmidt, M. F., Rakoczy, H., & Tomasello, M. 2012 Sep. Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition124*(3), 325–333. https://doi.org/10.1016/j.cognition.2012.06.004. Epub 2012 Jul 4. PMID: 22766522.

Schmidt, M. F. H., & Tomasello, M. 2012. Young children enforce social norms. *Current Directions in Psychological Science 21*(4), 232–236. https://doi.org/10.1177/0963721412448659

Schmidt, R. C., & Richardson, M. J. 2008. Dynamics of interpersonal coordination. In A. Fuchs, & V. K. Jirsa (Eds.), *Coordination: Neural, Behavioral and Social Dynamics* (pp. 281–308). Berlin: Springer. https://doi.org/10.1007/978-3-540-74479-5_14

Sperber, D., & Wilson, D. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Thompson, E. 2001. Empathy and consciousness. *Journal of Consciousness Studies 8*(5–7), 1–32.

Thompson, M. 2012. Ms "You and I." Unpublished manuscript. Retrieved from https://sites.pitt.edu/~mthompso/i+you.pdf.

Tollefsen, D. 2005. Let's pretend! Children and joint action. *Philosophy of the Social Sciences* 35–1, 75–97.

Tomasello, M. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. 2008. *Origins of Human Communication*. Cambridge: MIT Press.

Tomasello, M., & Rakoczy, H. 2003. What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language, 18*(2), 121–147. https://doi.org/10.1111/1468-0017.00217

Trevarthen, C. B. 1979. Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before Speech* (pp. 321–347). Cambridge: Cambridge University Press.

Trevarthen, C., & Hubley, P. 1978. Secondary intersubjectivity: Confidence, confiding, and acts of meaning in the first year. In J. Lock (Ed.), *Action, Gesture and Symbol* (pp. 183–229). London: Academic Press.

Warneken, F., Chen, F., & Tomasello, M. 2006. Cooperative activities in young children and chimpanzees. *Child Development* 77–3, 640–663.

Warneken, F., & Tomasello, M. 2007. Helping and cooperation at 14 months of age. *Infancy 11*(3), 271–294.

Zahavi, D. 2014. *Self and Other*. Oxford UP: Oxford.

Zahavi, D., & Satne, G. 2015. Varieties of shared intentionality: Tomasello and classical phenomenology. In J. Bell, A. Cutrofello & P. Livingston (Eds.), *Beyond the Analitic-Continental Divide: Pluralist Philosophy in the Twenty-First Century* (pp. 305–325) Routledge, London.

12 Linguistic relativity and embodiment

Fernando Martínez-Manrique

12.1 Introduction

The relation between language and thought is one of the traditional issues in analytic philosophy and cognitive science. In its early phases, analytic philosophy was committed to the ideal of a universal, formal language capable of capturing the logical structure of thought and the world. Natural languages were typically seen as messy, ambiguous, and prone to confusion, and thus of limited philosophical interest except as obstacles to be overcome through logical analysis. In a similar vein, classical cognitive science envisaged language mostly as a means to express thought. Cognitive processes took place in an independent computational medium of representations with their own language-like combinatorial and compositional properties—an idea that famously took shape in the influential language of thought hypothesis (Fodor, 1975). In this intellectual climate of emphasis on the universality of thought processes and ontology, linguistic relativity (LR)—the thesis that thought is systematically influenced in a non-trivial way by the specific language that a thinker speaks—played a marginal role, even though it was a hypothesis that had been entertained for a long time in Western thought.[1]

LR gained renewed traction by the development of studies in cross-linguistic differences and new experimental designs to assess the impact of such differences in nonlinguistic cognition (e.g., Gumperz & Levinson, 1996). One conclusion that can be gathered from these studies is that LR does not reflect a univocal view: it can be construed in different terms, as a family of

1 As Leavitt (2010) shows, the idea that language might shape thought has recurred in various forms across modern thought, long before it was assimilated as the Sapir-Whorf hypothesis. For instance, enlightenment philosophers such as Wilhelm von Humboldt argued that each language embodies a unique worldview, a "world-making" capacity that structures perception and cognition. Leavitt traces how this strand of thought influenced later developments in European philosophy, anthropology, and even psychology, suggesting that linguistic relativity is not a radical or isolated idea. Additionally, the work of Quine can be regarded as a sort of precursor of linguistic relativity developed from analytic philosophy postulates (see García Llorente, 2024).

DOI: 10.4324/9781003649359-15
This chapter has been made available under a CC BY 4.0 license.

proposals that differ in their hypotheses about how language affects thought and how to interpret the empirical evidence (Wolff & Holmes, 2011). In turn, these hypotheses will depend on assumptions and theoretical stances on how to understand and investigate cognitive processes. It is not strange, thus, that the hypothesis needs to be reevaluated with the advent of 4E cognitive science and its emphasis on embodied and situated views of cognition.

In this chapter, I examine the relation between LR and embodiment, one of the four "E".[2] Section 12.2 begins by distinguishing three possible ways to understand the relation between those sources of influence: independence, constraint, and interdependence. In Section 12.3, I examine more closely a fourth relation recently defended by Kemmerer, namely, that embodiment (aka grounded cognition) entails LR. I contend that Kemmerer's account has implications regarding the expected effects of concrete vs. abstract words, and that research on this field reveals a tension with his account. Moreover, to sustain his conclusion, it is necessary that the effects of embodiment on cognition are not trivial, in the sense of being functionally relevant for a number of tasks. I examine that question in Section 12.4 based on research on the cognitive differences between people with remarkable bodily differences. Finally, Section 12.5 returns to abstract words to examine the prospects of treating them in embodied terms and the consequences of this treatment for LR.

12.2 The relation between embodiment and LR

The relation between embodiment and LR is not a straightforward one. We can distinguish in the literature at least four ways in which the influence of language and the influence of embodiment could be related: independence, constraint, interdependence, and entailment. Let me review briefly each of them.

The first possible relation is independence. On this view, language and embodiment would exert distinct and relatively autonomous influences on cognition. Language might shape thought through structural mechanisms, such as grammar or morphology, or by lexical distinctions, whereas embodiment would exert its influence through sensorimotor and affective interactions with the environment. For example, consider those studies on grammatical gender which suggest that speakers of languages with gendered nouns (e.g., Spanish or German) ascribe stereotypical masculine or feminine attributes to inanimate objects (Boroditsky et al., 2003). If there is a real influence of

2 This is reasonable for two reasons. One is that embodiment, i.e., the idea that cognitive processes are deeply rooted in the body's sensorimotor systems and interactions with the environment, arguably plays a foundational role in the different variants of 4E (Gallagher, 2005). The second is that the lines of research about the influence on cognition from language and from embodiment converge at several points, as this chapter will show.

language on conceptualization in this domain, one may contend that it is largely independent of perceptual experience. Conversely, research on how perceptual and motor systems guide understanding and reasoning even in the absence of language (e.g., Beilock & Goldin-Meadow, 2010) would suggest that embodied processes may shape cognition in a way autonomous from language. The independence view thus suggests that the contributions of language and embodiment reflect different dimensions of overall cognitive architecture.

A second possible relation between linguistic and embodied influences is one of constraint. The idea is that embodiment may set fundamental boundaries on the possible effects of language on thought (Mondal, 2021). On this view, because cognitive processes are rooted in bodily structures and sensorimotor experience, the scope of linguistic influence is constrained by what the embodied mind can represent or process. For instance, cognitive linguistic theorists typically argue that the core architecture of conceptual systems is grounded in bodily experience, which sets a foundation for what can be linguistically encoded (Evans, 2009), implying that linguistic variation operates only within a framework of shared sensorimotor constraints. This perspective is particularly prominent in accounts that treat embodiment as universal across human populations. Actually, the constraint view could sustain the thesis of a certain universality of cognition, if sensorimotor experience tended to be roughly the same across individuals. For instance, Gallagher (2005) may be taken as support for the idea that embodied processes set fundamental limits on the cognitive influence of language. Basic interpersonal understanding and early social interaction, according to Gallagher, are grounded in nonlinguistic embodied capacities such as affective attunement and motor coordination. This suggests that language operates within boundaries established by more primary bodily processes. In this view, LR is plausible only within the range permitted by universal features of bodily experience.

A third and most dynamic relation between language and embodiment is interdependence, where linguistic and bodily experience mutually influence and co-constitute each other in shaping thought. Talking of sensorimotor experience as if it worked in an independent manner from linguistic experience is an oversimplification that does not capture the complex interactions that occur in reality. Moreover, linguistic capacities can be themselves understood as embodied, both at the level of acquisition (Reggin et al., 2023) and of language processing (Körner et al., 2023). In this respect, Casasanto (2016b) presents a compelling case for the interdependence between language, culture, and embodied experience, proposing a shared mechanism underlying their mutual influence. Rather than treating LR and embodiment as separate explanatory frameworks, or regarding one of them as the most basic, he argues that they can shape cognition through the same underlying cognitive processes—most notably, through the construction of experience-dependent mental representations and strengthening some implicit associations in

long-term memory while weakening others. For example, habitual patterns of bodily interaction with the environment (such as left- vs. right-handedness) can influence abstract reasoning and spatial metaphors in ways analogous to how linguistic structures do. This convergence suggests that language and embodiment are not parallel but interacting forces that jointly modulate thought, forming a dynamic feedback loop grounded in lived experience. So, to reassess the thesis of LR, one must take into account the many possible ways in which language may interact with perception and cognition (Casasanto, 2016a).

A different, more radical way to understand the interdependence relation comes from enactivist conceptions (e.g., Rodríguez Jordá & Di Paolo, 2025). In this view, language and diverse cognitive skills are seen as entangled processes that cannot be strictly isolated so as to ask about their respective influences. Framing the question this way, they contend, retains fundamental assumptions from classical cognitive science—such as the representational view of mind, a certain commitment to modularity, a linear view of interaction, or an excessive focus on formal properties of language over its dynamic use—that they outwardly reject. At the same time, they counter the possible universalist inclinations behind embodiment theories. While some embodied approaches have been skeptical of the importance of linguistic diversity, the enactive perspective, particularly through the concept of "linguistic bodies", suggests that human bodies are shaped by history and culture, leading to the conclusion that human diversity itself might be the universal.

Finally, there is a fourth and stronger relation between embodiment and LR: a relation of entailment from the former to the latter. I will examine this thesis more closely in a separate section because it raises issues, which, I think, can be illuminating of the problems involved in accounting for the relation.

12.3 Does embodiment entail LR?

Working from the neuroscience of language, David Kemmerer has recently argued for a stronger form of relation between embodiment and LR. In his view, what he calls the Grounded Cognition Model[3] entails LR (Kemmerer, 2023a). Let me quote his own summary of the argument:

> Premise 1: The Grounded Cognition Model (GCM) claims that the sensory and motor features of concepts, including word meanings, are stored directly within modal brain systems for perception and action. More precisely, the GCM holds that these semantic features are identical

3 Kemmerer regards "grounding" and "embodiment" as two closely related notions. Even though I acknowledge that there are differences between them, they are irrelevant for the purposes of this paper, so I will use the terms interchangeably.

> to some of the modality-specific representations that are engaged when the relevant kinds of entities and events are processed for nonlinguistic purposes, such as object recognition and action planning.
>
> Premise 2: The sensory and motor features of word meanings vary greatly across the roughly 7000 languages in the world. These extensive cross-linguistic differences are manifested in many conceptual domains, including numerous categories of entities and events.
>
> Conclusion: To accommodate this diversity, the GCM must assume that modal brain systems for perception and action are significantly shaped by the idiosyncratic semantic details of particular languages. This in turn entails linguistic relativity—that is, the view that language specific concepts influence other mental processes.
>
> (Kemmerer, 2023b, 699)

Kemmerer's two premises constitute embodiment versions of the two typical requirements for LR: linguistic determinism (i.e., the idea that language exerts a significant influence on cognition) and linguistic diversity (i.e., the idea that there are significant differences between languages in aspects that are relevant for linguistic influence).[4] Premise 1 is a claim about how embodiment determines cognition based on the idea of a direct impact in brain representational systems. Contrary to the idea that sensory-motor features are too raw to be part of the semantics of cognition—i.e., that they need to undergo a process of refinement or abstraction to figure in conceptual structures—Kemmerer contends that mental semantics retains the modalities experienced in constituting the relevant contents. This thesis is not enough by itself to support LR, as it could be the case that the sensory and motor features conveyed by the different languages were roughly the same. Premise 2 attacks this universalist view by contending that languages exhibit large differences with respect to the relevant features that they are able to convey.

How strong is the entailment argument? I contend that the thesis that grounded cognition entails LR depends on answers to two different questions: one has to do with the strength of premise 2, i.e., the degree to which linguistic differences involve differences in embodiment; the other has to do with the depth of premise 1, i.e., the degree to which differences in embodiment are functionally relevant for cognitive processes.

Let me begin with the strength of premise 2, which contends that there are linguistic differences that involve differences in embodiment. As Kemmerer points out, there are manifest differences in how word meanings partition a

4 Neither requirement is enough by itself: we could have linguistic determinism without linguistic relativity because it could be the case that language determined thought but that all natural languages were superficial variations of a Chomskyan universal language; and we could have linguistic diversity without linguistic relativity if natural languages turned out to be mere systems of expression of a preexistent language of thought.

number of perceptually and motor-related domains. To make his point, he reviews cross-linguistic findings that reveal a diversity of ways of partitioning reality in domains such as color, household containers, nominal classification systems, motion events, and events of cutting, breaking, and opening. According to his conclusions, these differences involve different ways of categorizing and of creating equivalences that otherwise may not exist. Notice that these are some of the domains that have been the focus for some time in LR research. What is relatively new in Kemmerer's view has to do with the more direct way in which these sensory-motor differences would be reflected in mental organization, given that, as premise 1 contends, they are stored directly within modal brain systems. Kemmerer backs this conclusion resorting to differences between event-related potentials in speakers of languages that exhibit some relevant difference. For instance, Event-Related Potential studies have shown that Greek speakers (whose language distinguishes between dark and light blue) exhibit different brain responses to shades of blue compared to English speakers (whose language has one term for blue), even during nonlinguistic color tasks.

Now, a corollary of Kemmerer's argument would be that we should expect that (i) those words that are more closely related to sensorimotor experience should exhibit greater relativistic effects. In other words, if language L and L* differ in the way their respective word meanings capture a certain aspect of sensorimotor experience for the domain of, say, motion, then one could predict that speakers of L should differ from speakers of L* in how they nonlinguistically relate to motion, e.g., in recognizing or remembering motion events. Conversely, (ii) one should expect weaker differences for those words that capture aspects of experience that are far removed from sensorimotor experience. In a first approximation, (i) would include typically more concrete words while (ii) would include more abstract words. The reason is that concrete words refer to objects, events, or experiences that can be perceived directly by the senses, while abstract words refer to entities, qualities, or events that cannot be so directly perceived. So, if Kemmerer is right, we should expect stronger relativistic effects for concrete words, and milder effects for abstract ones.

However, this expectation does not seem to be confirmed by evidence. In a previous review of the literature in LR with respect to the concrete vs. abstract contrast, Borghi (2019) argues that the influence of language on thought is higher in the domain of abstract words than in the domain of concrete ones. Indeed, some of the domains she reviews are the same as those considered by Kemmerer, e.g., containers, breaking/cutting events, and motion verbs. Even though linguistic variation has an impact on these domains, it is not such a large one when compared to more abstract domains, such as numbers, emotion, or time. To put but one example, she reviews a study by Majid et al. (2015) comparing naming in four domains (color, body parts, containers, spatial relations) in 20 Germanic languages. This study found that color, body parts, and containers are similar across languages, while spatial

relations show the most variation. Borghi interprets this as supporting the hypothesis that, since spatial relations are generally considered more abstract than the other domains, they show more variation across languages. In a nutshell, Borghi's review reveals a pattern of results that is difficult to accommodate within Kemmerer's broad argument.

One may wonder whether the different reviews are due to the fact that each author is putting the focus on different aspects of the evidence they examine. Let us consider Borghi's explanation for the differences she finds between concrete and abstract words. In concrete words, she says, "even if language plays a role, it is mostly confined to linguistic tasks, because the structure of the environment has an important influence and puts many constraints on how categories are formed" (2019, 441). Abstract words, on the other hand, would have an influence beyond linguistic tasks. However, this cannot be the source of the difference between her review and Kemmerer's. It is not the case that Kemmerer merely focuses on linguistic tasks, given that he contends that language-specific conceptual representations are also activated during the nonlinguistic processing of the relevant objects and events (e.g., during perception or action planning). There is a real tension between the two accounts regarding their interpretation of the evidence.

Still, one may dismiss the difference between the respective reviews by noticing that their conclusions are not exactly incompatible. So, one could say that Borghi is not claiming that differences in concrete words do not have an impact—her claim is only that the impact is smaller than for abstract words. Indeed, the fact that she claims to agree with Kemmerer's arguments (Borghi & Mazzuca, 2023), and that he comments on his work approvingly (Kemmerer, 2023b), could be taken in support of the idea that their views are closely associated. Nevertheless, I contend that the tension between their interpretations persists: if Kemmerer is right, LR should be stronger for concrete words, which are those that have stronger grounding. I will return below to the concrete/abstract distinction and its relevance for the relation between LR and embodiment. Before doing so, I want to examine another potential source of tension for Kemmerer's entailment approach, related to the question of the depth of premise 1. This has to do with the functional relevance of the embodiment differences that one may encounter in languages.

12.4 The functional relevance of differences in embodiment

One of the problems that has lurked in LR since its revival is whether the putative effects of language on cognition are not trivial (Casasanto, 2016a). "Non-trivial" can be understood in different ways, most typically that linguistic variables can have an effect in nonlinguistic tasks, and that these effects are neither shallow nor transient. Now, similar worries have been raised in concern with the embodied cognition paradigm. Indeed, there is still debate about the scope of embodied cognition explanations, with some authors outwardly regarding them as vague, trivial, or even nonsensical

(Goldinger et al., 2016). I won't pay attention in this chapter to such general criticisms. As Shapiro (2019) contends, embodied cognition is better considered a research program than a well-defined theory; hence, one might anticipate inconsistencies between the different versions of embodiment. This translates to the hypothesis of bodily relativity, which can be formulated in a way that runs parallel to the classical debate on LR. Just as LR means that differences between natural languages lead to differences in cognition, body relativity would mean that differences between bodily experiences lead to differences in cognition. So, a way to assess the hypothesis would be to compare individuals with significant differences in their embodied processes (how they experience and interact with the world physically) to see if these differences correlate with variations in cognitive processes that are believed to rely on those physical experiences.

As Shapiro (2011) contends, the questions for bodily relativity[5] would be roughly the same as for LR: do bodily experiences exert a non-trivial influence in cognition? Do different bodily experiences lead to differences in cognition? One would need positive answers to these questions if one wanted to endorse Kemmerer's conclusion of embodiment entailing LR. It is easy to see that the truth of the premises in his argument—(1) that sensory-motor features are directly stored in the semantics of brains, (2) that there is linguistic diversity regarding those features—does not warrant the conclusion—(3) that language-specific concepts influence other mental processes—unless one can establish that the influence of sensory-motor representations is not trivial.

A way to address this question is to investigate whether embodied processes are functionally relevant for different cognitive processes. This is the aim of Ostarek and Bottini (2021). Focusing on cognition-as-simulation accounts, they consider three alternative hypotheses about the role of sensorimotor processes in high-level cognition: "simulations may be strictly necessary and functionally relevant; they may not be strictly necessary but have functional relevance when they are in place; or they might be neither necessary nor functionally relevant" (2020, 4). To provide a preliminary assessment of these alternatives, they review three different kinds of studies: congenital sensory-motor disorders, acquired sensory-motor deficits, and interference paradigms with healthy participants. An interesting consequence for the purposes of this chapter is that, at the same time, they also provide a preliminary assessment of the thesis of bodily relativity. The rationale for this is that the groups of subjects they examine provide sources of bodily diversity, which is the premise analogous to linguistic diversity needed by the thesis of bodily relativity. In other words, if we found systematic differences in the cognitive processes of people who exhibit great variations in bodily

5 Shapiro refers to the thesis as linguistic "determinism" rather than "relativity", but the distinction does not have any import in this discussion. What he says can be extended to linguistic/bodily relativity in general.

experience, we would have some basis to hold the thesis that cognition is relative to embodiment. Let me summarize the three kinds of groups examined by Ostarek and Bottini.

First, we have studies on congenital sensory-motor disorders, such as individuals born without limbs or without sight. The idea is to assess whether typical conceptual cognition can develop without corresponding sensorimotor experiences. Findings show that such individuals can still acquire and use concepts related to actions or visual phenomena and that typical sensorimotor brain regions are often recruited even in their absence, suggesting that embodied simulations are not strictly necessary for conceptual thought. However, subtle differences in conceptual richness or structure indicate that sensorimotor experience can enhance or modulate cognition, supporting a view in which embodiment plays a functionally relevant—but not indispensable—role in high-level thought.

Second, acquired sensory-motor deficits are cases in which individuals lose sensory or motor functions due to brain damage or disease. Examining their performance on tasks requiring understanding of action or perception-related concepts, some patients show deficits in processing concepts tied to their damaged modality while others perform normally, suggesting compensatory mechanisms or partial redundancy in conceptual systems. Overall, the findings provide mixed evidence but point toward a moderate role of sensorimotor systems in supporting conceptual knowledge: they contribute functionally when intact but are not strictly necessary, as cognition can be preserved through alternative neural pathways.

Finally, interference paradigms with healthy participants test the functional relevance of sensorimotor processes by temporarily suppressing them during conceptual tasks. For instance, applying transcranial magnetic stimulation to the motor cortex can selectively disrupt the processing of action-related words, and visual interference can impair the recognition of visual features like shape or color. While such studies often show that disrupting sensorimotor systems leads to modest impairments, the effects are not always consistent or strong, suggesting that while sensorimotor processes can enhance conceptual cognition, they are not strictly necessary.

Considering the three types of findings, Ostarek and Bottini regard them as lending support to a moderate causal influence of embodied processes in cognition. They tentatively rule out the alternative that simulations are necessary for cognition, as well as the alternative that they are not even functionally relevant, concluding that simulations are functionally relevant but not indispensable for high-level cognition.[6]

The upshot for Kemmerer's argument would be thus: if language shapes, as he claims, the embodied, modal systems for perception and action, but

6 There are other studies that study different parameters of bodily diversity, such as Casasanto's findings of differences between left-handed and right-handed individuals (Casasanto, 2016b).

the functional relevance of the latter for higher cognition is moderate and non-mandatory, the consequence is a modest version of LR. Now, we would have a way of partially reconciling his view with Borghi's review of the evidence. The lesser impact of language in thought in the domain of concrete words would be simply a reflection of the fact that sensorimotor experiences are capable of exerting a functional influence on cognitive tasks, but not in a deep or generalized manner.

Still, the second part of the tension in Kemmerer's and Borghi's accounts remains unresolved: why is it that abstract words apparently exhibit stronger relativistic effects? I turn now to this question to examine its implications for the relation between embodiment and LR.

12.5 Abstract words and embodiment

As I said above, Borghi (2019) concluded that there are stronger relativistic effects for abstract words than for concrete ones. She points out herself that her review is not exhaustive and there is always room for newer results, inviting further reinterpretations. Still, it is worthwhile to take her conclusion at least as a plausible hypothesis and reflect about what this would involve for the relation between LR and embodiment. Using as a framework the types of relations that I distinguished in Section 12.2, one could be inclined to compose the following scenario: sensorimotor environmentally grounded experiences act as constraints for conceptualizations of concrete domains, with little room for linguistic divergences, while the influence of language appears as largely independent from such bodily constraints when it is directed to abstract domains. This scenario would receive further support from evidence showing that abstract and concrete words are represented by distinct systems in the mind (Borghi et al., 2018; Desai et al., 2018). Some contents enter the mind via bodily experiences, others enter through language, and they result in different kinds of representations. The upshot is a pluralistic view of cognition, defended among others by Dove (2009, 2024), or Kompa (2021).

However attractive this division of roles may appear, I think that it still misses the complexity of the own notion of embodiment. The fact that the meanings of abstract words are less embodied in sensorimotor experience does not necessarily mean that they are not embodied at all. As Borghi characterizes abstract concepts, they activate fewer exteroceptive and more interoceptive modalities than concrete ones. In her view, abstract concepts are uniquely characterized by two aspects. First is their relational nature, meaning they are defined by their connections to other concepts. Second, they exhibit heterogeneity: the types of properties characterizing their members are more diverse, and context plays a significant role in linking them. This entails that characterizing abstract concepts in terms of embodiment is a more complex task, but not an impossible one.

In this respect, Borghi et al. (2017) review eight theoretical frameworks that posit different sources of embodiment for abstract concepts. The theories differ in the level of embodiment, with the strongest relying only on a sensorimotor basis for abstract concepts, and the weakest suggesting a mix of representations. One theoretical framework that advocates a thorough embodiment of abstract concepts is cognitive linguistics. For instance, drawing on their work on conceptual metaphors, Johnson and Lakoff (2002) contend that these metaphorical constructions play a fundamental role in constituting the ontology of our abstract concepts. Similarly, in a recent overview of embodiment and language from a cognitive linguistic perspective, Pelkey (2023) contends that the use of metaphor in abstract reasoning demonstrates how language reconfigures bodily experience into new conceptual domains. Rather than simply reflecting embodiment, language re-channels and scaffolds it, supporting a model of mutual influence between linguistic structures and sensorimotor systems.

Of course, the whole idea depends on the controversial thesis that our conceptual system is structured by metaphors through and through. There are alternative frameworks that, while regarding abstract concepts anchored in grounded or embodied experience, contemplate concept construction and deployment as a more opportunistic ad hoc understanding. Barsalou and Wiemer-Hastings (2005) aimed to demonstrate that abstract concepts share what they call "situational content" with concrete concepts. The differences, they argued, lie in the situational forms and the greater complexity of abstract concepts, which are much more distributed across brain regions than concrete ones. The main assertion, in any case, is that abstract concepts could also be simulated by recreating the complex situation that grounds them. Borghi's own theoretical stance—the Words As Social Tools theory—regards abstract concepts as based on sensorimotor, linguistic, emotional, and social information. In other words, this study of abstract words calls for a revision of how the effects of embodiment take place, not for an outright dismissal of their embodied nature.[7]

Indeed, the concrete/abstract dichotomy is under revision as well. The consensus view is to abandon it and acknowledge the diversity and complexity of abstract concepts that cannot be possibly captured by a single theoretical account. The upshot is that "the heterogeneity of results does not offer a clear-cut answer to the universalist-relativist debate" regarding cross-linguistic differences in abstract concepts (Banks et al., 2023, 9). For some domains, like actions and colors, biological and physical factors play a prominent role, whereas for others, such as emotions and gender, linguistically shaped sociocultural factors turn out to be more relevant.

7 For an analysis (in Spanish) of abstract concepts regarding the problem of their grounding, see Martínez-Manrique (2020).

12.6 Conclusion

The link between embodiment and LR is not a straightforward one. There are different kinds of embodiment and different sorts of causal influences that language may exert. I think it is safe to conclude that the interactions between them cannot be captured by a general statement such as Kemmerer's "grounded cognition entails linguistic relativity". Even though I am inclined to regard interdependence as the most plausible model for the relation between linguistic and bodily influences, it is beyond the goals of this paper to try to adjudicate between the different ways to envisage this relation. Indeed, the picture could even be more complicated if one takes into account the possibility that the patterns of influence differ from domain to domain. So, there could be domains in which language and embodiment acted in independent ways, others where embodiment set a universal constraint, and finally others where the relation is much more interactive. What current research reveals is the need to address the complex interactions between the multiple factors involved in shaping cognitive processes.

Acknowledgments

This work is supported by research projects PID2023–152118NB-I00 & PID2021–126826NA-I00 of the Ministerio de Ciencia e Innovación, Spain. The author wishes to thank the audiences at the workshops on Relatividad Lingüística y Filosofía Experimental (Universidad Complutense de Madrid, January 2024) and Analytic Philosophy and E-Cognition (Universidad de Granada, June 2024) where versions of this paper were presented. Special thanks to Patricia Hernando, Carlota García Llorente and Manuel Heras Escribano.

References

Banks, B., Borghi, A. M., Fargier, R., Fini, C., Jonauskaite, D., Mazzuca, C., Montalti, M., Villani, C., & Woodin, G. (2023). Consensus paper: Current perspectives on abstract concepts and future research directions. *Journal of Cognition*, 6(1), 62, 1–26.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher, & R. A. Zwaan, (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 129–163). Cambridge: Cambridge University Press.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, 21(11), 1605–1610.

Borghi, A. M. (2019). Linguistic relativity and abstract words. *Paradigmi*, XXXVII, 429–448.

Borghi, A. M., & Mazzuca, C. (2023). Grounded cognition, linguistic relativity, and abstract concepts. *Topics in Cognitive Science,* 15, 662–667.

Borghi, A. M., Barca, L., Binkofski, F., & Tummolini, L. (2018). Varieties of abstract concepts: Development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170121.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, 143(3), 263–292.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–79). Cambridge, MA: MIT Press.

Casasanto, D. (2016a). Linguistic relativity. In N. Riemer (Ed.), *Routledge handbook of semantics* (pp. 158–174). New York: Routledge.

Casasanto, D. (2016b). A shared mechanism of linguistic, cultural, and bodily relativity. *Language Learning*, 66(3), 714–730.

Desai, R. H., Reilly, M., & van Dam, W. (2018). The multifaceted abstract brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170122.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, 110(3), 412–431.

Dove, G. (2024). Embodied cognition needs an upgrade. In L. Shapiro, & S. Spaulding (Eds.), *The Routledge handbook of embodied cognition* (pp. 485–494). London: Routledge.

Evans, V. (2009). *How words mean: Lexical concepts, cognitive models, and meaning construction*. Oxford: Oxford University Press.

Fodor, J. (1975). *The language of thought*. New York: Crowell.

Gallagher, S. (2005). *How the body shapes the mind*. London: Clarendon Press.

García Llorente, C. (2024). The ontology of natural language(s) and linguistic relativity: A deflationary approach. *Forum Philosophicum*, 29(2), 293–315.

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, 23(4), 959–978.

Gumperz, J. J., & Levinson, S. C. (Eds.) (1996). *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

Johnson, M., & Lakoff, G. (2002). Why cognitive linguistics requires embodied realism. *Cognitive Linguistics*, 13(3), 245–263.

Kemmerer, D. (2023a). Grounded cognition entails linguistic relativity: A neglected implication of a major semantic theory. *Topics in Cognitive Science,* 15, 615–647.

Kemmerer, D. (2023b). Grounded cognition entails linguistic relativity: Response to commentators. *Topics in Cognitive Science,* 15, 698–708.

Kompa, N. A. (2021). Language and embodiment—Or the cognitive benefits of abstract representations. *Mind & Language*, 36(1), 27–47.

Körner, A., Castillo, M., Drijvers, L., Fischer, M. H., Günther, F., Marelli, M., Platonova, O., Rinaldi, L., Shaki, S., Trujillo, J. P., Tsaregorodtseva, O., & Glenberg, A. M. (2023). Embodied processing at six linguistic granularity levels: A consensus paper. *Journal of Cognition*, 6(1), 60, 1–28.

Leavitt, J. (2010). *Linguistic relativities: Language diversity and modern thought*. Cambridge: Cambridge University Press.

Majid, A., Jordan, F., & Dunn, M. (2015). Semantic systems in closely related languages. *Language Sciences*, 49, 1–18.

Martínez-Manrique, F. (2020). Conceptos abstractos, lenguaje e hibridación. *Análisis. Revista de investigación filosófica*, 7(2), 239–261.

Mondal, P. (2021). The limits of language-thought influences can be set by the constraints of embodiment. *Frontiers in Psychology*, 12, 593137.

Ostarek, M., & Bottini, R. (2021). Towards strong inference in research on embodiment – Possibilities and limitations of causal paradigms. *Journal of Cognition*, 4(1), 5, 1–21.

Pelkey, J. (2023). Embodiment and language. *Wiley Interdisciplinary Reviews: Cognitive Science*, 14(5), e1649.

Reggin, L. D., Franco, L. E. G., Horchak, O. V., Labrecque, D., Lana, N., Rio, L., & Vigliocco, G. (2023). Consensus paper: Situated and embodied language acquisition. *Journal of Cognition*, 6(1), 63, 1–16.

Rodríguez Jordá, U., & Di Paolo, E. A. (2025). Linguistic relativity from an enactive perspective: The entanglement of language and cognition. *Language Sciences*, 108, 101702.

Shapiro, L. (2011). Embodied cognition: Lessons from linguistic determinism. *Philosophical Topics*, 39(1), 121–140.

Shapiro, L. (2019). *Embodied cognition*. London: Routledge.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science,* 2(3), 53–65.

Part 3

Ethical and political implications

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

13 Situated agency

A Wittgensteinian exploration

Annemarie Kalis

13.1 Introduction

The fact that we can understand ourselves as free agents, in other words, as individuals who make decisions that are genuinely up to us, is an important aspect of the human condition. This seems true even if we acknowledge that this is certainly not the only way in which we understand ourselves: seeing oneself as a victim of fate, or of the circumstances, is just as familiar to most of us. Nevertheless, the notions of human freedom and agency have received a lot of attention in philosophical thought, with the core questions being why this self-understanding is so important for us, and to what extent it reflects a metaphysical fact about human beings.

These questions are particularly challenging for recent situated approaches to the human mind. After all, situated approaches understand the human mind to be fundamentally embodied, embedded, extended, and enacted (a set of assumptions taken together as 4E approaches to cognition (Gallagher, 2009; Newen et al., 2018). On these assumptions, the idea that individuals make decisions that are free in the sense of genuinely being up to them does not naturally arise. Does this mean that from a situated point of view, the notions of freedom and agency should be understood as illusory or at least somewhat misleading? Or should situated approaches to the mind just understand these notions differently? In this chapter, I will argue for the latter. To provide some background, I will start by providing a brief overview of existing situated approaches to agency.

In a very general sense, agency indicates the capacity of individuals to shape their environments: to do things in the world in order to realize their goals. Situated approaches to the mind have developed accounts of agency that start from the assumption that individuals and their environments are deeply interdependent (Bevir, 2017; Di Paolo et al., 2018; Segundo-Ortin, 2020). A key feature of many discussions surrounding agency in the situated cognition literature is their focus on a predominantly scientific and biological perspective. For example, enactivist approaches ascribe agency to organisms insofar as they act within their environment in order to achieve goals, with survival being the most fundamental guiding aim. Barandiaran et al. (2009)

DOI: 10.4324/9781003649359-17
This chapter has been made available under a CC BY 4.0 license.

highlight three central aspects of agency: "(a) there is a system as a distinguishable entity that is different from its environment, (b) this system is doing something by itself in that environment, and (c) it does so according to a certain goal or norm" (p. 369). Notably, this definition of agency extends to all organisms, including even the simplest forms of life, such as bacteria. Similarly, Maiese (2022) sees autonomy as the integration of an "agentic entity" that acts and perceives and the norms that the entity establishes for itself.

In these two enactive accounts, norms are understood primarily in relation to the overarching goal of self-maintenance. As Mojica (2021) explains, "[n]ormative behavior is established by living organisms whose autonomous nature implies that they can differentiate events and regulate their responses in terms of what is better or worse for the maintenance of their own precarious identity". This norm of self-maintenance shapes how organisms move through and interact with their environment. Importantly, an agent is always situated in and interacting with its environment, and enactivists see agency as the modulation of this interaction by the organism: "agency is not something that occurs within the agent. It is, rather, something that emerges in the interaction between the agent and its environment" (McGann, 2014, p.219).

Ecological psychology similarly views agency as fundamentally interactional. Segundo-Ortin (2020) conceptualizes agency as a property of the relation between the organism and its environment. However, unlike enactivist accounts that emphasize norms, his perspective highlights the role of perceptual information that organisms exploit to guide their actions. Segundo-Ortin and Kalis (2024) have also extended this framework to explore distinctively human forms of agency: their work examines how shared social practices and behavior settings influence and enhance individual agency.

What these situated approaches do not address are more traditionally philosophical or existential questions such as: what does it mean to be not just an agent but an agent that is *free*? Do we have reason to think we are to some extent really in control of our actions, rather than merely subject to external events? Addressing these questions requires a focus on the first-person perspective on agency. Within the field of enactivism, several philosophers have focused on the first-personal perspective in their work on the phenomenology of agency (Buhrmann & Di Paolo, 2017; Gallagher, 2012). They analyze the phenomenological aspects of volitional control and try to relate them to cognitive and neuroscientific accounts (for example, sensorimotor organization). However, such cognitive-phenomenological accounts primarily focus on the specific experience of being the source or initiating cause of a concrete action (such as a button-press), as studied in the lab. So even though recent work on the "sense of agency" provides important insights into what it means to understand ourselves as agents, there are still important questions to answer. Most notably, what is missing is an understanding of a more general form of self-understanding that is not tied to a concrete action or moment of decision but is a feature of our everyday life. My proposal is that Wittgenstein's thoughts about freedom of the will can shed light on this form

of self-understanding. They show how freedom could be a socially mediated first-personal feature in a way that aligns well with the central tenets of situated approaches to the mind. Therefore, I take his work to have the potential to contribute to a situated understanding of agency.

Even though Wittgenstein is not frequently brought up in debates on situated agency, his work has been acknowledged as an important resource in the sociological debate on structure versus agency (King, 2009; Pleasants, 1997; Plotica, 2013). This debate concerned the fundamental question of how the sociological assumption that human action should be explained in terms of social structures could be squared with some notion of individual agency. Here, Wittgenstein's work turned out to be a valuable resource for several authors, as Wittgenstein shared the sociologists' starting point that the primary locus of human action is the social practice. For example, Pleasants (1997) argued that

> The idea of 'freedom' and 'agency' as inherent possessions of individuals in abstraction from any particular context is utterly unintelligible to a Wittgensteinian. Freedom can only be 'freedom from' something, or 'freedom to do' something. These 'freedoms' vary historically, geographically, and across classes of people.
>
> (p.24)

But nevertheless, Wittgenstein leaves room for individuality within a social practice. Plotica quotes McGinn in stating that

> 'insofar as [Wittgenstein] has a view on the individual/social opposition, he is an individualist'. This is not to discount the social dimensions of language use and agency, but to take a perspective on them that prioritizes the individual as a linguistic agent.
>
> (Plotica, 2013,p.64)

I want to take inspiration from the way Wittgenstein's ideas have been used in sociology to develop a situated understanding of agency and explore to what extent his work might also provide a situated account of what it means for human beings to understand themselves as free agents. In the next section, I will provide an analysis of Wittgenstein's most important writings on freedom of the will, the *Whewell Court Lectures*. Subsequently, I will show (in Section 13.3) how these remarks suggest a view on freedom as a stance that human beings can, but need not always, take—a stance they take for practical purposes. In the conclusion, I will connect this Wittgensteinian perspective to contemporary situated thinking about the mind.

13.2 Wittgenstein on freedom of the will

Wittgenstein did not write extensively about freedom of the will, but the topic comes up at several places in his work: in the *Tractatus Logico-Philosophicus*

(1922) (more specifically, in 5.1362) and in several lectures that he gave in Cambridge around 1940 (Wittgenstein, edition 2017). The most systematic and focused analysis of the notion is found in the *Lectures on Freedom of the Will*, which will be the core resource referred to in this chapter. However, I will connect several points made in these lectures to relevant passages in other texts from his oeuvre.

In the lectures, Wittgenstein (in his characteristic style) brings up a wide variety of examples, ideas, questions, and comparisons. One specific passage is particularly illuminating to show Wittgenstein's thoughts on what it means to understand oneself as a free agent. I will quote it at length, dividing it into several parts. In the first part, Wittgenstein presents the following example:

> Suppose I were about to do something of great consequence to myself and to someone else. I may get a very strong sense of what I may call 'freedom of the will'. I may say: 'I can't say that I am forced to do this, or not to do it. I choose freely to do it, if I do it.' And I could also imagine saying to myself: 'I am not free. What can I do? I haven't chosen these circumstances. Why should I do this? No one would. I am not a hero.'
>
> (2017, p.291)

Here, Wittgenstein describes how there are two possible ways in which someone could understand one's own situation. One *could* understand oneself as a free agent, but one *could* also understand oneself as a victim of the circumstances, or fate. This shows that whereas for Wittgenstein the subjective experience of freedom of the will can be present in human beings, it need not be—at least not always. Wittgenstein continues with the following questions:

> In this case, what actually am I saying to myself? Am I saying something about scientific law, or about what will probably be found when they discover more about the human mind?
>
> (2017, p.291)

The latter question is clearly rhetorical. Here, Wittgenstein is pointing out that understanding oneself as being either free or not should not be interpreted as the taking of a scientific position, or as a belief about empirical facts about the mind (the irrelevance of the sciences of the mind for human experience and self-understanding is a recurring theme in his later work [Williams, 1985]). Of course, this doesn't yet tell us how we should interpret the subjective experience of freedom. The third part of the passage sheds some light on this thanks to a question from his student Lewy:

> *Lewy*: Is the feeling of being free a sufficient ground for saying you are free?

> *Wittgenstein*: I don't know what feeling you are talking about. Instead of these words 'He had a feeling' I might just as well say, 'He had the thoughts.'
>
> *Lewy*: Suppose I ask: what are the grounds for this conviction of being free?
>
> *Wittgenstein*: I might say: there are no grounds. And as for feelings, you can choose whatever you consider most interesting.
>
> (2017, p.291)

This exchange develops at least two interesting ideas. First, whereas understanding oneself as free supposedly is something we can be aware of, it cannot be analyzed as a specific kind of mental occurrence, such as a feeling or a thought. He even suggests it is up to you how you characterize the experience of freedom: "whatever you consider most interesting". And second, understanding oneself as free is not grounded in any other conviction or thought.

As I will now show, taking these insights together leads to the idea that for Wittgenstein, freedom of the will is *a stance human beings can take*. Wittgenstein's suggestion is that freedom should not be understood as a metaphysical feature, but as a way to understand ourselves which is a possible mode of self-understanding for human beings, but not a necessary one. Importantly, this stance is not grounded in anything else, such as an assumption of rationality or scientific knowledge. The latter idea (that understanding oneself as free is not grounded in scientific knowledge) is developed more explicitly when Wittgenstein discusses the relation between freedom of the will and the laws of nature. His main point in these passages is that freedom is unrelated to regularity:

> There is no reason why, even if there was regularity in human decisions, I should not be free. There is nothing about regularity which makes anything free or not free. The notion of compulsion is there if you think of the regularity as compelled; as produced by rails
>
> (2017, p.284)

Here, Wittgenstein's core insight is that *regularity* is not the same as *compulsion* and that only compulsion stands in opposition to freedom. Natural laws do certainly constrain what human beings can do, but they do not *compel* behavior:

> What on earth would it mean ⟨to say⟩ that the natural law compels a thing to go as it goes? The natural law is correct, and that's all
>
> (2017, p.283)

This is an understanding of the laws of nature that can also be found in the writings of some of Wittgenstein's contemporaries, such as Elizabeth

Anscombe (Anscombe, 1971) and Gilbert Ryle (Ryle, 1954). They all rallied against the outlook that the laws of nature are like ropes that prevent us from doing what we would really like to do:

> there is a certain outlook: 'We are all the time being determined. We think we decide, but all the time we are being shoved about, our decisions too. This means that we are misled into thinking that we do what we want.'
>
> (2017, p.287)

Instead, they argued, the laws of nature are not rules but mere regularities that we know as perfectly ordinary features of our daily lives. Not being able to change the course of the atoms from which we are built constrains us only in the sense that gravity constrains us from diving to the moon. These types of constraints have nothing to do with the way in which we can understand ourselves as free in the everyday decisions we make: "Normally, unless we philosophize, we don't talk this way. We talk of making decisions" (2017, p.287). So, taking the stance of freedom with regard to one's actions and decisions is not threatened by the laws of nature nor by a scientific position like causal determinism. However, Wittgenstein qualifies this claim in an important way by pointing out that sometimes, science in fact does change the way we understand ourselves:

> My point was that these statements were not scientific statements, not corrected by experience. […] This is not to say that scientific discoveries have no influence on statements of this sort. […] A discovery might influence what you say on the freedom of the will, if only by directing your attention in a particular way.
>
> (2017, pp.293–294)

Here, Wittgenstein describes how scientific developments can change the way we look at the world and ourselves, even if these changes do not necessarily follow from scientific knowledge itself. Scientific ideas can direct our attention in a way that develops into a novel framing of our general experience and outlook. As an example, he discusses the impact of economic models:

> If your attention is drawn for the first time to the fact that economic states of affairs have enormous and obvious consequences, whereas such things as general states of mind of people do not […] it is very natural to think that *all* explanations can and should be given a form like economic explanation of historic states of affairs
>
> (2017, p.294)

Similarly, scientific insights in the movements of atoms and the laws of nature can make it more likely that human beings start to look at others

and themselves as being at the mercy of the atoms clashing inside them. But, as Wittgenstein emphasizes, this is not a change in scientifically grounded beliefs, but a change in attention. The way Wittgenstein here uses the notion of attention opens up a fascinating perspective according to which understanding ourselves as either free or as a victim of circumstance are two different *ways of seeing*, none of which in itself represents "the correct way" to understand oneself. In the next section, we will explore this idea in more detail, emphasizing that the stance of freedom is a practical stance.

13.3 Freedom as a practical stance

A few pages after Wittgenstein provides the example of someone who takes a fatalistic stance to themselves by saying "I am not a hero", he makes some interesting remarks about the status of such a self-ascription:

> He says, 'I am not a hero', as he might say, 'This is a cake. How could it be anything else?' Where is this comparison taken from? What sort of analogy is he making? How does he know he is not a hero? Because he has always acted in this way? In the case of the hero, there is nothing analogous to the case of the cake. Why are you making a point of this analogy at all?'
>
> One thing is: not to be made responsible.
>
> Another might be: a particular attitude of seeing what is tragic in a human being. You may be driving at this if you say: 'What do you want? That is how he is made.'
>
> Among other things, saying this rules out certain expectations.
>
> (p.293)

What is the idea Wittgenstein is developing here? First, he points out that at least sometimes we make self-ascriptions as if we are describing *obvious facts* about ourselves. He then asks the question: why do we self-ascribe in this way? And his answer is that this has practical purposes: we do it in order "not to be made responsible", or to express "a particular attitude of seeing what is tragic in a human being" and "to rule out certain expectations". All these phrases indicate that even though these ascriptions seem factual, we actually make them for practical reasons that have to do with the way we relate to one another within a social practice. We understand one another as free or determined only insofar as doing so makes a difference to our everyday lives (Stokhof, 2022). This shows that, according to Wittgenstein, we are actually hardly ever interested in freedom *as such*. As described in the introduction, this point was also taken up in the debate on structure versus agency in sociology (Pleasants, 1997): for Wittgenstein, freedom is always *freedom from or to* something.

Now, how should we understand this practical stance? At the end of the previous section, we discussed Wittgenstein's suggestion that understanding

oneself as a free agent is a way of directing one's attention. More specifically, understanding oneself as a free agent involves focusing on particular features of the world and of our own experience that suggests that our decisions and actions are up to us. Examples of such "suggestive features" might be the language that we use to describe our decisions and actions (phrases like "I did this because I felt like it", or "I could easily stop smoking if I set my mind to it"), or legal practices of holding citizens criminally responsible in a way that presupposes their freedom (for discussion on whether legal practices presuppose a metaphysical notion of free will, see (Greene & Cohen, 2004; Kolber, 2015). In principle, scientific developments might also guide our attention in this direction: for example, both complexity theory and the principle of indeterminacy as discussed in quantum mechanics seem to have had some such effect (Conway & Kochen, 2009; Wolfram, 1997).

Wittgenstein's point is that we should not understand such effects as a change of beliefs grounded in scientific evidence or other kinds of facts, but as a change in attention or in "the way we see things". In order to better make sense of this point, I want to propose that it might be useful to connect Wittgenstein's remarks here with his work on *aspect perception*, which he developed in several of his later writings (Wittgenstein, 1953, 1982). There is a large body of secondary literature on this work (Baz, 2000, 2020; Budd, 1987; Day & Krebs, 2010; Eldridge, 2010; Glock, 2016; Hagberg, 2010; Melzer, 2002; Mulhall, 1990, 2001) that I will not even try to do justice here. The interpretation I will build on here is Wittgenstein's idea that human beings sometimes see things "under an aspect", or see-as-something. One of his examples is the famous duck-rabbit figure (Wittgenstein, 1953, pp.194–199): we can see the figure either as a duck or as a rabbit, but neither of them is obviously the "correct" way to see it. This phenomenon not only tells us something about how human perception works: it also points to an important connection between perception and thought. As Budd (1987) states, aspect perception shows us "the juncture of the sensory and the intellectual" (p.2).

So, how is aspect perception related to the way we understand ourselves? Several authors working in the Wittgensteinian tradition have suggested that seeing aspects plays a fundamental role in human forms of life (Baz 2000; Stokhof, 2022). Stokhof even explicitly argues that aspect perception plays a crucial role in understanding ourselves as free: "the ability of aspect seeing reflects our awareness of the intrinsic plurality of our engagement with the world, and aspect change is the execution of the inherent freedom that comes with that pluralism. Things can be seen in different ways, and we are able to make use of that" (Stokhof, 2022). This provides the idea that different ways of understanding ourselves are, in a sense, different aspects under which we can see ourselves. "Free agents" and "victims of fate or circumstance" would be like the duck and the rabbit: by changing the way we look at ourselves we can understand ourselves one way or the other. However, here it must be noted that this only works when we interpret aspect perception (like at least both Baz and Stokhof seem to do) as being not just about visual perception.

Except when looking into a mirror, we do not *see* ourselves in the literal way we *see* an ambiguous figure. So, aspect perception should be seen as a broader capacity: the capacity to engage with reality in different ways.

However, the idea that understanding oneself as free is a manifestation of aspect perception leads to a problematic form of circularity. The problem is that perceiving aspects *already involves* understanding oneself as free in some sense (Eldridge, 2010). This is pointed out by Wittgenstein himself when he remarks that "An aspect is subject to the will. [..] It is essential that one can say "Now see it like *this*" " (1982, pp.897–902). This does not need to be true for every particular instance of aspect perception. As Wittgenstein acknowledges, sometimes we do not manage to perceive certain aspects that, according to others, should be perceivable, and often it seems impossible *not* to see certain aspects (as for example when we see an object in a field as a cow). However, Wittgenstein's point is that the general idea that human beings can perceive "under an aspect" is only intelligible insofar as we take aspect perception to be *to some extent* subject to the will (see also Stokhof, 2022). Like his other ideas about freedom, I do not think this should be read as a metaphysical point. My proposal is that it should be considered a claim about the first-personal perspective that comes with being able to engage with reality in different ways. This suggests that the human capacity for aspect perception would necessarily involve understanding ourselves as free.

However, if the stance of freedom *necessarily* comes with our capacity for aspect perception, how can this be squared with the idea that the stance of freedom is an optional stance? And doesn't it imply that understanding oneself as a victim of fate or circumstance involves adopting a paradoxical self-relation? After all, if you see yourself under the aspect of a victim, this means that you are capable of aspect perception, and this means that you *must* also understand yourself as free. Are these different self-understandings merely paradoxical, or actually contradictory? I want to suggest an interpretation that might solve these difficulties. The proposal starts from the acknowledgment that in learning to see aspects, agents necessarily learn to take the stance of freedom. By acquiring language and becoming "conceptually fluent", we learn to see reality under different aspects, and because aspect perception is "subject to the will" in Wittgenstein's sense, in this developmental process, we simultaneously learn to understand ourselves as free. Of course, this can only happen because seeing aspects doesn't take place in a vacuum: it is embedded in a social and communicative practice. As Stokhof notes: "The freedom that aspect change allows is made meaningful by its relationship with what makes a practical difference. And practical differences are strongly concerned to what concerns us in our everyday lives" (Stokhof, 2022, p.12). But this doesn't entail that the stance of freedom then becomes the *only* way in which we can understand ourselves. In learning to see different aspects, we "expand the experience of the ordinary and the familiar" (Baz, 2000), and the consequence is that aspect perception creates room for understanding ourselves as victims of fate and circumstance.

To summarize: my proposal is that understanding oneself as either free or as a victim of fate and circumstance should be seen as different stances we can take toward ourselves. Our sense of freedom comes first, as it immediately comes with the seeing of aspects. But once we come to see aspects, we come to learn to expand our experience, the result of which is that we can *also* come to understand ourselves in more fatalistic ways. This means that such forms of self-understanding are indeed paradoxical but not contradictory: the basic experience of oneself as free that comes with aspect perception can be the foundation that enables us to also understand oneself in certain contexts as victims of fate and circumstance. Such a fatalistic self-understanding does not undermine the basic freedom that comes with aspect perception, but actually presupposes it: when in a specific situation you understand yourself as a victim, this presupposes that you do not *always* understand yourself as such.

13.4 Conclusions

In conclusion, I will recapitulate the Wittgensteinian ideas developed on what it means to understand oneself as free, and end with some brief suggestions on how these ideas could help make sense of the concept of situated agency, and more generally might be valuable for situated perspectives on cognition.

First, Wittgenstein has developed the idea that freedom indicates a stance that human beings can, but need not always, take toward themselves. This means that freedom is neither a metaphysical nor a scientific concept. For Wittgenstein, the question is not whether we are genuinely free or not (in fact, he would argue that this question is a classic example of philosophers misleading themselves into conceptual confusion) but what it means to understand ourselves as free or as not-free. He argues that understanding ourselves as free is "optional", in the sense that it is not a stance we must necessarily or always take, but a stance that it makes sense for us to take in specific contexts and practices.

Second, I have argued that for Wittgenstein, the stance of freedom is a thoroughly practical stance: we take such a stance only because, and in so far as, doing so makes a practical difference in our lives. This clarifies the way in which understanding ourselves as free is "optional": in contexts where it would not have any practical implications, the question of how to understand ourselves won't even come up. It is only in situations where we are called on to explain or justify ourselves, that it becomes necessary to understand ourselves. And importantly: the particular stance we take will also be guided by practical requirements and implications. After all, in adopting a stance of either "being free" or "being a victim", we are not adopting a metaphysical position: we are attempting to deal with practical problems of living. I developed the suggestion that taking the stance of freedom can be related to Wittgenstein's thinking about aspect perception. Seeing aspects "reflects our awareness of the intrinsic plurality of our engagement with the world"

(Stokhof, 2022) and this already entails a sense of freedom. However, once we experience this sense of freedom, aspect perception also enables us to take other stances toward ourselves, such as understanding ourselves in relevant contexts as victims of fate and circumstance.

Now, how could these ideas contribute to an understanding of agency from situated perspectives on cognition? My first suggestion here is that there is value in Wittgenstein's perspective on how agency relates to environmental constraints. Yes, our agency is constrained by nature, but only in the obvious sense: namely, in the sense that we are constrained by regularities like the laws of gravity. In this way, Wittgenstein offers an account of what it means to understand ourselves as agents that steers clear of metaphysical commitments and is compatible with whatever science tells us (and will tell us in the future) about the world. Especially, the latter feature will be appealing for many 4E perspectives on agency and cognition, as these perspectives usually hope to bring about a stronger integration between philosophical work and the empirical sciences of the mind (Clark, 2013; Newen et al., 2018; Segundo-Ortin, 2020). At the same time, an important message conveyed by the Wittgensteinian point of view might precisely its resistance to the intuition that our self-understanding is grounded in scientific facts and knowledge. In my view, emphasizing that the sense of agency is not a scientific or factual understanding might be crucial in order to create room for a genuinely first-personal perspective on agency, something many 4E perspectives are struggling with.

A second suggestion is that Wittgenstein's ideas provide support for specifically those situated approaches to agency that emphasize the importance of narratives in agents' self-understanding (Fiebich, 2016; Hutto, 2012, 2016; McConnell, 2016). To briefly return to the example of the person telling themselves "I am not a hero": Wittgenstein's analysis of examples like these can be used to clarify how such "micro-narratives" play an important role in negotiating social expectations. It could be argued that we use narrative discourse when we have practical reasons to direct the attention of others and/or ourselves toward certain aspects or precisely away from certain aspects. This interpretation could also help narrative approaches in analyzing cases in which one's self-narrative gets stuck, for example, in certain cases of addiction (McConnell, 2016). Getting stuck in a "passive" narrative as frequently happens in addiction might be analyzed as the development of a stable understanding of oneself as a victim of fate or circumstances, showing how this self-understanding can impact an agent's participation in certain meaningful social practices.

With these brief suggestions, I have hoped to show that 4E approaches could draw various kinds of insights from Wittgenstein's ideas about agency and freedom. However, this is not to say that his perspective can be simply imported into debates on situated cognition. I want to end by pointing out a feature of Wittgenstein's approach that might be in tension with certain strands of thought in 4E cognition. Many approaches in situated cognition

emphasize the need to conceptualize mind and cognition in terms of features that are found (although in different ways) in different kinds of organisms (Barrett, 2015; Costall, 2004; Godfrey-Smith, 2016). Wittgenstein, on the other hand, was primarily concerned with human beings in their social practices. He even states that:

> When we talk of volition, we first of all talk of the human form of it […] Distinctions which we make are made for the sake of the normal human being. Then there are also applications to all sorts of borderline cases.
> (edition 2017, p.264)

This raises the question whether Wittgenstein's perspective on freedom and agency might only be helpful to illuminate human agency, or even whether his perspective might be vulnerable to the bias of anthropectomy, or the illegitimate denial of certain valued features to non-human beings (Andrews & Huss, 2014). To evaluate whether this problem is real or merely apparent would require a paper of its own; but it is important to point out that it is precisely Wittgenstein's focus on the human form of life that enables him to shed light on the question of what it means *for us* to understand ourselves as free.

Acknowledgments

This research was supported by the Dutch Research Council (grant VI.VIDI.195.116). I want to thank the organizers of and participants in the workshop *Analytic Philosophy and 4E Cognition* (Granada, June 2024) for their helpful feedback and support, and Josephine Pascoe for incredibly valuable ideas and comments on an earlier draft.

References

Andrews, K., & Huss, B. (2014). Anthropomorphism, anthropectomy, and the null hypothesis. *Biology & Philosophy*, *29*(5), 711–729. https://doi.org/10.1007/s10539-014-9442-2

Anscombe, G. E. M. (1971). *Causality and determination*. Cambridge University Press.

Barandiaran, X., Di Paolo, E. & Rohde, M. (2009). Defining Agency. Individuality, Normativity, Asymmetry and Spatio-temporality in action. *Journal of Adaptive Behavior*, 17(5), 367–386. https://doi.org/10.1177/1059712309343819

Barrett, L. (2015). *Beyond the brain: How body and environment shape animal and human minds* (First paperback printing). Princeton University Press.

Baz, A. (2000). What's the point of seeing aspects? *Philosophical Investigations*, *23*(2), 97–121.

Baz, A. (2020). *Wittgenstein on aspect perception*. Cambridge University Press.

Bevir, M. (2017). Situated agency: A postfoundational alternative to autonomy. In: Kevin M. Cahill , Martin Gustafsson and Thomas Schwarz Wentzer (Eds). *Finite but Unbounded: New Approaches in Philosophical Anthropology* (pp. 47–66). De Gruyter.

Budd, M. (1987). Wittgenstein on seeing aspects. *Mind*, *96*(381), 1–17.

Buhrmann, T., & Di Paolo, E. (2017). The sense of agency – a phenomenological consequence of enacting sensorimotor schemes. *Phenomenology and the Cognitive Sciences*, *16*(2), 207–236. https://doi.org/10.1007/s11097-015-9446-7

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Conway, J., & Kochen, S. (2009). The strong free will theorem. *Notices of the AMS*, *56*(2), 226–232.

Costall, A. (2004). From Darwin to Watson (and cognitivism) and back again: The principle of animal-environment mutuality. *Behavior and Philosophy*, *32*(1), 179–195.

Day, W., & Krebs, V. J. (2010). *Seeing Wittgenstein anew: New essays on aspect-seeing*. Cambridge University Press.

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). Sense-making and agency. In *Linguistic bodies: The continuity between life and language* (pp. 32–34). MIT press.

Eldridge, R. T. (2010). *Wittgenstein on aspect-seeing, the nature of discursive consciousness, and the experience of agency*. In: W. Day and V.J. Krebs (Eds). *Seeing Wittgenstein anew: New essays on aspect-seeing* (pp. 162–180). Cambridge University Press, https://doi.org/10.1017/CBO9780511750663.010.

Fiebich, A. (2016). Narratives, culture, and folk psychology. *Phenomenology and the Cognitive Sciences*, *15*(1), 135–149. https://doi.org/10.1007/s11097-014-9378-7

Gallagher, S. (2009). *Philosophical antecendents to situated cognition*. In P. Robbins and M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 35–51). Cambridge University Press, https://doi.org/10.1017/cbo9780511816826.003.

Gallagher, S. (2012). Multiple aspects in the sense of agency. *New Ideas in Psychology*, *30*(1), 15–31.

Glock, H.-J. (2016). Aspect-perception, perception and animals: Wittgenstein and beyond. *Wollheim, Wittgenstein, and Pictorial Representation* (pp. 77–100). Routledge, https://doi.org/10.4324/9781315640983-4.

Godfrey-Smith, P. (2016). *Other minds: The octopus, the sea, and the deep origins of consciousness*. Farrar, Straus and Giroux.

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1451), 1775–1785. https://doi.org/10.1098/rstb.2004.1546

Hagberg, G. L. (2010). *In a new light: Wittgenstein, aspect-perception, and retrospective change in self-understanding*. In: W. Day and V.J. Krebs (Eds.). *Seeing Wittgenstein anew: New essays on aspect-seeing* (pp. 101–119). Cambridge University Press, https://doi.org/10.1017/cbo9780511750663.007

Hutto, D. D. (2012). *Folk psychological narratives: The sociocultural basis of understanding reasons*. MIT press.

Hutto, D. D. (2016). Narrative self-shaping: A modest proposal. *Phenomenology and the Cognitive Sciences*, *15*(1), 21–41. https://doi.org/10.1007/s11097-014-9352-4

King, A. (2009). Overcoming structure and agency: Talcott parsons, ludwig wittgenstein and the theory of social action. *Journal of Classical Sociology*, *9*(2), 260–288. https://doi.org/10.1177/1468795X09102125

Kolber, A. J. (2015). *Free will as a matter of law*. In D. Patterson and M. S. Pardo (Eds.), *Philosophical Foundations of Law and Neuroscience* (pp. 9–28), https://doi.org/10.1093/acprof:oso/9780198743095.003.0002.

Maiese, M. (2022). *Autonomy, enactivism, and mental disorder: A philosophical account*. Routledge. https://doi.org/10.4324/9781003174103.

McConnell, D. (2016). Narrative self-constitution and recovery from addiction. *American Philosophical Quarterly*, *53*(3), 307–322.

McGann, M. (2014). *Situated agency: The normative medium of human action*. *Synthesis philosophica*, 29(2), 217–233.

Melzer, T. (with Melzer, T.). (2002). Atlas of Aspect Change: A book on Shifting Meanings with Tine Melzer. In *Melzer, Tine (2002). Atlas of Aspect Change: A book on Shifting Meanings with Tine Melzer Zürch: Rollo Press*. Rollo Press. https://www.ideabooks.nl/9783906213361-tine-melzer-atlas-of-aspect-change

Mojica, L. (2021). The enactive naturalization of normativity: From self-maintenance to situated interactions. *History and Philosophy of the Life Sciences*, *43*(4), 127. https://doi.org/10.1007/s40656-021-00479-3

Mulhall, S. (1990). *On being in the world: Wittgenstein and heidegger on seeing aspects*. Routledge.

Mulhall, S. (2001). Seeing aspects. In H.- J. Glock (Ed.), *Wittgenstein: A critical reader* (pp. 246–267). Blackwell Oxford.

Newen, A., Bruin, L. D., & Gallagher, S. (Eds.). (2018). *The Oxford handbook of 4E cognition*. Oxford University Press.

Pleasants, N. (1997). Free to act otherwise? A Wittgensteinian deconstruction of the concept of agency in contemporary social and political theory. *History of the Human Sciences*, *10*(4), 1–28. https://doi.org/10.1177/095269519701000401

Plotica, L. P. (2013). "This is simply what I do": Wittgenstein and oakeshott on the practices of individual agency. *The Review of Politics*, *75*(1), 45–68.

Ryle, G. (1954). *Dilemmas: The Tarner lectures 1953*. Cambridge University Press.

Segundo-Ortin, M. (2020). Agency from a radical embodied standpoint: An ecological-enactive proposal. *Frontiers in Psychology*, *11,* 1319, https://doi.org/10.3389/fpsyg.2020.01319.

Segundo-Ortin, M. and Kalis, A. (2024). Intentions in ecological psychology: An Anscombean proposal. *Review of Philosophy and Psychology* 15, 69–89. https://doi.org/10.1007/s13164-022-00661-x.

Stokhof, M. (2022). Philosophy as Change. In T. Melzer (Ed.), *Atlas of Aspect Change* (pp. 61–79). Rollo Press.

Williams, M. (1985). Wittgenstein's rejection of scientific psychology. *Journal for the Theory of Social Behaviour*, *15*(2), 203–223.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Blackwell.

Wittgenstein, L. (1982). *Last writings on the philosophy of psychology, Vol. II* (GH von Wright and Heikki Nyman, Ed.; CG Luckhardt and Maximilian AE Aue, Trans.). University of Chicago Press.

Wittgenstein, L. (2017). *Wittgenstein's Whewell Court lectures: Cambridge, 1938–1941* (Y. Smythies, Ed.). John Wiley & Sons.

Wolfram, S. (1997). *A new kind of science*. Free Press.

14 Recognition and the grounding of normativity

Laura Mojica

14.1 Normativity: from analytic philosophy to e-cognition

Ecological and enactive approaches characterize human cognition as a normative phenomenon that unfolds in agent-environment interactions. Various authors within these traditions have adopted a Wittgensteinian framework, where the normativity of characteristically human forms of cognition is understood as constituted by social practices. The correctness, meaningfulness, and intentionality of what we do and say are determined by the social practices we participate in. This has been a decisive step away from internalist and individualist accounts of meaning and mind, one that has proven especially attractive to analytically inclined proponents of the e-approaches.

Wittgenstein's insight responds to a central question of the analytical tradition: what determines the meaning of language and the contents of the mind? The dominant answer has long relied on representationalism. Language, on this view, represents the world; its meaning consists in the state of affairs it depicts, and this representational content can be true or false. Analogously, we humans make sense of the world by forming accurate mental representations. Against this picture, Wittgenstein argued, most notably in the *Philosophical Investigations*, that the meaning of linguistic expressions is determined by their use, that is, by their public, practical consequences. These consequences are constituted in our shared forms of life. Two crucial implications for the philosophy of mind follow. First, that expressions referring to mental states, such as desires and beliefs, do not derive their meaning from inner mechanisms (e.g., mental representations or neural activations), but from their public expression and practical role in social life. Second, that meaningfulness is not confined to language: actions and gestures are themselves meaningful as part of a shared form of life.

This idea has been central to contemporary work in e-cognition, especially within ecological psychology and, to a lesser extent, enactivism. In ecological psychology, Rietveld and Kiverstein, drawing explicitly on Wittgenstein, have argued that the skills which enable us to engage with affordances, that is, opportunities for action, are acquired through participation in social practices. These practices are our relatively stable ways of living together.

DOI: 10.4324/9781003649359-18
This chapter has been made available under a CC BY 4.0 license.

Participating in them grounds what it means to engage correctly with an affordance: "the abilities that are acquired by participating in practices are abilities to act adequately according to the norms of the practice" (Rietveld & Kiverstein, 2014, p. 330). From an opposing perspective, Heras-Escribano (2019) has argued that affordances cannot be normative in themselves, since they are lawful regularities, and lawfulness excludes normativity. Nonetheless, he acknowledges that affordances are shaped by our normative practices. It is through participating in practices, he argues, that we acquire the capacity to distinguish between "it is correct" and "this seems correct to me", a distinction that plays a fundamental role in our human form of life.

From a radical enactive perspective, Hutto and Satne (2015) have maintained that so-called content-involving attitudes (e.g., beliefs, desires) are possible only through social ascriptive practices—practices in which we make sense of reasons for action, both our own and those of others. Even within enaction, which tends to lean on phenomenological resources, Wittgenstein's idea of meaning as use within social practices has been instrumental to support a clarification between two normative poles: that of the agent, who acts from a first-person perspective and for whom outcomes matter, and that of the social practices in which the agent participates and which regulate, shape, and determine the appropriateness of her actions (Mojica, 2021).

Across these perspectives, we find convergence on a key Wittgensteinian idea: social practices establish the normativity that determines whether an agent's engagements with the world are correct, skillful, or intentional. But this convergence raises an important and often underexplored question: what does it mean to act in accordance with a practice? What precisely is involved in being normatively aligned with a practice's patterns of behavior?

14.2 Brandom's recognition and normativity

We can begin to answer this question with the intuition that *recognition* of an individual's behavior as correct, intentional, or skillful by established members of a practice makes it so. Thus, we can take as an initial claim that acting in accordance with a practice means *being recognized* as a participant by an established participant of the practice. Brandom has addressed this idea through an exhaustive interpretation of Hegel's philosophy. He offers a pragmatist reading of Hegelian philosophy that answers classic questions of the analytic tradition such as the nature of actions and linguistic meaning. In this sense, he aims to offer answers assuming a "pragmatism about semantics" and cognition (2019, p. 753; Houlgate, 2020). Particularly relevant for our question is that Brandom sees his reading of Hegelian reciprocal recognition and its consequent self-consciousness as grounding "an account of what it means for norms to be implicit in social practices" (Brandom, 2019, p. 761).

To offer a Hegelian account of recognition within pragmatic semantics and cognition, Brandom addresses the challenge to interpret it as an activity

motivated by desire. This is important not only for exegetical rigor, but because we are fundamentally desiring animals and recognition emerges from this basis. Our desiring nature gives us a basic form of awareness, given how it relates to our activities and how we give significance to things. According to Brandom, these elements establish a triadic structure: a motivating *attitude* (a desire), a responsive *activity* motivated by the attitude (to fulfill the desire), and a practical *significance* of an object as what allows the individual to successfully carry out the activity (what fulfills the desire). Given Brandom's pragmatism, acting out of a desire and taking the world as significant in fulfilling that desire are equivalent in practice. This triadic structure attitude-activity-significance gives rise to a basic form of awareness, orectic awareness, where the world becomes significant to an individual in terms of what satisfies its attitudes. Importantly, a basic form of normativity also arises with this triadic structure, as the subjective significance the individual attributes to a thing is correct or incorrect depending on whether or not it satisfies the attitude that motivated the response.

In Brandom's reconstruction of Hegel's *Phenomenology of the Spirit*, recognition results from the triadic structure of desire. Following Hegel, we all fundamentally have the desire of being recognized, which moves us to recognize others and signify them as someone who can attribute significances to the world and act intentionally motivated by desires (i.e., motivating attitudes). As will be shown below, in recognizing the other as a desirer we acknowledge her normative authority, and by recognizing her as a recognizer we acknowledge her desire for recognition. Recognition does not occur abstractly; rather, we recognize others in their concrete activities, that is, as someone who does, a specific activity *K*, in order to satisfy her *K*-desire and gives *K*-significance to things in the world.

Brandom labels "simple recognition" to recognizing someone as a desirer and "robust recognition" to recognizing her as a desirer that can recognize others. This distinction is important because only by recognizing the other both as a desirer and as someone who recognizes others, our desire for recognition is satisfied and we have successfully carried out the activity. Let us unpack these dense claims through an example.

I see you putting in your mouth something that looks like a small rock to me, chewing and swallowing it. I see you eating the small rocks and stopping at some point, presumably when your hunger is satiated, and I see you eating them later again when your hunger seems to have reappeared. I can recognize you as a desirer, in this case, as someone who is hungry (a desire), who eats that small rock motivated by that desire (a motivated response), and that takes that small rock as food (a significance). Note that simple recognition allows me to understand your activities as part of the triadic structure of desire (Figure 14.1). But more importantly, it allows me to acknowledge that some things are still food when I am not hungry. In doing so, I see that things have significance beyond my desires and, should anyone have the same desire, those things will satisfy it. In that sense, I attribute you normative

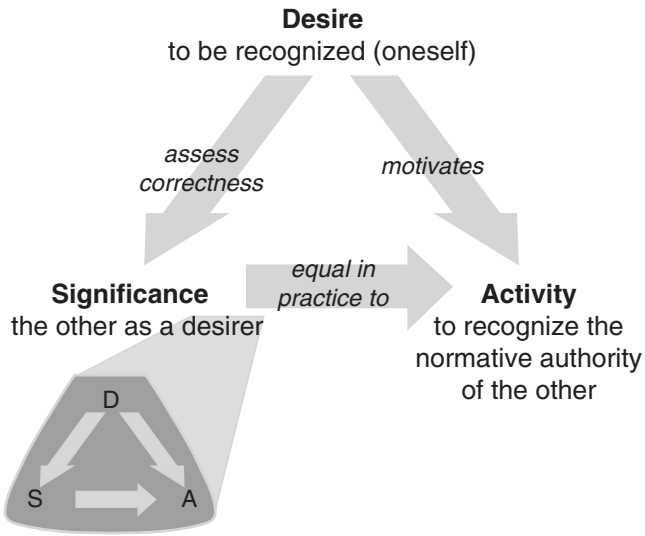

Figure 14.1 Simple recognition.

authority in that I take you as committed to the fact that the small rocks you eat will satisfy anyone's hunger, which practically implies that I will eat them when I am hungry and I will hold you responsible of consistently taking them as food. In virtue of my acknowledging your normative authority, I see you as an individual who does not act on the basis of impulse, but with a commitment to how things are, in this case, to taking these rocks as food even when you are not hungry. However, this simple recognition does not satisfy my desire of being recognized myself by others.

Recognition becomes robust when I take you both as a desirer and someone who can recognize others as recognizers. This means that I take you as aware that the there are others who desire, act motivated by that desire, and for whom the world has significance. Your capacity to recognize others, which crucially include me, allows you to see that the world has significance beyond your immediate desires and to attribute normative authority to them and to me. Note that my acknowledgment of you as someone who can recognize puts you in a position of being able to recognize me as a desirer thereby fulfilling my desire (Figure 14.2). I thus become a normative authority to you and you can hold me accountable for my actions and the significances I give to the world.

Important to our purposes is the normative attribution involved in recognition. In general, the authority implied in recognition is two-fold. On the one hand, I recognize that your desire gives you license to respond as you do. I also recognize your commitment to see things with the significance you do, which in practice means that I hold you responsible to be consistent and to give reasons for your judgments and actions. On the other hand, I recognize

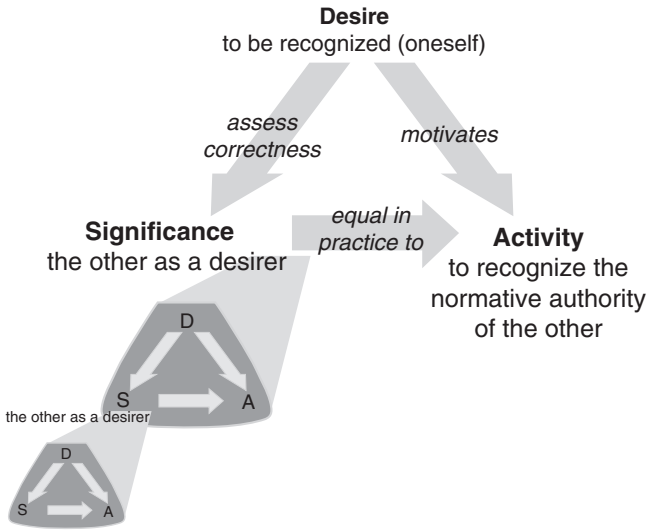

Figure 14.2 Robust recognition.

your authority over me to assess the correctness of the significance I attribute to something as a thing that can satisfy my desire and the success of my response to fulfill that desire. Brandom's explicit account of how normativity arises in recognition allows us to come back to our main question, namely, what it means to act in accordance with a practice so that it normatively legitimates an individual's action as skillful, intentional, or correct.

Recall that the initial claim was that acting in accordance with a practice means being recognized as a participant by an established participant of the practice. Using Brandom's conception of recognition, we can begin to clarify what this claim means and, crucially, why it implies a normative binding. Following Brandom's insights, we can establish that recognition is an activity, in this activity, the significance of the other is established, and it implies the attribution of normative authority to the other over herself and myself. However, there are two limitations with Brandom's account of recognition: (i) it restricts the normativity, authority, and meaningfulness established in recognition to what is reasonable, neglecting forms of meaningful behavior socially instituted that are not evaluable and understandable in terms of rationality and (ii) it restricts recognition to only those that desire and behave like me precluding the possibility of recognizing an individual as a different.

14.3 Normativity beyond reasons

In the context of Brandom's philosophy, it makes sense to understand the normative authority and demands attributed to the other in recognition through his inferentialist account of meaning. When Brandom claims that

"[t]o recognize someone is to take her to be the subject of normative statuses, that is, of commitments and entitlements, as capable of undertaking responsibilities and exercising authority" (2019, p. 245), he is emphasizing that recognition involves treating the other as a participant in practices governed by *reasons*. These commitments and responsibilities involve being responsible to justify one's claims and actions by citing reasons and being responsible for what follows from those claims. For example, if someone is committed to the claim that a certain kind of small rock is food, she is also committed to its practical and inferential consequences, such as being entitled to offer it to others or to defend its edibility if challenged. A crucial consequence of recognition, then, is that the other is taken as occupying a position in the space of reasons, where she can be held accountable for her beliefs and the inferences they license. This space of reasons is the normative domain that rational beings share, and recognition consists in acknowledging the other as a co-participant in such a space (Brandom, 1994).

Brandom's account works well if one is interested in the human capacity to make sense through the lens of rational accountability, where meaning is structured by inferential relations between claims, and intelligibility is a matter of navigating the space of reasons. However, our capacity to make sense of one another extends beyond this inferential structure. We routinely understand others through expressions that are not reducible to commitments or entitlements: we engage through greetings, expressions of emotion, jokes, rituals, etc. These forms of intelligibility are not necessarily truth-evaluable, nor do they always belong to chains of inference. Their normative force does not derive from their contribution to rational discourse, but from shared expectations, affective attunement, or embodied routines that, as we will see later, constitute social practices too.

The diversity of what we do with language is precisely one of Wittgenstein's points in the *Philosophical Investigations*, where he urges us to abandon the idea that language has a single function, especially the function of portraying a truthful representation of the world. Brandom does take seriously the idea that semantic meaning is not primarily about representing; indeed, that is one of the central motivations behind his proposed inferentialism. But Wittgenstein's point goes further: his claim is not merely about whether language represents the world or not, but about the irreducible plurality of its uses despite their surface similarity (Wittgenstein, 2009, §§11, 12, 23). From this perspective, the meaning of an utterance is grounded not in its inferential role within a propositional structure, but in its use within a shared pattern of behavior. Brandom's inferentialism, despite its pragmatic orientation, remains focused on a subset of language: those expressions that can bear inferential weight and be held to standards of justification. As such, it risks neglecting the diversity of normative structures that govern non-propositional, affect-laden, or ritualized modes of interaction.

Note that recognition involves acknowledging the other as intelligible; then, we need an account of intelligibility that reflects this diversity. Once we recognize that not all intelligibility takes the form of reason-giving, we are led to a more inclusive conception of recognition: one that is not limited to inferential commitments but includes our capacity to respond appropriately to expressions of affect, embodied styles of interaction, and shared social expectations. This shift widens the normative space in which recognition operates, making room for forms of making sense and recognizing one another that Brandom's model cannot fully accommodate. Thus, let us explore wider forms of conceiving intelligibility.

14.4 Practices of the mind

Schatzki has proposed a Wittgensteinian account of social practices and, specifically, of how the intelligibility of social action is constituted. His proposal allows us to develop a more comprehensive account of intelligibility and then to specify how intelligibility constitutes recognition, while acknowledging the diverse ways in which we make sense of the world and each other. According to Schatzki (1996), intelligibility is constituted within social practices. Just as Wittgenstein-inspired authors in e-cognition, he embraces the idea that an individual's actions are intelligible when they go in accordance with a practice. Schatzki spells out further what it means to go in accordance with a practice: it involves both being able to perform the activities proper to the practice and to identify and react to others when they perform them. The adequacy of a performance as an instance of the practice and of the agent's ongoing participation in the practice is determined by whether other members identify and react to it as an activity of the practice, rather than, for instance, an imitation or an entirely different activity.

There are various kinds of practices in which an individual can participate. Schatzki offers a classification that distinguishes between practices that can appear in different contexts, which he calls *dispersive practices*, and those that integrate our actions into more complex orders of intelligibility, which he calls *integrative practices*. Examples of dispersive practices include describing, asking questions and giving answers, giving and receiving orders, etc. These practices are usually exercised as part of an integrative practice that organizes individuals' doings and sayings through explicit rules in combination with teleoaffective orders, that is, implicit configurations of what constitutes a legitimate purpose for us to engage in a particular activity (e.g., asking for and giving reasons), and which affective states are appropriate to hold and express. However, not all integrative practices include an affective dimension. For example, cooking practices have structured purposes: we follow established procedures to make desserts as opposed to main courses, but the practice does not require that we cook with love or out of hate. In contrast, some integrative practices have affectivity at their core, such as

child-rearing and marriage where loving is central in the way we practice them now (Schatzki, 1996, p. 101). Note that participating in a practice does not imply that the individual is recognized as a minded subject capable of making sense of others and the world: we would not recognize an entity that is only capable of asking questions and giving answers as a minded individual. Thus, we can ask: what kinds of social practices must one participate in to achieve recognition as an agent that makes sense?

Schatzki's account of human mind and agency can help to begin to offer an answer. According to him, human beings are not reducible to our participation in an array of practices; rather, we are embodied creatures with unfolding life conditions that we tend to refer to as how things are going for us. Yet, in Schatzki's words, "human life transpires within social practices" (1996, p. 125). This means two things. First, it is within social practices that we learn to express those life conditions: from the basis of replacing crying with words, to sophisticated affective and mind expressions such as signing petitions to express a belief, expressing disagreements by leaving the room, or expressing through poetry and the arts complex dimensions of life that are tied together. Second, integrative practices are an especially important kind of context in which our conditions of life are both constituted and unfold, and where the chains and hierarchies of purposes and affectivities are organized (Schatzki, 1996, pp. 131–132). These include, for example, practices of marriage and romantic love, or corporate practices where our daily jobs are situated. The practices we participate in hang together, sometimes clash, sometimes complement each other, but together constitute a crucial constitutive aspect of our social identity. Let us call *practices of mind* those practices that constitute intelligible expressions of beliefs, feelings, and intentions and that are attuned to the structure of integrative practices. These include, for instance, sending flowers to express love, or furiously applying to job vacancies to express the belief that one's job is unfair.

A straightforward answer to the question we left open above would be that *practices of mind* are those we are to participate in to achieve recognition. This potential answer is not only clear but also avoids the problem of Brandom's account of recognition, namely, that it limits recognition and its consequent attribution of normative authority to reasonability. Under the Schatzki-inspired account, recognition would consist in identifying and engaging with the other as expressing an aspect of her life conditions, that is, of how things are going for her. This includes identifying individuals as expressing beliefs and holding them responsible for the inferences that follow from those beliefs. But crucially, it also includes identifying others as expressing affective states, such as grief, and adequately engaging with them by comforting and accompanying them, rather than holding them accountable to standards of reasonability.[1] Within Schatzki's account, identification

1 Emotions are indeed subject to reasonability and coherence (Helm, 2009): we hold people accountable for consistency with their affective states. Yet, our engaging and recognizing the

and engagement also imply a normative attribution: I attribute to the other the skill to correctly express her life conditions she is undergoing. In this case, "correctly" amounts to publicly intelligible, and I acknowledge that she does so in the way "we", the members of the practice, do. This attribution is what enables me to engage with her in her expression of grief. Thus, in the tentative definition of recognition as participation in practices of the mind, recognition also implies a normative attribution of the other's capacity to appropriately act and respond within our shared forms of expression.

Moreover, this tentative answer has the advantage of implying, as Brandom's account also does, that recognition should be mutual. In general, when I recognize an individual as a member of a practice, I identify her doings and sayings as correct enactments of the practice, and I attribute to her the capacity to respond to such doings and sayings when encountering another member of the practice. This means that I expect her to identify my actions and engage with me in the way we do in the practice. In this sense, I recognize in her the authority to recognize me as a member of the practice. This is also the case in practices of the mind, where not only do I recognize you as a participant of an integrative practice like cooking or academia, but also as someone who skillfully expresses life conditions and is thus able to skillfully engage with others when they express theirs.

However, the Schatzki-inspired account of recognition I just sketched does not integrate at least two points that are fundamental to recognition. First, normativity and intelligibility under this account are limited to "us", more precisely, one's understanding of another as a human being risks reducing her to what fits within what one does and "we" do within our shared practices of the mind. As pointed out above, this is also a limitation of Brandom's account. Second, the other, as a human being, is more than what is intelligible through social practices. This point is anticipated by Schatzki's own claim about the irreducibility of the human being to the practices she participates in. But equating recognition with participation in practices of mind confines the recognition of the other to her participation in practices. This practice-bounded understanding of recognition risks obscuring her humanity, namely, her particularity and uniquely embodied perspective on the world, beneath the uniformity of shared social practices. Schatzki's account is not necessarily incapable of accommodating these two concerns. After all, grounding recognition in intelligibility does not mean that recognition is equal to intelligibility, and he explicitly acknowledges the embodied irreducible nature of human beings. Yet, his framework does not provide the conceptual tools to overcome the shortcomings just discussed.

---

other as an affective being is not reducible to accountability, but we deem correct to engage affectively with emotional expressions.

14.5 The [im]possibilities of recognizing the other in her difference

Let us begin unpacking why Brandom's account does not leave room for recognizing the other in her difference. Consider Brandom's characterization of the most basic case of recognition:

> [I]n the most basic case, one can take another to be a K-taker only if one is oneself a K-taker. Taking the other to be a K-taker will then be attributing to him activity of the same sort in which one oneself engages in response to things one (thereby) takes to be Ks.
>
> (Brandom, 2019, p. 252)

Brandom's point in this passage is that we do not recognize others abstractly but only practically in their responsive activities. Recognition fundamentally occurs in taking the other as desiring something specific (e.g., being hungry), acting on that desire, and seeing significance in things as potential fulfillers of that desire. What I would like to highlight about Brandom's characterization is that the other's triad desire-activity-significance must be something I myself engage in to be able to recognize her as a desirer. The rationale seems to be that I might not have the sensitivity and conceptual tools to understand the other's actions as intentional and reasonable unless the other acts, feels, and makes sense of the world and others in a way that I myself would do. Following this reasoning, the only possibility of recognizing the other is to assimilate her into my own ways of engaging with the world.

This assimilation also occurs in the attribution of authority that comes with recognizing the other. Brandom is right in understanding the authority I recognize in the other as such insofar as it has practical implications for me. But he takes these implications to consist in authorizing my responses and forms of signifying should I have the same desire. In his words:

> My specific K-recognitive response to you is to acquire the disposition: if I have the K-desire, then I will K-respond to the things to which I (thereby) take you to have successfully K-responded. My acknowledging *your* K-desire as *authoritative* in the dual sense of licensing *your* responsive K-activity and serving as a standard of *normative* assessment of its *success* or *correctness* consists in my treating it as *authorizing* my *own* K-takings, should I have a K-desire.
>
> (Brandom, 2019, p. 252)

Note that recognizing the other and attributing normative authority to her consists in you authorizing me to behave like you if I have the same desire. This assumes that I am capable of having the same desire and having the skills to behave like you. It presupposes, in Schatzki-inspired terms, that we share life conditions and participate in similar practices of mind.

The problem with this idea of recognition is that it leaves no room for acknowledging the other as genuinely different, that is, someone whose way of making sense of the world is not already available to me. This limitation becomes especially clear when it comes to people who do not share a form of life in fundamental ways. For example, recognition between an atheist and a Mormon, who rejects blood transfusions, motivated by the desire to go to heaven in the afterlife, would be impossible, because they do not share the fundamental desire for eternal salvation. Each would likely regard the other as unreasonable, foreclosing the possibility of recognition and dialogue.

This problematic scenario also occurs *within* shared practices. Individuals' participation in a practice is shaped by their gender, race, sexual orientation, and other social positionalities. Because of this, the desires, significances, and other life conditions that, e.g., a racialized woman expresses may be unintelligible, unreasonable, and therefore unrecognizable to a non-racialized man, and vice versa, within corporate, rearing, or religious practices. It is deeply problematic that a conception of recognition precludes the possibility of encountering and recognizing the humanity and legitimacy of those who experience the world differently than we do. Difference, in this context, does not imply irrationality, but a way of making sense that falls outside what dominant frameworks count as intelligible.

Avoiding this problem demands that we recognize the normative authority of those who are different from me, from "us". In Brandom's terms, this means that normative authority should not be reserved for those whose desires and activities are like mine. While in Schatzki's terms, it means recognizing others whose behaviors I do not know how to engage with or that I do not fully identify. What is needed is a conception of recognition that allows for intelligibility across difference, rather than through assimilation.

14.6 A way out: engaging epistemologies and recognition

A core part of the problem is that recognition in these accounts follows an image of knowledge and intelligibility as the imposition of an individual's ways of making sense onto the object of knowledge, even when that object is another human being. De Jaegher (2021) develops this point. She argues that the primordial form of human knowledge is an *implicated* relation in which both knower and known are transformed. Knowing is an active relation we establish with the world, in which we simultaneously tend to determine the other and are ourselves determined by the other. This applies both to our knowledge of the inert world and of others, though only the latter concerns us here. Yet, philosophical characterizations of knowledge have neglected its reciprocity, reducing knowledge to a unidirectional determination of the knower over the known.

This asymmetry is at the center of Brandom's abstract conception of recognition, where no interaction occurs between recognizer and recognized. In his account, recognition is framed as a detached attribution of beliefs,

desires, and normative authority from an uninvolved perspective, where my own desires remain unchallenged and unchanged. In this framework, I project my motivational attitudes and desires onto others and recognize them only if they conform to it, which is a form of overdetermining the other by not letting herself be, that precludes mutual implication.

De Jaegher advocates for an engaged, or engaging, form of knowing, where the tension between over- and under-determining the other is explicitly acknowledged and navigated. The guiding principle is letting the other and myself be, while letting ourselves be transformed together in our interaction. Her proposal, grounded in enaction, acknowledges that we are always already implicated in our encounters with the world, including knowing the world, and we imbue them with significance. According to this view, significance arises from our embodied nature and its inherent vulnerability: the world is significant to us because it threatens or supports our continued existence. This applies not only to biological life but also to sensorimotor and social identities, which must be actively maintained to persist. This ongoing engagement is known as sense-making.

Crucially, in our interaction with other human beings, this basic form of significance is not mainly about maintaining our living identities, but rather our social identities. Activities like dancing or conversing acquire a dynamics of their own and must be actively sustained to continue. This requirement introduces a shared normativity: what allows the interaction to persist shapes how individuals make sense of themselves and the other. Individual's identities and their concerns do not disappear: they are implicated as they are actively maintained in the interaction or are potentially or actively threatened. The interaction is thus co-constituted by each individual's concerns and sense-making processes and by the emerging dynamics between the two, which in turn feed back into their individual ongoing sense-making processes. Insofar as an individual's agency is maintained while both the other and the dynamic itself participate in the individual's sense-making process, there is a genuinely social interaction; what De Jaegher and Di Paolo (2007) call participatory sense-making.

With this framework, we can return to the Schatzki-inspired notion of recognition. While it involves real interaction, where both expression of mind and the engagement it elicits are enactments of a shared practice, it risks overdetermination: recognition is only possible if I already possess the skills to identify your expressions of mind. The concept of participatory sense-making helps refine how the engagement within practices of the mind occurs explicitly leaving room for navigating and engaging *in* the unknown. They imply an ongoing negotiation of over- and underdetermination, where the practices of the mind provide established choreographies for agents to establish an interaction where complex forms of sense-making can emerge (Di Paolo et al., 2018). Reconceptualizing recognition in this way opens the door to relating to the other not just as a participant in a practice, but as a

unique person with whom novel, idiosyncratic forms of interaction, expression, and participatory sense-making can arise.

As De Jaegher has rightly observed, implicated ways of knowing share their core structure with loving relations. In her words, "in their most minimal, stripped down form, loving and knowing are manifestations of the same basic, existential way of relating" (2021, p. 860). In both, one is transformed by loving or knowing the other, the other is also transformed by being loved or known, and both are vulnerable to losing themselves in the relationship by being overdetermined by the other while trying to maintain their own individual becoming into who they are. The same vulnerabilities and dialectics appear in interactions within practices of the mind where I engage with you as expressing how your life is going. Note that this structure of vulnerabilities, and dialectical transformation gives us a normative standard for interactions, whether of loving, knowing, or recognizing: they should aim for a balance between letting implicated individuals be and letting them be transformed. When this balance is systematically lost, loving relations are abusive, knowledge is biased, and recognition is rather a form of repression, silencing, or alienation.

What does this tell us about the social normativity that constitutes meaning, skill, or intentionality? To answer this, let us recapitulate what has been discussed so far. Brandom's account helped us clarify how recognition implies acknowledging the other's normative authority; Schatzki's practice theory helped us restrict the claim to recognition within practices of the mind. Now, through enaction and De Jaegher's engaged/-ing epistemology, we see that recognition within such practices is a dialectical process of letting oneself be transformed by the other and maintaining one's continued identity. Recall that our initial claim was that acting in accordance with a social practice involves being recognized as a participant by other participants and that acting in accordance with a practice is what it means to act in accordance with a norm of intelligibility, skillfulness, or correctness. So, under our refined conception of recognition, correct, skilled, and intelligible behavior is an ongoing dialectical construction that occurs in interaction within practices of the mind. Their dialectic character gives interactions the potential of surpassing the established forms of recognition within such practices.

There is one last worry that this account raises. It is impracticable to demand implicated recognition in all our routine social interactions and political spheres. We do not have to undergo a deep, implicated transformation in our interaction with the supermarket cashier or the immigration policeman to recognize him as a human being and his doings as intelligible, nor do we need to do so with people whose precarious life conditions demand political action to recognize their humanity, empathize with their suffering, and demand or take political action. Furthermore, we are not required to maintain an implicated, engaging relationship that hurts us or systematically diminishes our agency and well-being, but this does not mean we must stop

recognizing them as participants in our practices of mind. We are able to recognize the other's humanity while deciding to actively avoid engaging with them. Note that the inverse can also be the case and would constitute a normative failure: we have the possibility of not recognizing the humanity of the cashier, the policeman, and people in need, and consequently finding them unintelligible or overdetermining them with our fixed prejudices.

This worry motivates the last tweak to our working definition. Becoming a participant of a practice requires implicated and sustained participatory sense-making with another, typically a mother, a mentor, a loved partner, etc. However, once we are members of shared practices of mind, we can establish interactions with others where we do not deeply and actively engage in the dialectics of mutual transformation and individual continued identity, but instead recognize this implicated interaction as a possibility for the other. This means that recognition lies in taking the implicated interaction with the other as a *real possibility*; not as an abstract, remote, but practically unrealizable one. This implies that one should have the affective disposition to see the other as such, that is, as someone who can transform and be transformed by another, and who is entitled to seek the balance between our together transformation and an individual, creative maintaining of her identity. To continue with De Jaegher's parallelism, we should see the other as *lovable*.

Situating recognition in the field of the virtual allows us to see the other, e.g., the cashier in the supermarket, as a member of practices of the mind that may or may not coincide with mine, to understand his gestures, actions, and locutions as expressions of those practices that currently only have a transactional practical significance to me. Even if our interaction is merely transactional, he still deserves to sustain implicated relations with others, just as I do. Crucially, recognizing him as a subject of virtually implicated interactions allows me to engage with him by letting him be and allowing myself to be determined by him, as he should do toward me, should a breakdown arise. Many questions remain open with the preliminary conception of recognition I have sketched. Still, further developing the embodied dialectics of implicated interactions and the virtuality of recognition has the potential to explain the mechanisms through which social normativities change. Understanding these mechanisms would allow us to understand and channel the possibilities of social change.

References

Brandom, R. B. (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Brandom, R. B. (2019). *A Spirit of Trust: A Reading of Hegel's Phenomenology*. Harvard University Press.

De Jaegher, H. (2021). Loving and knowing: Reflections for an engaged epistemology. *Phenomenology and the Cognitive Sciences*, 20(5), 847–870. https://doi.org/10.1007/s11097-019-09634-5

De Jaegher, H., & Di Paolo, E. A. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6(4), 485–507. https://doi.org/10.1007/s11097-007-9076-9

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic Bodies: The Continuity between Life and Language*. The MIT Press. https://doi.org/10.7551/mitpress/11244.001.0001

Helm, B. W. (2009). Emotions as evaluative feelings. *Emotion Review*, 1(3), 248–255. https://doi.org/10.1177/1754073909103593

Heras-Escribano, M. (2019). *The Philosophy of Affordances*. Springer.

Houlgate, S. (2020). [*Review of Review of A Spirit of Trust: A Reading of Hegel's Phenomenology*, by R. B. Brandom]. Harvard University Press. https://ndpr.nd.edu/reviews/a-spirit-of-trust-a-reading-of-hegels-phenomenology/

Hutto, D. D., & Satne, G. (2015). The natural origins of content. *Philosophia*, 43(3), 521–536. https://doi.org/10.1007/s11406-015-9644-0

Mojica, L. (2021). The enactive naturalization of normativity: From self-maintenance to situated interactions. *History and Philosophy of the Life Sciences*, 43(4), 127. https://doi.org/10.1007/s40656-021-00479-3

Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology*, 26, 325–352. https://doi.org/10.1080/10407413.2014.958035

Schatzki, T. R. (1996). *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge University Press.

Wittgenstein, L. (2009). *Philosophical Investigations*. John Wiley & Sons.

15 Embodied, embedded, enactive, extended… and exclusionary? Toward an inclusive E-Cognition for cognitive diversity

Víctor Fernández-Castro and Miguel Núñez de Prado-Gordillo

15.1 Introduction

> In speaking of neurodiversity as the diversity among minds, I use the word mind in the broadest possible sense, to encompass the totality of every aspect of perception, cognition, emotion, memory, psyche, and consciousness. Mind is an *embodied* phenomenon. (…) So when I say that neurodiversity is *the diversity among minds*, I'm really saying that it's the *diversity among bodyminds*.
>
> (Walker, 2021, pp. 53–54)

This claim by neurodiversity activist Nick Walker echoes the (embodied) specter that is haunting cognitive science—the specter of embodied, embedded, enactive, extended, and/or ecological views of mind (henceforth E-Cognition). None of the powers of old, *traditional cognitivism*—which represent the mind as an internal, representation-hungry, information-processing machine (see Menary, 2010)—have been capable of exorcizing this specter. Over the last three decades, E-Cognition views have increasingly pervaded discussions about the mind, promoting a non-reductionist, holistic, and relational analysis of agents' mental abilities as scaffolded or even directly constituted by their bodies, environments, and the interaction between the two (see Newen et al., 2018). This specter has recently started to haunt mental health (see Núñez de Prado-Gordillo & López-Silva, 2025). E-Mental Health views claim that mental health and disorder cannot be analyzed solely by examining what brains do. Rather, mental pathologies are seen as disturbances in the complex, multilevel dynamics that characterize the brain-body-environment system.

The purpose of this chapter is to analyze the merits of this E-Mental Health framework from a *neurodiversity* perspective (Chapman, 2023b; Walker, 2021). Multiple authors, including neurodivergent scholars, have emphasized how the E-framework brings forth a more inclusive, neurodiversity-affirming approach to mental healthcare (e.g., Chapman, 2021; Hoffman, 2019; Jurgens, 2020; Legault et al., 2024). Despite these strong affinities, however,

DOI: 10.4324/9781003649359-19
This chapter has been made available under a CC BY 4.0 license.

many E-Cognition scholars seem to remain problematically tied to exclusionary neuronormative assumptions that perpetuate unjust stereotypes and practices (Jurgens, 2023; Parra Rubio, 2025; van Grunsven, 2020). In this chapter, we take this critique further: E-Cognition perspectives do not only *de facto* exclude cognitive divergence, but also risk reinforcing this exclusion through their foundational assumptions about mental normativity—ultimately obstructing progress toward inclusive mental health.

Our critique stems from a non-ideal, activism-oriented metaphilosophical perspective (e.g., Keiser, 2022; see Bordonaba-Plou et al., 2022; Hänel & Müller, 2022). This perspective, which prioritizes the ameliorative goal of building theoretical frameworks that contribute to develop the insights from oppressed collectives, shares much with the E-Cognition framework: it sees theory production as an inextricably situated practice, deeply rooted in the socio-material niches of knowledge producers, not their pristine Cartesian minds. Our hearts thus lie close to the E-Cognition framework: our aim is to develop it in ways consistent with the self-proclaimed liberatory ends that both its proponents and neurodiversity advocates see in it.

The structure of this chapter is as follows. Section 15.1 introduces the Neurodiversity movement and its critique of traditional cognitivist accounts of neurodivergence. Section 15.2 then explores the main benefits and limitations of the E-Mental Health framework from a neurodiversity-affirming perspective. Section 15.3 argues that such limitations stem from a remaining Cartesian commitment at the core of E-Cognition understanding of mind: descriptivism, or the idea that mental language primarily serves a descriptive, fact-tracking, or representational function. This commitment reinforces a certain epistemic attitude toward neurodivergent contributions to discussions about their own minds, which conditions their appreciation to their accommodability within experts' preferred model of mind. Sections 15.4 and 15.5 then discuss how an alternative *expressivist* account of mental language can help us turn knowledge production in mental health right side up again, questioning the epistemic privilege often attached to mental health experts and placing neurodivergents' evaluative accounts of their own minds at its basis.

15.2 The traditional exclusion of cognitive diversity

Traditional cognitivism, typically associated with the computational and representational theory of mind (Fodor, 1989), is probably the still dominant perspective in mental health science. On this view, the mind is conceived as an internal network of brain-based computational processes that mediate perception and action, coordinating an agent's interaction with the world via the generation, manipulation, storage, and retrieval of mental representations. Despite the different developmental and social trajectories of different groups and individuals, traditional cognitivism assumes that the underlying functional architecture of human cognition is universally shared across the

species, something characteristic of *the* human mind. What counts as *normal* cognitive functioning can thus be studied and described somewhat independently of specific social norms and customs.

Mental health science often adopts this underlying commitment. ADHD is a suboptimal attentional profile that significantly deviates from normal executive functioning (Douglas, 1972). Schizophrenic hallucinations involve a radical breakdown of normal sensory processing; delusions a failure of normal metacognitive mechanisms that monitor one's thoughts (Coltheart et al., 2011; Frith, 2003). Autism is characterized by various deficits in normal social cognition (Baron-Cohen et al., 1985) or, furthermore, in the "universal", "normal information processing (…) tendency to draw together diverse information to construct higher-level meaning in context" (Frith & Happé, 1994, p. 121). This rationale is the characteristic tenet of cognitive neuropsychiatry (David & Halligan, 2000), which views psychopathology as a unique opportunity to map the anatomic and information-processing structure of *normal* cognition: by analyzing the neurobiological underpinnings of computational failures associated with different mental disorders, we may come to understand how normal brains work. Other cognitivist views (traditional Cognitive Behavioral Therapy; Beck, 1979; Bentall, 2003) reject this assumption, instead placing normal and clinical cognition on a *continuum*—differing only in the extent to which cognitive biases and distortions characteristic of *all* human cognition are expressed. However, the underlying assumption is the same: normal human cognition can be described in terms of universal information-processing mechanics. What counts as proper cognitive functioning is a matter of *empirical discovery*, of contrasting agents' cognitive profiles with the benchmark of normal cognitive functioning established by cognitive science.

This core commitment is exactly what the *Neurodiversity movement* considers an unfounded, contestable, and stigmatizing idealization that systematically marginalizes neurodivergent cognition (Chapman, 2023b; Rosqvist et al., 2020; Walker, 2021). This civil rights movement, born within the autistic community (Botha et al., 2024), challenges "The idea that there is one 'normal' or 'healthy' type of brain or mind, or one 'right' style of neurocognitive functioning" (Walker, 2021, p. 36)—the cornerstone of what neurodiversity scholars call the *pathology* or *normalcy paradigm* (Chapman, 2023b; Walker, 2021). This paradigm sees mental health as inextricably linked to cognitive typicality, portraying deviations from typical cognition as necessarily *deficient*; the bare thought of a healthy, flourishing, yet divergent mode of cognitive functioning is almost a logical conundrum—more likely a byproduct of the very cognitive dysfunctions posited by cognitive neuropsychiatrists (e.g., in self-awareness; Chapman & Carel, 2022).

Against this view, neurodiversity activism advances the *neurodiversity paradigm*. This paradigm sees mental diversity—i.e., diversity in the ways bodyminds function and express themselves—as a valuable form of diversity, worthy of respect and celebration alongside other forms of sexual, racial, or

cultural diversity. The paradigm also reclaims *neurodivergences*—i.e., systematic deviations from normal or typical cognitive standards—as potentially constitutive of personal and political identity (Jeppsson, 2022). In line with Mad and other radical mental health activisms, many neurodiversity activists view these divergences as revealing the limits of assimilation under hegemonic cognitive norms and their underpinning socioeconomic hierarchies, which can therefore be weaponized to foster collective liberation (Adler-Bolton & Vierkant, 2022; Chapman, 2023b). As Chapman (2023b) puts it, neuronormative standards of healthy mindedness are a byproduct of shifting capitalist production needs that configure an "Empire of Normality", i.e., "an apparatus of material relations, social practices, scientific research programmes, bureaucratic mechanisms, economic compulsions, and administrative procedures that […] bring a much more restricted bodily, cognitive, and emotional normal range than those seen in any previous society" (p. 15).

This revindication does not, however, entail denying the disabilities and limitations often associated with neurodivergence. Rather, in line with the social model of disability, the paradigm points to the *relational* mismatch between individual cognitive traits and their socio-material niches as their main origin (Chapman, 2021; Jurgens, 2020; Milton, 2012; Walker, 2021). What this movement opposes is the *ableist* assumption that neurotypical ability sets the benchmark for a life worth living, that cognitive disability necessarily stands in the way of a fully satisfactory and flourishing life (Chapman & Carel, 2022; Higashida, 2018; Sinclair, 1992). Furthermore, moving beyond the relational analysis of disability, neurodiversity proponents advance an *ecological* understanding of cognitive diversity, which pays attention not only to the individual contributions of certain cognitive traits, but also to the collective benefits resulting from the interaction of diverse cognitive styles. On this view, the reduction of cognitive diversity within a population can undermine collective functioning, much like a loss of biodiversity weakens an ecosystem's resilience (Chapman, 2021; Hoffman, 2017; Jurgens, 2023).

Importantly, this relational analysis of disability does not rule out understanding divergent traits through a pathological or medical lens. Most neurodiversity advocates distance themselves from more traditional critical perspectives that reject "the language of disorder" (Kinderman et al., 2013; see also Szasz, 1961) as an *intrinsically* harmful conceptualization of mental divergence. These critiques adopt a *normalizing* discourse that, in line with CBT's continuity thesis, reconceptualizes mental conditions as *normal,* typical deviations from social standards that are expectable given certain difficulties in living. This narrative can be sometimes beneficial, sure; but its underlying neuronormative assumption, that there is one fundamental mode of normal cognitive functioning, systematically neglects the distinct experience, functioning, and perspectives of neurodivergent individuals—sometimes to the point of caricaturing the neurodiversity paradigm as the tragic fantasy of "naïve victims who have 'been sold [a] lie'" (Chapman, 2023a, p. 113),

or worse, as a "Trojan horse" at the service of Big Pharma (Cromby & Johnstone, 2024).

The Neurodiversity movement condemns both the wrongful pathologizing and the wrongful normalizing of cognitive divergence as two sides of the same oppressive assumption: that neuronormative standards reflect natural or universal facts about human cognition, and therefore that their contestation is little more than unscientific opinion, if not moral perfidy. Neurodiversity activism promotes instead a person-centered and context-sensitive approach to decision-making and knowledge production, centering neurodivergents' standpoints and experiential knowledge in discussions about what should be treated or pathologized (Catala et al., 2021; Chapman & Carel, 2022; Legault et al., 2024; Walker, 2021).

15.3 Embodied, embedded, enactive, extended… and exclusionary?

Against the traditional cognitivist framework, recent E-Cognition approaches to mental health characterize mental health conditions as phenomena that cannot be understood solely by looking inside individuals' brains. Instead, E-Mental Health promotes a multilevel and relational view that takes the organism-environment system as the fundamental unit of analysis; determining an agent's mental status requires analyzing their dynamics of interaction with the environment from various scales of analysis (e.g., de Haan, 2020; Gallagher, 2024; Hoffman, 2016; Maiese, 2022; Nielsen, 2023; Roberts et al., 2019). This holistic perspective has reinvigorated the old, yet often under-implemented biopsychosocial model of health (Aftab & Nielsen, 2021), contributing to develop a much-needed middle ground between traditional neuro-reductionist medical accounts—which characterize mental pathology as mere internal, brain-based dysfunction—and traditional antipsychiatry—which, in its rejection of neuroreductionism, undermines any meaningful distinction between psychopathology and social deviance, thereby throwing the baby out with the bathwater.

We may distinguish two different developments of the E-Mental Health framework (Núñez de Prado-Gordillo & López-Silva, 2025). *Extended* views often implement a traditional functionalist, representational, and computational understanding of mind (e.g., Bray, 2008; Sprevak, 2011). On this view, mental health and disorder are still defined in terms of function and dysfunction; however, the functional abilities under scope may extend beyond the brain, incorporating elements from the environment. *Embodied enactivist* views, by contrast, reject functionalism in favor of enactivism (e.g., de Haan, 2020; Maiese, 2022; Nielsen, 2023). On this view, living, autonomous systems must constantly strive to maintain a balance between *self-maintenance* and *adaptivity* to an ever-changing environment. This precarious equilibrium gives rise to *sense-making*, i.e., the agent's intrinsically affective and normative construal of their environment and possibilities for action as valuable

or detrimental to their self-maintenance and adaptivity (Di Paolo et al., 2017). From biology to culture, through the psychological realm of *habits* and *regional identities*, these scales bring forth their own normativities; that is, their own norms determining what forms of sense-making patterns are valuable or detrimental. Mental health and disorder are defined in terms of an agent's *meta-flexibility* and *meta-stability* (García & Barandiaran, 2025), i.e., their ability to coordinate these different and sometimes conflicting normativities in a way that optimally preserves the agent's identity while at the same time allowing them to flexibly adapt to the environment (see also Gallagher, 2024).

Despite these differences, both share the non-reductionist idea that mental (dis)abilities cannot be identified with internal (dys)functions in brain-bound information-processing mechanics. This core assumption aligns closely with the Neurodiversity movement. Indeed, the affinities between E-Mental Health and neurodiversity have been actively explored by many scholars, including prominent neurodiversity advocates (Catala et al., 2021; Chapman, 2021; Hoffman, 2019; Jurgens, 2020; Legault et al., 2024; Roberts et al., 2019). First, E-Mental Health views share the relational analysis of mental (dis)ability, often exemplified by the so-called *double empathy* conception of autism (Jurgens, 2020; Krueger & Maiese, 2018; Roberts et al., 2019; van Grunsven, 2020). First introduced by neurodiversity scholar Damian Milton (2012), this conception opposes traditional cognitivist characterizations of autism as involving an inner deficit in social cognition (e.g., Baron-Cohen et al., 1985). Instead, Milton proposed that difficulties in socio-cognitive interpretation are experienced not only by (some) autistic individuals, but also by allistic ones when immersed in social environments predominantly shaped by allistic and autistic norms and customs, respectively. E-Mental Health explains this double empathy problem as the result of a mismatch between individual cognitive traits and the *mental institutions*—the norms, practices, and conventions—that predominate in a given social context (Krueger & Maiese, 2018; Roberts et al., 2019).

Furthermore, in line with neurodiversity advocates' emphasis on the need for co-productive strategies in mental healthcare, enactive approaches promote a participatory and person-centered approach to such relational analysis (de Haan, 2020; Gallagher, 2024; García & Barandiaran, 2025; Nielsen, 2023). On traditional and extended functionalist accounts, the mental abilities that configure mental health—whether implemented solely in the brain or in the brain-environment coupling—are given by abstract, universal computational requirements, defined independently of specific agents' histories. By contrast, enactivists assume that the norms relevant for mental health are the ones individual agents, with their specific embodiment and developmental histories, *enact* or bring forth in their interaction with their environment. Distinguishing mental health and disorder requires paying attention to individuals' perspectives and values, emerging from their specific way of navigating the socio-material world. This approach also grounds the

vindication of *experiential knowledge*, based on lived experience of a mental condition, as an irreplaceable epistemic good in mental healthcare (Dings & Tekin, 2023).

Finally, another important complementarity between E-Mental Health and the Neurodiversity movement, related to the latter's ecological model, concerns the development of the notion of *extended* or *collective* dysfunctions: against the traditional definition of mental disorder in terms of individual dysfunctions (e.g., American Psychiatric Association, 2022, p. 15), social dynamics may also count as functional or dysfunctional depending on how they impact group adaptability (Cooper, 2017; Dengsø, 2024; Hoffman, 2016, 2017; Jurgens, 2023; Roberts et al., 2019). On this view, oppressive social dynamics that systematically undermine a population's diversity—including neurodiversity—could be conceptualized as extended or collective dysfunctions that disrupt the collective's ability to cope with a broader variety of environmental challenges.

Despite these strong complementarities, recent criticisms have pointed out actual or potential neuronormative biases in E-Mental Health theories. Russell (2023) notes that the use of notions like "(dys)function" in enactive accounts leaves much room for interpretation, potentially preserving or reinforcing stigmatizing attitudes. Others point out how E-theorists often actively rely on undue pathologizing views of neurodivergent traits or assume neurotypical understandings of core capacities considered crucial for mindedness (Jurgens, 2023; Parra Rubio, 2025; van Grunsven, 2020). Van Grunsven (2020) illustrates this with Gallagher's interactionist view of autism, which, despite emphasizing second-person interactions to characterize social cognition, switches to third-person explanatory accounts when discussing autism, merely giving "sketches of [autistic] behaviors just insofar as they deviate from the neurotypical norm. The autistic perspective itself never really announces itself" (p. 9).

Another example is provided by Bray's (2008) extended account of borderline personality disorder (BPD), here characterized as an extension of executive metacognitive functions to other people's brains—which would allegedly explain why people with BPD have difficulties in coping with others' rejection. However, the default pathologizing of BPD is left unquestioned: such extension of metacognitive functions is not understood as an expression of an alternative mode of functioning—akin to how, for instance, digital natives extend other executive functions to their smartphones in ways non-natives do not; rather, Bray sees it as the necessary result of brain deficits hindering the *normal*, internal implementation of executive metacognitive processes, even in the absence of independent evidence for such deficit interpretation. In this vein, Jurgens (2023) has argued that although E-Cognition views dispute the internalist assumptions of traditional cognitivism, they still proceed on the unquestioned and unfounded assumption that it is neurodivergent cognitive styles that need correction and accommodation to neuronormative standards of proper cognitive functioning.

On the one hand, this is hardly surprising: theorists, no matter how well-intentioned, are equally immersed in the social structures that sustain and reinforce oppressive ideologies. But it is still somewhat striking that *these* theorists in particular are sometimes just as prone to ableist and neuronormative biases as their cognitivist counterparts. For one, they are far more aware of their own situatedness as knowledge producers; more importantly, their own theoretical frameworks provide the very conceptual tools needed to radically break with the internalist, individualistic assumptions that underpin the ideological structure of the Empire of Normality—the structure that enforces ableism and neuronormativity in the first place. How is it, then, that E-Mental Health so often falls so short of its own liberatory potential?

15.4 Descriptivism and (unwarranted) epistemic symmetry

As we view it, notwithstanding other sociocultural and material influences, this apparent difficulty of some E-Mental Health thinkers to break through the lenses of ableism and neuronormativity stems from an often implicit, residual commitment to the Cartesian theory of mind: what we may call *mental descriptivism*, the thesis that the interpretation of an agent in mental terms is a practice primarily geared toward the description of facts about agents' activities or their causes (Fernández Castro, 2023, 2024; Heras-Escribano et al., 2015; Heras-Escribano & Pinedo-García, 2018; Pinedo García, 2020).[1] This descriptivist commitment is what we take Ryle (1949/2009) to refer to as "the logical mould into which Descartes pressed his theory of the mind", which represents "differences between the physical and the mental […] as differences inside the common framework of the categories of 'thing', 'stuff', 'attribute', 'state', 'process', 'change', 'cause' and 'effect'. Minds are things, but different sorts of things from bodies" (p. 9). Note that this is a more fundamental and subtler commitment than dualism, internalism, or mechanicism: even if we reject these, conceptualizing minds in the language of objects, properties, and relations keeps philosophy of mind within the bounds of Cartesian coordinates (Pinedo García, 2020; Rorty, 1979).

The main consequence this position poses for the recognition of cognitive diversity is that it represents mindedness, and mental health in particular, as a primarily factual matter. Debates about what a functioning mind is can be, ultimately, settled *empirically*: empirically informed models of the mind thus set the authoritative criteria to resolve such disagreements. They establish a

1 In fact, we don't think this commitment is independent from the same social and material conditions that Chapman (2023) ties to the emergence of Descartes' mechanical view of mind and nature, which eventually gave rise to the Empire of Normality: that is, the historical antecedents of industrial capitalism and its ever-growing need to monitor productive ability. Descriptivism, after all, underpins a conception of minds as *assets*, as goods and resources to be predicted, controlled, and managed (Pinedo, 2020).

cognitive archetype of normal, healthy functional ability, which then serves as a standard for assessing individual cognitive profiles. Assessments of proper cognitive functioning are, in this sense, descriptive—something that can be neutrally represented in aperspectival terms, based on the degree of archetype-profile fitness.

We believe this view contributes to naturalizing certain epistemic assumptions that neurodiversity scholars criticize as the foundation of their stigmatization and systematic exclusion from knowledge production about their own minds: what we may call the *epistemic (a)symmetry* assumption (see Catala, 2023; Catala et al., 2021; Chapman & Carel, 2022; Legault et al., 2024). Focusing on autism, Legault et al. (2024) describe this as "the presupposition that allistics are as well positioned as Autistics to speak about autism, perhaps even better positioned, prompting allistics' refusal to acknowledge the epistemic limits of their own positionality" (pp. 83–84). Importantly, although Legault et al. (2024) talk about allistics in general, their target examples concern neuronormative scientific *experts*—specifically, *experts-by-training*, as opposed to those with *expertise-by-experience* (Dings & Tekin, 2023). Note however that, from a descriptivist standpoint, it is hard to see why the epistemic (a)symmetry assumption is problematic: if what is a properly functioning mind is a matter of scientific discovery and description, scientists, or empirically-informed experts-by-training more broadly, will naturally be in at least a symmetrical, if not *privileged* epistemic position in disagreements about what sort of cognitive abilities configure a (healthy) mind. It's not just that scientific experts are a particularly well-informed contending party; rather, they are the closest thing to the *tribunal* that may adjudicate who's right and wrong in such disagreements. Any substantial challenge to how the mind is portrayed in expert opinion should come from, or at least be endorsed by, expert opinion.

This constellation of commitments is characteristic of traditional cognitivism, which, as we saw, assumes that normal information-processing sets the benchmark of cognitive health. But it is, we think, equally present in most E-Cognition views. This is particularly clear in the case of extended functionalist accounts, which merely *extend* the possible realizers of information-processing dynamics. Enactive proposals do emphasize the need to consider agents' specific histories and perspectives—which various scholars see as akin to a neurodiversity-affirming framework (Legault et al., 2024; see García & Barandiaran, 2025; Maiese, 2024). However, at least in the case of autopoietic or autonomy enactivists, this apparently evaluative and perspectival view of mindedness ultimately defaults to a descriptivist account of mind within their broader project of naturalizing normativity (e.g., Di Paolo et al., 2017; see Heras-Escribano & Pinedo, 2018; Heras-Escribano et al., 2015; Pinedo, 2020). The norms governing appropriate mindful organism-environment interaction are, on this view, *natural* ones, resulting from the viability conditions imposed by the autonomy and adaptability requirements of the multiple, hierarchically nested operational loops that configure a living system (García & Barandiaran,

2025). On this view, what is to be properly, healthily minded is still a factual matter, something to be discovered via scientific analysis. This may require setting not just one single cognitive archetype, as traditional cognitivists think, but multiple ones, each set by the normativities emerging at various embodied domains, from metabolism to interpersonal relations. However, once we decide which domain to look at, what sorts of mindedness are valid, what specific abilities are required for a particular creature to count as (healthily) minded is, again, a matter of empirical discovery.

Now, one could ask: which domain *should* we look at in the first place? Should we focus on organismic self-maintenance and adaptivity (Nielsen, 2023)? Should we rather look at the individual's specific habits or *regional identities* (Maiese, 2022), or even at the social level, considering the self-maintenance and adaptivity prospects of the individual's social ecosystem (Dengsø, 2024)? Or should we rather view mental health as some sort of precarious balance of meta-stability and meta-flexibility across *all* domains (Gallagher, 2024; García & Barandiaran, 2025)?

What would exactly amount to a sufficient balance of meta-stability and meta-flexibility—or how to *weight* different and potentially contradicting normativities emerging from different scales of analysis—is somewhat difficult to conceive. But what we want to highlight here is that, whatever answer we give, it will be intrinsically *open* to contestation and renegotiation. This openness signals the deeply *evaluative* nature of the sort of question at hand. As in Moore's (1903) classic open-question argument against the naturalist fallacy,[2] no matter which factual properties one associates with what is good or correct, one may always ask: but is it *really* good? Likewise, no matter how multi-scalar and context-dependent the framework under which we wish to describe the relevant facts, no matter what specific assortment of capacities our preferred theories of cognition set as crucial for healthy cognition, one may always ask: "but is this *really* a healthy mind?" Take again the meta-balance of stability-flexibility as an example: even if the emphasis on this sort of capacity reflected deep-seated, prevailing normative views of what counts as properly minded, the question remains: are these views correct? Rather than a *natural* benchmark of mental health/disorder, may not the meta-balance criterion merely reflect the "evermore restrictive cognitive norms of the age" (p. 117), i.e., the *exact kind of double bind* one would expect from production dynamics that increasingly necessitate of simultaneously predictable and resilient (i.e., stable) as well as malleable and versatile (i.e., flexible) labor?

2 In this line, Heras-Escribano and Pinedo (2018) understand the enactive naturalization of normativity as a form of naturalistic fallacy—not in the realm of morality (Moore, 1903) or knowledge (Sellars, 1956), but cognition. Our concerns thus are a specific application of this general criticism to E-Mental Health.

15.5 The evaluative concept of mind

Our diagnosis of the problems with E-Mental Health is the opposite of Russell's (2023): the issue is not that they underdetermine the criteria for demarcating proper from improper minds, but precisely that they overdetermine them. What is needed is a conceptual framework that accommodates the open, *evaluative* nature of disagreements about an agent's mental constitution.

The alternative we want to propose is *mental expressivism*, i.e., the view that the primary function of mental concepts or the speech acts that use them is to express some sort of evaluative attitude or commitment. Here, we include expressivist views of ascriptions of folk-psychological attitudes (Fernández Castro, 2023, 2024; Pérez-Navarro et al., 2019) and related ones, such as ascriptions of self-knowledge (Villanueva, 2014), rationality (Frápolli & Villanueva, 2018; Gibbard, 1990), or mental pathology (Núñez de Prado-Gordillo, 2024; Wilkinson, 2020). Contemporary close friends of this view can also be found in non-factualist analyses of mind and agency (Heras-Escribano et al., 2015; Heras-Escribano & Pinedo-García, 2018; Pinedo García, 2020), as well as frameworks that emphasize the evaluative and regulative nature of mental language, including affective fictionalism (Demeter, 2022) and certain readings of *mindshaping* accounts of folk-psychology (Fernández Castro, 2020; McGeer, 2021; Strijbos & De Bruin, 2025; see Zawidzki & Tison, 2025).

Mental expressivism can be understood in terms of a negative tenet and a positive one (Fernández Castro, 2024). The negative one is the *non-descriptivist* thesis that mental language does not serve to describe or represent states of affair. The conceptual antecedents of this thesis go back to mid-20th-century analytic philosophy. It is nicely captured by Ryle's (1949) insistence that "the phrase 'there occur mental processes' does not mean the same sort of thing as 'there occur physical processes'" (pp. 11–12); Wittgenstein's argument against the possibility of a private language—which shows not that mental states are "a nothing", but "that a nothing would serve just as well as a something about which nothing could be said" (1953 Section 304); or Sellars' (1956) application of Moore's (1903) reasoning to the epistemic domain, which characterizes "the idea that epistemic facts can be analyzed without remainder—even in principle—into non-epistemic facts [as] a radical mistake—a mistake of a piece with the so-called 'naturalistic fallacy' in ethics" (p. 257) (see also Heras-Escribano & Pinedo-García, 2018; Tanney, 2009). Dennett's (1969) analysis of mental vocabulary as "non-referential" (p. 18), Davidson's (1991) emphasis on the disanalogy between psychological and physicalist vocabulary, or Rorty's (1979) critique of representationalism about both mind and language constitute other important antecedents of this idea in analytic philosophy of mind.

Rather than describing facts about an agent, mental interpretation is primarily geared toward the evaluation of each other's actions, reasoning, and

experiences in terms of their conformity to myriad norms structuring our social interactions, including norms of intelligibility, rationality, or morality. This is the positive thesis, *evaluativism*. When we attribute mental states like the belief that neurodiversity merits respect, the delusion that one's partner is an impostor, or, more generally, when we attribute mindedness to a creature, we are not primarily tracking facts about individuals considered as objects of prediction and control. Rather, we are adopting an intrinsically evaluative and ethical attitude toward them (Ramberg, 2000; see Pinedo García, 2020). Specifically, we are expressing a commitment to viewing them as effective *rule-followers*, i.e., as agents capable of responsibly responding to normative concerns and expectations, treating them and their behavior as not only falling under certain norms, but as effectively complying (or not) with them.

This distinction between *describing* an agent's activity as falling under certain norms and *evaluating* them as effectively complying with them is well-captured by Heras-Escribano and Pinedo's (2018) Rylean distinction between *conditions of satisfaction*—which determine whether some goal-directed behavior *successfully* meets some norm or standard—and *correctness criteria*—which determine whether the agent's success in meeting the norm is the result of mere accident, reflex, or brute force, or an actual expression of their intelligent and responsible *application* of the norm; that is, an instance of actual *rule-following*, to put it in Wittgenstein's (1953) terms. As Ryle (1949/2009) puts it, "To be intelligent is not merely to satisfy certain criteria, but to apply them; to regulate one's actions and not merely to be well-regulated" (pp. 28–29).

This helps to capture the main difference between the sort of evaluativist approach we are defending here and normativist, yet still descriptivist accounts, such as enactivism (Di Paolo et al., 2017) or classical interpretivism (Davidson, 2001; Dennett, 1987). Especially in Dennett's variant, interpretivists still assume that mental interpretation is primarily a descriptive practice, oriented toward causal explanation and prediction. Sure, these views grant a moral dimension to mental interpretation, emphasizing its link with assessments of agents as intelligent, intelligible, responsible, and therefore subject to normative considerations. Our view, by contrast, is that the moral dimension of mental interpretation is not a mere byproduct of the role it *happens* to play in our society—as if we could separately conceive the (allegedly) descriptive practice of mental interpretation from the evaluative practice of assessing each other in morally loaded terms. Rather, this moral dimension is a *constitutive* aspect of mental interpretation: in assessing each other in mental terms, we are assessing each other as intelligible, intelligent, responsible *rule-followers*, and thus as potential subjects of reactive attitudes.

Finally, this is what we think underpins the *regulative* function of mental interpretation emphasized by mindshaping approaches (Fernández Castro, 2020). Precisely because mental interpretation expresses a commitment to treating an agent's activity as the product of intelligent and responsible

rule-following, it sets up a series of normative expectations about how that activity ought to unfold. These normative expectations exert normative pressure on ascribees to offer some intelligible response to them, whether this entails conforming to the expected trajectories, offering excuses for norm-deviances, or—crucially for our purposes here—contesting the normative expectations in the first place (see McGeer, 2019).

15.6 Toward an expressivist, non-exclusionary E-Cognition

How can this expressivist framework help advance toward a more inclusive E-Mental Health for neurodiversity? In general terms, we think that its main contribution is that it offers a way to fully liberate the epistemically and politically empowering potential of E-Cognition frameworks from the self-defeating grip of descriptivism. As Pinedo García (2020) puts it:

> (…) enactivism, ecological psychology and the non-descriptivist and antirepresentationalist strands of analytic philosophy share a commitment with the need to start with the vocabulary of agency, of meaning and value, if we are to do justice to life and cognition. But (…) what gives philosophical and ethical edge to this commitment may be lost if we present ourselves as pursuing a project that competes with standard, non-normative forms of approaching nature, as merely offering richer redescriptions of a world that others may try to describe with the vocabulary of the physical sciences. The reduction of living beings to things, of value to price, of the sphere of normative negotiation amongst legitimate options regarding how to live to a technocratic calculus of benefits, are all equally threatening consequences of blurring the distinction between evaluation and description.
>
> (Pinedo García, 2020, p. 8)

Specifically, this recognition of the irreducibly evaluative nature of mental interpretation brings, to our view, three main inter-related benefits for the recognition of cognitive diversity: (i) the recognition of the intrinsically open nature of debates about neurodivergents' cognitive makeup, health, and flourishing; (ii) the recalibration of existing epistemic (a)symmetries that systematically downplay neurodivergents' contributions to such debates; and (iii) the potential avoidance of elite capture of the Neurodiversity movement.

These three contributions stem from the expressivist reinterpretation of debates about an agent's mental makeup—and, relatedly, mental health, and flourishing—as primarily evaluative rather than factual. As we saw in Section 15.3, treating these debates as factual commits us to the idea that they should be resolved by appeal to empirical evidence, to facts about an agent's internal or extended computational machinery, about the allegedly natural norms underpinning life itself, etc. But this doesn't seem to be the sort of dynamic governing the most pressing disagreements about

neurodiversity. Of course, neurodiversity activists challenge the evidential status of claims concerning neurodivergent traits and dispositions or their default pathologization/normalization. Typically, however, this is not their primary concern. For instance, autistic activists don't deny that autistic individuals often engage in repetitive patterns of self-stimulatory behavior (i.e., stimming) in stressful or emotionally challenging circumstances; that they often disregard implicit social codes of conduct when addressing their criticisms at perceived moral failures; or their tendency to hypersystemizing. What they reject is the default *evaluative interpretation* of these traits as "bizarre", "meaningless", "suboptimal", "deficient", or "dysfunctional". Instead, they reclaim their value, meaning, and overall *positive* status as instances of cognitive functions or abilities, e.g., as an ability to self-regulate in the case of stimming (Legault et al., 2024); to transmit information efficiently (Crompton et al., 2020) and stick to one's moral standards in the face of social pressures (Chapman, 2021); or to display "excellent attention to detail" (Baron-Cohen et al., 2009).

What these examples show is that the language of *mental functions* and *abilities* with which standard models of mental health operate is primarily evaluative rather than descriptive: to frame a certain disposition as an instance of cognitive function or ability, rather than a mere disposition or reflex, expresses an endorsement of performances so characterized (Fernández-Castro & Núñez de Prado-Gordillo, forthcoming; Heras-Escribano et al., 2015). Recognizing this does not mean renouncing to the possibility of a genuinely scientific mental health science—a consequence that only those committed to the neutrality or aperspectivality of scientific practice will read into our account (see Toole, 2022). It only implies recognizing that our moral and political *perspectives* lie at the core of scientific practice, and that what is primarily at issue in debates about the mental makeup and health status of neurodivergent individuals is what perspective should be endorsed. Again, this is an intrinsically open issue—one that cannot be foreclosed by merely "offering richer redescriptions of a world that others may try to describe with the vocabulary of the physical sciences" (Pinedo, 2020, p. 8).

The first contribution of our expressivist framework is precisely that it makes this openness evident. In doing so, it helps us to remain aware of the deeply and inescapably political basis of theories of mind and mental health. It promotes an E-Mental Health more self-aware of its positionality and perspectivality, more open to *unknow* and transform its own fundamental theoretical assumptions about the mind in ways that make room for neurodivergent perspectives; a self-critical stance that can be instrumental in resisting the tendency to naturalize neuronormative ideological assumptions as fundamental facts about human nature.

Second, this is closely related to the issue of epistemic asymmetry. As we saw in Section 15.3, treating the debates raised by neurodiversity activists as factual makes it natural to grant scientists and other experts-by-training—those typically seen as most acquainted with the relevant facts—a privileged

epistemic standpoint. But this naturalizes the very epistemic assumptions that neurodiversity scholars identify as the underlying force behind their epistemic oppression. By contrast, understanding these debates as evaluative and inherently open to contestation helps recalibrate the epistemic relationship between experts-by-training and experts-by-experience, clarifying how epistemic asymmetries should be redistributed depending on the issue at hand. Especially when it comes to fundamental questions about the alleged boundaries of healthy mindedness—or about research and treatment priorities—those whose mental makeup is under scrutiny should occupy a privileged epistemic position.

Finally, this relates to the need to avoid elite capture—that is, the process by which the interests of marginalized groups are co-opted or represented primarily by their most privileged members (Táíwò, 2022). If experts-by-training are positioned as holding epistemic authority, then only neurodivergent individuals with expertise-by-training themselves will count as capable of fully participating in decision-making and knowledge production about their own minds. Mere deference to neurodivergent scholars, however, will almost certainly leave crucial decisions in the hands of a privileged minority—if only because access to academic and scientific discourse remains accessible only to a few. Reconceptualizing these debates as primarily evaluative contributes to opening them to all neurodivergent individuals, whether their social position or cognitive makeup enables them to fully partake in current academic conversation or not. Indeed, it reinforces the idea that these debates should be shaped by neurodivergent *collectives*, and that their contributions in non-academic spaces—such as those that gave rise to the very concept of neurodiversity in the first place (Botha et al., 2024)—are just as significant, if not more so, than those emerging from academic institutions. No academic saviors or messiahs are necessary or sufficient for constructing a more just and inclusive conception of mind.[3]

15.7 Conclusion

Despite the numerous conceptual benefits brought by E-Mental Health for the recognition of cognitive diversity, a residual commitment to the Cartesian mind—mental descriptivism—forecloses the full liberatory potential of these approaches. By factualizing mindedness, mental descriptivism forecloses the possibility of contesting what is an intrinsically contestable, open matter, thereby potentially reifying hegemonic neuronormative assumptions about the mind and contributing to the perpetuation of the very social dynamics that give rise to them. In knowledge production systems, this reinforces the self-perceived privileged epistemic status of experts-by-training, conditioning the consideration of neurodivergents' perspectives on their own minds

3 We thank Neftalí Villanueva for drawing our attention to this possible application of the expressivist framework.

and mental flourishing to the approval of a pretendedly neutral scientific standpoint. We may say that E-Mental Health philosophers have hitherto only reinterpreted the mind in various ways, some in line with the Neurodiversity movement; the point, however, is to change it—and to uplift the epistemic status of neurodivergent people in decisions about how to do so. Mental expressivism, which proposes to reject mental descriptivism in favor of an evaluativist account of our mindmaking practices, can help liberate the transformative potential of E-Mental Health.

Acknowledgments

This work is supported by research grant HRZZ-IP-2022-10-1788, funded by the Croatian Science Foundation; by research grants PID2021-126826NA-I00 and PID2024-155717NB-I00, funded by the Spanish Ministry of Science and FEDER, UE; and by a Juan de la Cierva grant JDC2023-052608-I, funded by MCIU/AEI/10.13039/501100011033 and the FSE+.

References

Adler-Bolton, B., & Vierkant, A. (2022). *Health Communism: A Surplus Manifesto*. Verso Books. https://www.versobooks.com/products/2801-health-communism

Aftab, A., & Nielsen, K. (2021). From engel to enactivism: Contextualizing the biopsychosocial model. *European Journal of Analytic Philosophy*, *17*(2), 17_2_3. https://doi.org/10.31820/ejap.17.2.3

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders:* DSM-5-TRTM (Revised 5th edition). American Psychiatric Publishing, Inc.

Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., & Chakrabarti, B. (2009). Talent in autism: Hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1522), 1377–1383. https://doi.org/10.1098/rstb.2008.0337

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Beck, A. T. (Ed.). (1979). *Cognitive Therapy of Depression*. Guilford Press.

Bentall, R. P. (2003). *Madness Explained: Psychosis and Human Nature* (New Ed edition). Penguin.

Bordonaba-Plou, D., Fernández-Castro, V., & Torices-Vidal, J. R. (2022). *The Political Turn in Analytic Philosophy: Reflections on Social Injustice and Oppression*. De Gruyter.

Botha, M., Chapman, R., Giwa Onaiwu, M., Kapp, S. K., Stannard Ashley, A., & Walker, N. (2024). The neurodiversity concept was developed collectively: An overdue correction on the origins of neurodiversity theory. *Autism*, *28*(6), 1591–1594. https://doi.org/10.1177/13623613241237871

Bray, A. (2008). The extended mind and borderline personality disorder. *Australasian Psychiatry*, *16*(1), 8–12. https://doi.org/10.1080/10398560701636930

Catala, A. (2023). Epistemic injustice and epistemic authority on autism. In S. L. Tremain (Ed.), *The Bloomsbury Guide to Philosophy of Disability* (pp. 246–267). Bloomsbury Academic. https://www.bloomsburycollections.com/collections/monograph-detail

Catala, A., Faucher, L., & Poirier, P. (2021). Autism, epistemic injustice, and epistemic disablement: A relational account of epistemic agency. *Synthese*, *199*(3–4), 9013–9039. https://doi.org/10.1007/s11229-021-03192-7

Chapman, R. (2021). Neurodiversity and the social ecology of mental functions. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *16*(6), 1360–1372. https://doi.org/10.1177/1745691620959833

Chapman, R. (2023a). A Critique of critical psychiatry. *Philosophy, Psychiatry, & Psychology*, *30*(2), 103–119.

Chapman, R. (2023b). *Empire of Normality: Neurodiversity and Capitalism* (1st edition). Pluto Press.

Chapman, R., & Carel, H. (2022). Neurodiversity, epistemic injustice, and the good human life. *Journal of Social Philosophy*, josp.12456. https://doi.org/10.1111/josp.12456

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology*, *62*(1), 271–298. https://doi.org/10.1146/annurev.psych.121208.131622

Cooper, R. (2017). Where's the problem? Considering Laing and Esterson's account of schizophrenia, social models of disability, and extended mental disorder. *Theoretical Medicine and Bioethics*, *38*(4), 295–305. https://doi.org/10.1007/s11017-017-9413-0

Cromby, J., & Johnstone, L. (2024, August 5). Part 4: Neurodiversity: New Paradigm, or Trojan Horse? *Mad In America*. https://www.madinamerica.com/2024/08/part-4-neurodiversity-new-paradigm-or-trojan-horse/

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., & Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. *Autism*, *24*(7), 1704–1712. https://doi.org/10.1177/1362361320919286

David, A. S., & Halligan, P. W. (2000). Cognitive neuropsychiatry: Potential for progress. *Journal Neuropsychiatry Clinical Neuroscience*, *5*, 506–510.

Davidson, D. (1991). Three Varieties of Knowledge. *Royal Institute of Philosophy Supplements*, *30*, 153–166. https://doi.org/10.1017/S1358246100007748

Davidson, D. (2001). Thought and talk. In D. Davidson (Ed.), *Inquiries into Truth and Interpretation* (p. 0). Oxford University Press. https://doi.org/10.1093/0199246297.003.0011

de Haan, S. (2020). *Enactive Psychiatry* (1st edition). Cambridge University Press. https://doi.org/10.1017/9781108685214

Demeter, T. (2022). A mental fictionalism worthy of its name. In T. Demeter, T. Parent, & A. Toon (Eds.), *Mental Fictionalism* (pp. 86–103). Routledge.

Dengsø, M. J. (2024). Enactive psychiatry and social integration: Beyond dyadic interactions. *Phenomenology and the Cognitive Sciences*, 1–25. https://doi.org/10.1007/s11097-024-09957-y

Dennett, D. C. (1969). *Content and Consciousness*. Routledge.

Dennett, D. C. (1987). *The Intentional Stance* (pp. xi, 388). The MIT Press.

Di Paolo, E., Bhurman, T., & Barandiaran, X. (2017). *Sensorimotor Life: An Enactive Proposal*. Oxford University Press.

Dings, R., & Tekin, Ş. (2023). A philosophical exploration of experience-based expertise in mental health care. *Philosophical Psychology*, *36*(7), 1415–1434. https://doi.org/10.1080/09515089.2022.2132926

Douglas, V. I. (1972). Stop, look and listen: The problem of sustained attention and impulse control in hyperactive and normal children. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *4*(4), 259–282. https://doi.org/10.1037/h0082313

Fernández Castro, V. (2020). Regulation, normativity and folk psychology. *Topoi*, *39*(1), 57–67. https://doi.org/10.1007/s11245-017-9511-7

Fernández Castro, V. (2023). Folk psychology without metaphysics: An expressivist approach. *South African Journal of Philosophy*, *42*(2), 128–143. https://doi.org/10.1080/02580136.2023.2273122

Fernández Castro, V. (2024). An expressivist approach to folk psychological ascriptions. *Philosophical Explorations*, *27*(1), 86–105. https://doi.org/10.1080/13869795.2023.2251491

Fernández-Castro, V., & Núñez de Prado-Gordillo, M. (forthcoming). Neurodiversity: a challenge for ability-based views on mental health. In G. Leder, S. Dembić, F. Hundertmark, T. Zawidzki, & F. Jotterand (Eds.), *Psychiatric Disorders: Advances in Ability-Based Approaches* . Springer

Fodor, J. A. (1989). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. The MIT Press.

Frápolli, M. J., & Villanueva, N. (2018). *Minimal Expressivism and the Meaning of Practical Rationality*. Brill. https://doi.org/10.1163/9789004359475_002

Frith, C. D. (2003). *The Cognitive Neuropsychology of Schizophrenia* (Reprinted). Psychology Press.

Frith, U., & Happé, F. (1994). Autism: beyond "theory of mind". *Cognition*, *50*(1–3), 115–132. https://doi.org/10.1016/0010-0277(94)90024-8

Gallagher, S. (2024). *The Self and its Disorders*. Oxford University Press.

García, E., & Barandiaran, X. E. (2025). Varieties of normativity and mental health: An enactive approach. *Synthese*, *205*(2), 53. https://doi.org/10.1007/s11229-024-04854-y

Gibbard, A. (1990). *Wise Choices Apt Feelings—A Theory of Normative Judgement*. Harvard University Press.

Hänel, H. C., & Müller, J. M. (2022). Non-ideal philosophy as methodology. *Theoria. A Journal of Social and Political Theory*, *69*(172), 32–59. https://doi.org/10.3167/th.2022.6917202

Heras-Escribano, M., Noble, J., & de Pinedo, M. (2015). Enactivism, action and normativity: A Wittgensteinian analysis. *Adaptive Behavior*, *23*(1), 20–33. https://doi.org/10.1177/1059712314557364

Heras-Escribano, M., & Pinedo-García, M. de. (2018). Naturalism, non-factualism, and normative situated behaviour. *South African Journal of Philosophy*, *37*(1), 80–98. https://doi.org/10.1080/02580136.2017.1422633

Higashida, N. (2018). *Fall Down Seven Times Get Up Eight*. Sceptre.

Hoffman, G. A. (2016). Out of our skulls: How the extended mind thesis can extend psychiatry. *Philosophical Psychology*, *29*(8), 1160–1174. https://doi.org/10.1080/09515089.2016.1236369

Hoffman, G. A. (2017). Collectively ill: A preliminary case that groups can have psychiatric disorders. *Synthese*, *196*(6), 2217–2241. https://doi.org/10.1007/s11229-017-1379-y

Hoffman, G. A. (2019). Chapter Thirteen - Public mental health without the health? Challenges and contributions from the Mad Pride and neurodiversity paradigms. In K. Cratsley & J. Radden (Eds.), *Developments in Neuroethics and Bioethics* (Vol. 2, pp. 289–326). Academic Press. https://doi.org/10.1016/bs.dnb.2019.07.003

Jeppsson, S. (2022). Solving the self-illness ambiguity: The case for construction over discovery. *Philosophical Explorations*, *25*(3), 294–313. https://doi.org/10.1080/13869795.2022.2051589

Jurgens, A. (2020). Neurodiversity in a neurotypical world: An enactive framework for investigating autism and social institutions. In H. B. Rosqvist, N. Chown, & A. Stenning (Eds.), *Neurodiversity Studies* (pp. 73–88). Routledge.

Jurgens, A. (2023). Body social models of disability: Examining enactive and ecological approaches. *Frontiers in Psychology*, *14*, 1128772. https://doi.org/10.3389/fpsyg.2023.1128772

Keiser, J. (2022). *Non-Ideal Foundations of Language*. Routledge. https://doi.org/10.4324/9781003188537

Kinderman, P., Read, J., Moncrieff, J., & Bentall, R. P. (2013). Drop the language of disorder. *Evidence Based Mental Health*, *16*(1), 2–3. https://doi.org/10.1136/eb-2012-100987

Krueger, J., & Maiese, M. (2018). Mental institutions, habits of mind, and an extended approach to autism. *Thaumàzein*, *6*, 10–41.

Legault, M., Catala, A., & Poirier, P. (2024). Breaking the stigma around autism: Moving away from neuronormativity using epistemic justice and 4E cognition. *Synthese*, *204*(3), 84. https://doi.org/10.1007/s11229-024-04730-9

Maiese, M. (2022). *Autonomy, Enactivism, and Mental Disorder: A Philosophical Account* (1st edition). Routledge. https://doi.org/10.4324/9781003174103

Maiese, M. (2024). Problems for enactive psychiatry? Mindshaping, social normativity, and neurodiversity. *Philosophical Psychology,* 1–26. https://doi.org/10.1080/09515089.2024.2446631

McGeer, V. (2019). Mindshaping is inescapable, social injustice is not: Reflections on haslanger's critical social theory. *Australasian Philosophical Review*, *3*(1), 48–59. https://doi.org/10.1080/24740500.2019.1705231

McGeer, V. (2021). Enculturating folk psychologists. *Synthese*, *199*(1–2), 1039–1063. https://doi.org/10.1007/s11229-020-02760-7

Menary, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 459–463. https://doi.org/10.1007/s11097-010-9187-6

Milton, D. (2012). On the ontological status of autism: The 'double empathy problem.' *Disability & Society*, *27*(6), 883–887. https://doi.org/10.1080/09687599.2012.710008

Moore, G. E. (1903). *Principia ethica*. Cambridge University Press.

Newen, A., Bruin, L. D., & Gallagher, S. (Eds.). (2018). *The Oxford Handbook of 4E Cognition* (4th edition). Oxford University Press.

Nielsen, K. (2023). *Embodied, Embedded, and Enactive Psychopathology: Reimagining Mental Disorder*. Springer International Publishing. https://doi.org/10.1007/978-3-031-29164-7

Núñez de Prado-Gordillo, M. (2024). Broken wills and ill beliefs: Szaszianism, expressivism, and the doubly value-laden nature of mental disorder. *Synthese*, *203*(1), 24. https://doi.org/10.1007/s11229-023-04427-5

Núñez de Prado-Gordillo, M., & López-Silva, P. (2025). Making sense of the 4E cognition turn in mental health research. *Philosophy, Psychiatry, & Psychology, 32*(2), 131–150. https://doi.org/10.1353/ppp.2025.a964179

Parra Rubio, F. (2025). Intersubjective, systemic, and sensory roots of autistic social difficulties: A critical evaluation of enactivist and phenomenological approaches. *Synthese*, *205*(4), 153. https://doi.org/10.1007/s11229-025-04995-8

Pérez-Navarro, E., Fernández Castro, V., González de Prado Salas, J., & Heras–Escribano, M. (2019). Not expressivist enough: Normative disagreement about belief attribution. *Res Philosophica*, *96*(4), 409–430. https://doi.org/10.11612/resphil.1794

Pinedo García, M. (2020). Ecological psychology and enactivism: A normative way out from ontological dilemmas. *Frontiers in Psychology*, *11*, 1637. https://doi.org/10.3389/fpsyg.2020.01637

Ramberg, B. (2000). Post-ontological philosophy of mind: Rorty versus davidson. In R. Brandom (Ed.), *Rorty and His Critics* (pp. 9–351). Wiley-Blackwell.

Roberts, T., Krueger, J., & Glackin, S. (2019). Psychiatry beyond the brain: Externalism, mental health, and autistic spectrum disorder. *Philosophy, Psychiatry, & Psychology*, *26*(3), E-51-E-68. https://doi.org/10.1353/ppp.2019.0030

Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton University Press.

Rosqvist, H., Chown, N., & Stenning, A. (Eds.). (2020). *Neurodiversity Studies: A New Critical Paradigm*. Routledge. https://doi.org/10.4324/9780429322297

Russell, J. L. (2023). Problems for enactive psychiatry as a practical framework. *Philosophical Psychology*, *0*(0), 1–24. https://doi.org/10.1080/09515089.2023.2174423

Ryle, G. (1949/2009). *The Concept of Mind: 60th Anniversary Edition*. Routledge.

Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (Eds.), *Foundations of Science and the Concepts of Psychology and Psychoanalysis* (Vol. 1, pp. 253–329). University of Minnesota Press. https://hdl.handle.net/11299/184083

Sinclair, J. (1992). Bridging the gaps: An inside-out view of Autism (or, do you know what I don't know?). In E. Schopler & G. B. Mesibov (Eds.), *High-Functioning Individuals with Autism* (pp. 289–306). Springer US. https://doi.org/10.1007/978-1-4899-2456-8_14

Sprevak, M. (2011). Neural sufficiency, reductionism, and cognitive neuropsychiatry. *Philosophy, Psychiatry, & Psychology*, *18*(4), 339–344. https://doi.org/10.1353/ppp.2011.0057

Strijbos, D. W., & De Bruin, L. (2025). Mindshaping and neurodiversity: Challenges and opportunities. In T. W. Zawidzki & R. Tison (Eds.), *The Routledge Handbook of Mindshaping* (pp. 458–472). Routledge.

Szasz, T. S. (1961). *The Myth of Mental Illness: Foundations of a Theory of Personal Conduct* (Rev. ed., reprinted, 27. [print.]). Perennial.

Táíwò, O. O. (2022). *Elite Capture: How the Powerful Took Over Identity Politics*. Haymarket Books.

Tanney, J. (2009). Rethinking Ryle: A Critical Discussion of the Concept of Mind. In G. Ryle (Ed.), *The Concept of Mind: 60th Anniversary Edition* (pp. ix–lix). Routledge.

Toole, B. (2022). Demarginalizing standpoint epistemology. *Episteme*, *19*(1), 47–65. https://doi.org/10.1017/epi.2020.8

van Grunsven, J. (2020). Perceiving "Other" minds: Autism, 4E cognition, and the idea of neurodiversity. *Journal of Consciousness Studies*, *27*(7–8), 115–143.

Villanueva, N. (2014). Know thyself: A tale of two theses and two theories. *Teorema: Revista Internacional de Filosofía*, *33*(3), 49–66.

Walker, N. (2021). *Neuroqueer Heresies: Notes on the Neurodiversity Paradigm, Autistic Empowerment, and Postnormal Possibilities*. Autonomous Press.

Wilkinson, S. (2020). Expressivism about delusion attribution. *European Journal of Analytic Philosophy*, *16*(2), 59–77. https://doi.org/10.31820/ejap.16.2.3

Wittgenstein, L. (1953). *Philosophical Investigations*. Wiley-Blackwell.

Zawidzki, T. W., & Tison, R. (Eds.). (2025). *The Routledge Handbook of Mindshaping*. Routledge. https://doi.org/10.4324/9781032639239

16 Adaptive preferences and extended cognition

Gloria Andrada

16.1 Introduction

Imagine the following case: Celia, a 15-year-old girl, desires to undergo an aesthetic medical procedure as, for example, dermal fillers to increase skin volume or so-called "preventive" anti-wrinkle injections intended to stop wrinkles from forming. This has become a fairly common scenario in many parts of the world today. To provide some context, the 2023 plastic surgery statistics report shows that there were 5.3 million dermal filler procedures performed in the US, indicating a noticeable upward trend, even among younger individuals.[1]

Let's continue building this story. To get what she wants, Celia plans to use the money she has been saving, which means she won't be able to join the final school trip, and has to take a three-month break from her dance classes, an activity which she really enjoys.

The question that concerns me here is: why does Celia prefer to undergo such interventions as opposed to taking the other courses of action?

From the perspective of analytic feminist theory, a plausible explanation for this preference comes from the concept of *adaptive preferences*. Adaptive preferences have been said to be preferences that are formed in response to oppression, and which "serve to uphold or reinforce the oppressive social structures that are imposed on the agent as a result of their situation" (Knowles, 2021, p. 2) The idea is basically that individuals who live under conditions of social inequality shape their desires to fit oppressive contexts rather than resist them (Nussbaum, 2001; Superson and Brennan 2005).

In Celia's case, her preference for cosmetic enhancement may reflect the norms of sexist beauty standards, developed in response to persistent gendered pressures.[2] One plausible analysis of Celia's psychology is that she has shaped her preferences in light of aesthetic norms promoted by the current

1 https://www.plasticsurgery.org/documents/news/statistics/2023/plastic-surgery-statistics-report-2023.pdf [Last visited: May 15, 2025].

2 The extreme aesthetic pressure suffered by women is a long-treated topic in feminist theory, where some have characterized it as a form of self-objectification (see, for instance, Young

DOI: 10.4324/9781003649359-20
This chapter has been made available under a CC BY 4.0 license.

fashion and cosmetic industry. In so doing, she has turned away from values that would otherwise afford her a different, and potentially more empowering, sense of well-being.

Now, before moving on, I want to make two important clarifications. First, I will not take a stance in this chapter on the ethics behind cosmetic interventions and body modification. I do not intend to suggest that all of such cases result from adaptive preferences. Some cosmetic interventions may sometimes be carried out for reasons unrelated to oppression, and might in fact be liberating and empowering for some individuals. However, my focus here is on cases where the preference or choice is oppressive, in the sense in which "one's desires turn away from goods and even needs that, absent those conditions, they would want'"(Cudd, 2006, p. 181). And this is how I will interpret Celia's situation throughout this text.

Second, while the notion of adaptive preferences has been primarily developed within feminist theory to account for women's oppression in patriarchal societies, its scope is broader and can be applied to various social groups across different social contexts. The central concern is that individuals form preferences that sustain their oppression rather than promote their liberation. In other words, injustice permeates their desires. I will adopt here this more general perspective.

My goal is to offer an analysis of adaptive preferences through the lens of *extended cognition* (Clark, 2007; Sutton, 2010; Menary, 2007, 2010). As we will see, though the debate on adaptive preferences comes from discussions mostly within moral and political philosophy, in the literature, we do find references to the cognitive processes behind the development of such preferences. These cognitive accounts endorse an established *internalist* way of understanding cognition, where an individual's psychology is contained somewhere inside their head and can be distinguished from the environment, which is understood simply as the input for the formation of adaptive preferences.

In contrast, under an extended cognition view, human cognition is not something that happens within the boundaries of skin and skull, but it actually extends beyond an individual's brain to include external tools, environments, and social practices. As we will see, this entails that within this framework, adaptive preferences are not simply "in the head" but are dynamically co-constituted through and with extra-organismic structures.

In this chapter, I will show that this view is actually a very good fit for an analysis of adaptive preferences and can, in fact, help address some of the challenges within that debate. To do so, my plan is the following. I begin by looking into internalist cognition and current cognitive views of adaptive preferences (Section 16.2). I then present the main tenets of so-called second wave

(1990), Bartky (1990) Weiss (1999), Morgan (1991) or culturally imposed form of "feminine narcisissm" (Bartky, 1990)).

extended cognition, especially in the version which highlights cognitive enculturation as central, and I address adaptive preference formation from these lenses, focusing specially on the enculturation and extension of attention and memory (Section 16.3). I then examine the virtues of endorsing such a framework in key debates in adaptive preference intervention design (Section 16.4).

16.2 Internalist cognition and adaptive preferences

Traditional internalist cognitive science posits that mental processes, including preferences, are realized *internally,* within the boundaries of the brain. From this view, preferences are formed and modified through internal information-processing mechanisms, such as neuronal cognitive computations involving comparison, weighting, and information integration. They are understood as internal mental representations of values, desires, and goals, with the external environment playing no role beyond providing input for these internal operations. This includes the oppressive environment which explains the formation of adaptive preferences.

This internalist understanding of cognition also helps explain why the developmental process through which adaptive preferences are formed is often described as the "internalization" of oppression (e.g., oppressive social norms and expectations).[3] The very process of internalization reflects a conception of human cognition in which the mind is seen as located within a bounded internal world, contained by the physical body.

Although debate on adaptive preferences has its roots primarily in moral and political philosophy, the literature does contain some references to the cognitive mechanisms underlying its formation. As an example, consider the following three approaches.

1 **Value distortion**: In one of the first analysis of adaptive preferences, it is argued that individuals cognitively distort the value of goods or opportunities that are inaccessible to them due to their circumstances. They thus reduce psychological dissonance by lowering their value (Elster, 1983). This is usually called the "sour grapes" phenomenon, where individuals devalue the unattainable (e.g., they think, "I didn't want it anyway"). The cognitive mechanism in play is an internal form of self-deception where the person downplays the desirability of what they cannot attain.
2 **Misperception of interests and trade-offs**: In a more recent, and very influential analysis of adaptive preferences, Khader (2011) claims that some adaptive preferences are formed due to a misperception of interests. They are the result of a process in which a person's understanding of their needs is distorted, partially or globally. People who hold adaptive preferences often engage in trade-offs, accepting less of one good to attain more of

3 See Knowles 2021, p. 3.

another. In some cases, this internal cognitive operation is explained by the misperception of their true interests. For example, some women might harm their bodies in order to get social recognition, and in doing so, they believe in some sense that they are worthy of social recognition, but their understanding of their needs is distorted.

3 **Semantic shifts in evaluative reasoning**: More recently, it has been argued that adaptive preferences result from a semantic distortion that leads to faulty reasoning (Eftekhari, 2021). In this view, individuals may alter the meaning of key concepts (e.g., redefining what "respect", "well-being", or "freedom" means) to reflect an oppressive context. This redefinition enables rationalizations or justifications of their preferences that are cognitively flawed. In this way, adaptive preference formation is linked to changes in reasoning and reasoning errors.

These three approaches treat the formation of adaptive preferences as a failure of rationality or perception, caused by the harmful influence of external forces that have been internalized, subsequently, we might add, altering cognitive operations and representations. In the next section, we will see how viewing adaptive preferences through the lens of extended cognition can complement and enrich these perspectives.

16.3 Adaptive preferences: enculturated and extended

16.3.1 Extended cognition

Let us begin this section by saying something about the main tenets of an extended approach to human cognition. As we have just seen, one of the most established commitments of the standard image is that the mind (e.g., one's thoughts, desires, and memories.) is located somewhere in the head. This idea receives further support from cognitive neuroscience, according to which mental and cognitive processes are implemented only by the brain and the central nervous system. In this way, according to the default internalist position, mental processes take place somewhere *in* the brain and the central nervous system.

As part of the so-called 4e cognition framework (see Newen, De Bruin, & Gallagher, 2018), an extended cognition approach challenges this form of cognitive internalism. It does so by *extending* the cognitive realm so as to include not only the agent's body, but also some elements in their environment: mainly material culture (Clark & Chalmers 1998; Clark, 2008), but also cultural practices (Menary, 2007). This departure from the established image is not the result of mere philosophical speculation, or armchair reasoning. Rather, the extended cognition theory draws its philosophical conclusions from scientific insights about the nature of cognition, drawing on fields such as cognitive science, robotics, artificial intelligence, and biology.

We can distinguish two main lines of argument which have given rise to different agendas in extended cognition theorizing. The first of these emerges

from a particular way of understanding cognition, namely functionalism. The basic idea is that cognitive states and abilities are defined by the function they play within the overall cognitive system. It should not matter where that function is implemented (i.e., inside or outside the brain) or what it is made of (e.g., organic matter vs silicon). What matters is the functional contribution it makes to the cognitive behavioral structure of an agent.

But there is another way of thinking about extended cognition that does not directly rely on functionalism. This line of thinking, which, following John Sutton (2010), can be characterized as a *second-wave* approach to extended cognition, draws on a range of sources, from dynamical systems theory to evolutionary biology and niche construction theories, including cognitive anthropology.[4] This approach can outlined in three more specific claims, which I will characterize as follows:

1 ***Cognition is enculturated:*** Humans are able to acquire novel cognitive capacities because our capacities are augmented and transformed through the acquisition of cultural practices, i.e., patterns for action, which are transmitted both across members of a community and across generations. Cognitive enculturation can be defined as the process by which human cognitive abilities are altered and extended by social learning. Human minds are highly plastic and depend upon social learning and high-fidelity transmission to acquire knowledge, abilities, and develop and refine cognitive capacities. Individuals literally embody social practices by undergoing neural and bodily transformations in the process. These changes occur under an extended developmental history, enabling individuals to perform an array of tasks like reading, writing, and solving mathematical problems. Examples of neural transformations include the transformation of body schemas and the acquisition of different motor programs (Menary, 2007, 2018).
2 ***Cognition is extended***: Such capacities involve the skillful manipulation of culturally produced and maintained environmental resources, e.g., writing systems, number systems, other kinds of systems, and tools, with which humans interact in stable, consistent ways. These extra-organismic elements are essential elements of cognitive processes; therefore, cognition is extended beyond the organic boundary.
3 ***Cognition is integrated:*** Cognition is extended through the integration of cultural practices that guide interaction with the environment, along with the functionalities that technologies and devices afford. Cognition is

4 For the relation between Dynamical Systems Theory and extended cognition, see for instance, Menary (2007, pp. 42–48) and Palermos (2014). For the relation between niche construction and some versions of extended cognition, see Menary (2007). In relation to cognitive anthropology and extended cognition, see Sutton (2010). We find also reference to a third wave in Sutton 2010, which is further developed in Kirchhoff 2012.

> thus extended and enculturated through the integration within a cognitive system of an inherited cognitive capital which includes a structured socio-cultural cognitive niche and cognitive tools (Menary, 2007). This is afforded by a particular learning-driven neural flexibility/plasticity, and a bodily adaptability that allows individuals to acquire different motor programs (Fabry, 2018). Cognitive integration is produced and sustained by embodied (sensorimotor) engagements with the environment or with others via the environment. It is typically studied across a series of dimensions that help determine the degree of integration between organic processes, cultural practices, and external systems (Heersmink, 2014).

To sum up, the main idea against an internalist understanding of cognitive processes, including adaptive preference formation, is that these processes, rather than being confined within the boundaries of skin and skull, extend beyond the brain to include external tools and environments. This extension occurs through the acquisition of cultural practices that guide cognitive integration via sensorimotor engagement and embodied action. Within this framework, adaptive preferences are not simply "in the head" and the result of processes of "internalization", but are dynamically co-constituted through and with extra-organismic elements. From this point forward, I will refer to this view as an enculturated-extended approach to cognition.

Now, let me turn to the question: how can this framework contribute to our understanding of adaptive preference formation?

16.3.2 Enculturating and extending adaptive preferences

As we have just seen, within this understanding of human cognition, cognition is extended via the process of cognitive enculturation which allows for the integration of cultural and artifactual elements into an agent's cognitive capacities.

One of the first things that we need to notice is that this framework reveals that *preferences* are always in some sense *adaptive*, that is, they are always tuned to, constituted, and shaped within a specific social and cultural environment, since the boundaries between individual cognition and cultural environments are blurred. So, it makes sense that if the environment is oppressive, preferences would be formed in light of that oppression. The social group an individual is perceived to belong to (and self-perceives as belonging to), their specific developmental trajectory, individual differences, and the cultural and material environment they inhabit, all contribute to the the formation of specific preferences over others. In this sense, we could talk about *oppressive preference*s as the key phenomenon we are investigating here, since all preferences result from environmental adaptation. However, to avoid confusion, I will continue referring to these types of preference as adaptive preferences.

A key issue that then emerges is how to distinguish such oppressive preferences, from preferences *simpliciter*, since ultimately all preferences are shaped by the social and material environment in which individuals live. A common way of understanding adaptive preferences in the literature is that they cause harm by leading the subordinated agent to deviate from their "natural trajectory" (Knowles, 2021), which is typically understood as a deviation from a shared basic human flourishing (Nussbaum, 2001).

This is where we see one of the first contributions an enculturated-extended approach to human cognition can offer to the discussion on adaptive preferences: human flourishing, including forms of cognitive development, is always cultural relative and situated in specific contexts and embodiments. This insight contributes to important discussions around human's "natural trajectory" or "common basic flourishing", since the concept of "natural" trajectory is itself enculturated, and therefore open to cultural variation. Ultimately, cultural values provide a framework for evaluating desirability and appropriateness, and this includes flourishing and cultural practices. This warns us against a cultural form of imperialism, which is a well-known form of oppression (Young, 1990).

A key issue then becomes how oppression affects cognition such that adaptive preferences are formed. Before delving into the specific insights offered by an enculturated-extended approach, it is important to remember that the connection between cognition and oppression is not a new concern. For instance, Haslanger (2020) argues that cognition is shaped through socialization, and this process is permeated by power dynamics, and Maiese (2021) has focused on psychological ideological oppression. There is also a well-known lively debate on epistemic injustice, which connects social injustice with epistemic phenomena such as testimonial practices, and the interpretation of social experiences (Fricker 2007). Importantly, decolonial thinkers have long examined the psychological effects of oppression, as illustrated in the seminal works of Anzaldúa (1987) and Fanon (1961).

In this regard, we are engaging in a long-standing concern, since one of the effects of oppression on cognition is the development of adaptive preferences. As I will show, an enculturated-extended understanding of human cognition offers valuable insights into this process. To begin, it is helpful to examine the relevant cognitive abilities involved in the process of preference formation.

Preference formation is, in general, a complex process shaped by a dynamic interplay of different cognitive abilities such as attention. working memory, executive functions, perception, and introspection (Orquin & Mueller Loose, 1993; Hixon & Swann, 1993). At its core, preference formation involves the evaluation and comparison of options, often influenced by both deliberative and automatic processes.

Here, I won't review all of the abilities involved in this process, since addressing the enculturation and extension of each one of them and their role in adaptive preference formation would entail a much longer project. But I do want to review some of those cognitive abilities, and more importantly,

give a general blueprint that will hopefully guide future research. As I will then proceed to show in the next section, this blueprint also yields interesting conclusions for adaptive preference intervention design.

As we saw, a key feature of adaptive preferences that an enculturated-extended perspective sheds light on is not just the fact that oppression and social injustice permeate the enculturation of cognition, but that this takes place via the active role in cognition of structured environments, whole systems of material culture and cultural practices, not just through their internalization. This means that from a cognitive perspective, studying adaptive preferences requires not only identifying the mechanisms and abilities involved in their formation and maintenance, but also examining how oppressive circumstances reshape them, alongside the role of cultural factors that actively sustain and constitute those abilities. To do so, this approach gives us a structured roadmap: when investigating the cognitive abilities underlying adaptive preference formation, we must look not only at internal cognitive operations (which are grounded in neural and other embodied features), but also at the tools, artifacts, and cultural practices (i.e., patterns of action) with which individuals consistently interact and engage.

So, let's put this framework into action.

One key cognitive ability that has not received sufficient attention in the debate on adaptive preferences is attention itself. In fact, as we just saw, it is a fundamental capacity underlying the very formation of adaptive preferences.

Attention is composed among other sensory modalities of visual attention. Visual attention is the process by which the nervous system prioritizes certain locations, objects, or attributes within the visual scene. This can occur either through an eye movement that directs the object to the fovea, something known as overt attention, or by enhancing the neural processing of visual stimuli that appear in peripheral areas of the visual field, a mechanism referred to as covert attention (Bisley, 2011). Its neural basis lies in cortical visual regions and the parietofrontal network. Under an internalist understanding of attention, the analysis might end here. However, to fully understand visual attention, we also need to attend to eye movements and knowledge of sensorimotor contingencies (O'Regan & Noë, 2001), and then, importantly, we need to look into how culture transforms a person's attentional patterns.

From an enculturated-extended perspective, it is crucial to recognize that individuals learn culturally specific ways of attending to and interacting with the world, shaping their attention styles through interactions with others, cultural tools, and guided social participation. This process, which is a lifelong endeavor, influences how people perceive, process, and interpret information (Gavelek & Kong, 2012).

For example, recently it has been shown in an experimental setting how verbal attention guidance effectively influences children's attention styles (Jurkat, Gutknecht-Stöhr, & Kärtner, 2024). Verbal attention guidance is

thus an important cultural practice that contributes to the development of culture-specific visual attention patterns in childhood.

Given this, it is important to note that the enculturation of attention can explain key elements in adaptive preference formation. For example, revisiting our opening case, Celia's upbringing in a sexist culture, may have directed her attention at some things while neglecting others. The practices and narratives that she has been enculturated into explain why, given a certain circumstance, she focuses more on how her face looks than on other features. This might explain why she is more attentive to how other people react to her appearance, than to than what they say about her skills, or how she feels about a given situation.

This form of cultural cognitive railroading is not an internal activity: it not only happens but is sustained and activated through interaction with material culture, hence the cognitive extension thesis. There have already been some accounts of *extended attention*. For example, Bruineberg and Fabry (2022) address how digital technologies can extend and shape a person's attentional capacities. If this is understood in the context of attention economy, where attention is understood as a scarce commodity (Hendricks & Vestergaard, 2019; Williams, 2018;), some technologies, especially current digital technologies such as smartphones and apps like social media, compete for attention. In the context of adaptive preferences, we can argue that subjects can become oppressed through this competition for attention, as attention is culturally malleable. Take, for example, smartphone use nowadays, and its effect on attention, particularly through features like recommender systems on social media. These tools direct a person's attention to specially curated options thereby shaping and "enculturating" their desires and preferences.

In light of this, Celia's adaptive preference for undergoing cosmetic interventions at such a young age may be partly explained by the transformations digital technologies have caused in her attention. For example, as shown in my previous research, the use of a widespread type of photo filter, called beauty filters, in social media, has been linked with an increase in face dissatisfaction and the pursuit of cosmetic procedures (Andrada 2025). All this can be expressed in the table below (see Table 16.1).

A similar analysis can be applied to other cognitive abilities that contribute to adaptive preference formation. Take memory, for example. The enculturation of memory is evident, for instance, in the way adults' discussion of past events with children influence both what they remember and how they remember it. For example, in an influential article, Mullen and Soonhyung (1995) show how culture plays a crucial role in memory formation and recall. The central idea is that adults can significantly influence a child's memory by engaging in conversations about past events, emphasizing certain aspects of the experience, and teaching the child the value of remembering. This occurs through a process of memories "co-construction" where adults guide a child in verbalizing their experiences. Such discussions about past events influence

Table 16.1 Attention's role in Celia's adaptive preference from an Ex/Enc perspective

Adaptive preference	Celia's desire to undergo a certain aesthetic medical procedure (e.g., dermal fillers, and anti-wrinkle injections) instead of going to dance classes.
Cognitive ability under evaluation	Attention (visual attention)
Neural and other embodied features:	The parietofrontal network and visual cortex. Also, knowledge of sensorimotor contingencies.
Tools and artifacts:	AR Filters, smartphone apps, social media content.
Cultural practices:	Aesthetic practices, "feminine" narratives and expectations.
Cognitive mechanism/ specific distortion(s)	Celia's attention is drawn to socially valued traits, such as the appearance of her face, rather than to other aspects of who she is or how she feels, like the things she genuinely enjoys doing.

the development of autobiographical memory by teaching children which aspects of their experiences are deemed memorable.

Given this, it is easy to see how childhood enculturation of memory might explain the formation of adaptive preferences. For example, in a sexist culture, with strong aesthetic pressures and expectations on women, they may remember certain episodes of their lives, particularly those related to their appearance, rather than others. This selective memory might explain the formation of adaptive preferences, such as Celia's.

A similar dynamic can be seen in how artifacts and material culture contribute to the shaping of memory. As we saw, from an extended cognition perspective, material objects are not merely auxiliary tools for cognition (in this case, of memory) but can be constitutive elements of the cognitive ability itself. For example, Heersmink's account of *evocative objects,* that is, emotionally and autobiographically significant artefacts such as photographs, diaries, or souvenirs, suggests that such items are integral to extended memory, as they scaffold autobiographical reflection and identity construction (Heersmink, 2018). His point is that these objects are deeply integrated into personal cognitive routines and play a central role in how individuals access and shape past experiences.

Along these lines, there is also a tradition in material culture studies and feminist thinking which focuses on how everyday material culture, particularly in domestic settings, mediates identity formation. For example, Attfield (2000) shows that seemingly mundane and neutral objects often encode

gendered expectations and hierarchies, subtly guiding the development of preferences and behaviors. This supports the fact that adaptive preferences, such as Celia's, are partly formed by interactions with specific domestic artifacts and objects.

Summing up, the previous remarks show that an explanation of the formation of adaptive preferences requires paying attention not only to internal cognitive operations but also to the interplay between neural and other embodied features, cultural practices, and material culture. This can offer a comprehensive explanation of the intricate elements in play in adaptive preferences. In fact, this roadmap helps us complement the previously mentioned cognitive accounts.

For instance, as we briefly saw, Eftekhari (2021) argues that adaptive preferences result from flaws in evaluative reasoning, such as the shifting of the meaning of reasons to accommodate subordination. This process can be explained as the result of specific enculturating processes and developmental trajectories, reinforced by specific material environments. The distortion in reasoning could also stem from an oppressive enculturation of executive functions, including an individual's ability to response, plan, and reason. The same analysis applies to Khader's account of adaptive preferences as trade-offs which are liked to a misperception of interests (Khader, 2011). Enculturation, social learning, and material culture arrangements can lead individuals to misperceive their actual interests by directing self-worth toward specific domains and away from other alternatives. Exploring all the details of how this process unfolds can be a fruitful task for future research.

16.4 Implications for adaptive preference intervention design

In this last section, I want to show how an enculturated-extended cognition approach to adaptive preferences provides valuable insights for designing interventions to address adaptive preferences.

Adaptive preference intervention design refers to ethically grounded strategies aimed at addressing preferences that develop under conditions of oppression, particularly those that seem to perpetuate a person's own subordination. Since these preferences reflect oppression, a key issue in the pursuit of social justice is how to support individuals in forming preferences that are not oppressive but liberating. This task raises normative concerns about how to respect a person's agency while also promoting justice. This is one of the key challenges of this type of intervention design: developing strategies that disrupt the mechanisms producing and sustaining adaptive preferences, without dismissing the rational agency of those who hold them. As I will proceed to show, an enculturated-extended approach to human cognition, such as the one outlined in this chapter, offers interesting contributions to some of the issues raised in this debate.

For example, Nussbaum (2001) argues that interventions should aim to expand individuals' real freedoms such as capabilities that support dignity

and choice, and this involves challenging preferences that are shaped by social inequality and injustice. Nussbaum supports the idea that development should be oriented toward creating the conditions under which people can reflect on and potentially revise their preferences through democratic and deliberative processes. This capabilities framework is informed quite nicely by an enculturated-extended cognition mostly in two points.

First, understanding human cognition from this perspective helps avoid the risk of paternalism, as it underscores that adaptive preference intervention should respect cultural variation, and avoid assuming a universalist standard for the capabilities whose development should be promoted. Second, it reinforces the idea that acquiring and developing capabilities for contextual and culturally relative flourishing, requires focusing not only on the internal states of agents, but also on their material conditions, including the surrounding material culture and collective cultural practices. This last point, in turn, addresses a key issue raised by Khader (2011).

Khader argues that not all adaptive preferences require interventions targetting the psychological states of a person, as their cause may not lie in their psychology but in their material conditions.[5] This is a common mistake which she calls "psychologizing the structural", namely, incorrectly assuming that "a person is failing to flourish primarily because of problems with her psychology (her values, desires, etc.) rather than because of her structural environment" (p. 56). This, she continues,

> promotes ineffective development interventions. Interventions aimed exclusively at changing people's values and attitudes—like helping people build self-esteem, for instance—will likely do little on their own to improve the lives of people whose flourishing requires expanded options or structural change.
>
> (p. 59)

In such cases, Khader argues that interventions should begin by addressing structural and material constraints, such as poverty and disempowerment, rather than focusing on a person's psychological states. Only once those conditions are improved can individuals meaningfully deliberate and explore alternatives to their existing preferences.

I believe that this issue is strengthened and nuanced by an enculturated-extended cognition framework. If cognition itself, including preference formation, is shaped and constituted by a person's environment, then oppressive environments do more than restrict available options: they also constitute the very preferences individuals come to hold. In this light, while it is important to distinguish adaptive preferences that arise primarily from psychological

5 This aligns with relational autonomy models (Mackenzie & Stoljar, 2000).

attitudes from those that originate in material conditions (e.g., survival under deprivation), it remains crucial to recognize that individuals cannot meaningfully revise their preferences without also revising the external structures that sustain and constitute their cognition and identity. In other words, following Khader's terminology, the structural is "psychologized" (and vice versa) whether one wants it or not, so it is important to recognize that neither dimension can be changed in isolation from the other. That is why efforts to revise adaptive preferences must attend to both the psychological and the material dimensions of preference formation. For instance, reconfiguring material culture, institutional norms, or labor divisions should not be a secondary consideration but a central aspect of how an individual could change their psychological attitudes. In other words, to change how people think, we must change their context, and to change context, we must also change cognition.

16.5 Wrap up

We began this chapter by illustrating the phenomenon of adaptive preferences through the case of Celia, a student who desires to undergo a cosmetic medical procedure at a very young age. I have suggested that Celia's behavior can be understood through a key concept in feminist theory: adaptive preferences. To have an adaptive preference, that is, one that reflects her subordination or oppression, we would need to check whether, in the absence of oppressive conditions, Celia would have chosen another course of action. In Celia's case, this means that without the extreme aesthetic pressure placed on young women of their generation, she would not want to spend her money on those procedures. Instead, she would likely choose the school trip and dance lessons.

Celia's case, like any instance of this type of preference, can be explained by examining the cognitive processes and mechanisms behind its formation, including the psychological attitudes, capacities and the ways in which the subject interprets and justifies them, whether consciously or not. It is here that I have noticed, under a standard internalist understanding of cognition, a tendency is to focus on internal cognitive operations that are distorted by the so-called internalization of oppressive or unjust social structures.

The key of my argument has been that enculturated-extended cognition encourages us not to focus solely on the internal cognitive aspects of the holder of adaptive preferences, and contrast it with oppressive external conditions that are "internalized". Instead, it urges us to examine how such preferences are formed through cognitive transformations, that are themselves constituted by material devices and artifacts with which the person consistently interacts, and by the cultural practices that guide these and other social interactions. This broader view emphasizes the co-constitutive role of the material and cultural environment in the formation of preferences, rather than simply viewing them as internal psychological responses to external oppression.

This provides a more comprehensive picture of the cognitive dimension of adaptive preferences. As I have shown, the process of cognitive enculturation and extension transforms key cognitive abilities in preference formation, such as attention or memory. In Celia's case, her adaptive preference can be partly explained by by transformations in her memory and attention, which lead her to focus on and recall features and interactions related to her aesthetic appearance. These cognitive transformations are shaped by her upbringing, and actively sustained by a range of cultural artifacts and technologies with which she interacts, further directing her preferences.

Lastly, I have explored some of the implications that an enculturated-extended cognition perspective on human cognition has for adaptive preference intervention design. First, I have shown how it can contribute to developing a more culturally situated or contextualized capability approach like the one proposed by Nussbaum (2001). Additionally, I have argued (following an important point made by Khader (2011)) that an enculturated-extended cognition framework highlights the fact that there is no sharp distinction between psychological attitudes and structural constraints in intervention design. Rather, these are deeply interconnected such that preferences cannot change without changing a person's environment, and vice versa.

Consequently, although many details remain to be addressed, an enculturated and extended approach to cognition appears to be a valuable ally for both adaptive preference theorists and development practitioners seeking a more comprehensive understanding of adaptive preference formation. In this chapter, I have shown how this promising collaboration can begin.

References

Andrada, G. (2025). Beauty filters in self-perception: The distorted mirror gazing hypothesis. *Topoi, 44*(2), 367–378.

Anzaldúa, G. (1987). *Borderlands/La frontera: The new mestiza* (1st ed.). Aunt Lute Books.

Attfield, J. (2000). *Wild things: The material culture of everyday life*. Berg Publishers

Bartky, S. L. (1990). *Femininity and domination: Studies in the phenomenology of oppression*. Routledge.

Bisley, J. W. (2011). The neural basis of visual attention. *Journal of Physiology, 589*(Pt 1), 49–57. https://doi.org/10.1113/jphysiol.2010.192666.

Bruineberg, J., & Fabry, R. E. (2022). Extended mind-wandering. *Philosophy and the Mind Sciences, 3*, 1–30. https://doi.org/10.33735/phimisci.2022.9190

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333213.001.0001

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Cudd, A. (2006). *Analyzing oppression*. Oxford University Press.

Eftekhari, S. (2021). The irrationality of adaptive preferences: A psychological and semantic account. *Utilitas, 33*(1), 68–84. https://doi.org/10.1017/S0953820820000321

Elster, J. (1983). *Sour grapes: Studies in the subversion of rationality*. Cambridge University Press.

Fabry, R. E. (2018). Betwixt and between: The enculturated predictive processing approach to cognition. *Synthese, 195*, 2483–2518. https://doi.org/10.1007/s11229-017-1334-y

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Gavelek, J. R., & Kong, A. (2012). Learning: A process of enculturation. In Seel, N. M. (ed) *Encyclopedia of the sciences of learning*. Springer. https://doi.org/10.1007/978-1-4419-1428-6_868

Haslanger, S. (2019). Cognition as a social skill. *Australasian Philosophical Review, 3*(1), 5–25. https://doi.org/10.1080/24740500.2019.1705229

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences, 14*, 577–598. https://doi.org/10.1007/s11097-014-9355-1

Heersmink, R. (2018). The narrative self, distributed memory, and evocative objects. *Philosophical Studies, 175*(8),1829–1849.

Hendricks, V. F., & Vestergaard, M. (2019). *Reality lost: Markets of attention, misinformation and manipulation*. Springer. https://doi.org/10.1007/978-3-030-00813-0

Hixon, J. G., & Swann, W. B., Jr. (1993). When does introspection bear fruit? Self-reflection, self-insight, and interpersonal choices. *Journal of Personality and Social Psychology, 64*(1), 35–43. https://doi.org/10.1037/0022-3514.64.1.35

Jurkat, S., Gutknecht-Stöhr, A. C., & Kärtner, J. (2024). The socialization of visual attention: Training effects of verbal attention guidance in urban German children. *Developmental Psychology*. Advance online publication. https://doi.org/10.1037/dev0001746

Khader, S. J. (2011). *Adaptive preferences and women's empowerment*. Oxford University Press.

Khader, S. J. (2012). Must theorising about adaptive preferences deny women's agency? *Journal of Applied Philosophy, 29*(4), 302–317.

Kirchhoff, M. D. (2012). Extended cognition and fixed properties: Steps to a third-wave version of extended cognition. *Phenomenology and the Cognitive Sciences, 11*(2), 287–308. https://doi.org/10.1007/s11097-011-9237-8

Knowles, C. (2021). Beyond adaptive preferences: Rethinking women's complicity in their own subordination. *European Journal of Philosophy, 30*(4), 1–18. https://doi.org/10.1111/ejop.12742

Mackenzie, C., & Stoljar, N. (Eds.). (2000). *Relational autonomy: Feminist perspectives on autonomy, agency, and the social self*. Oxford University Press

Maiese, M. (2021). Mindshaping, enactivism, and ideological oppression. *Topoi, 41*(2), 341–354. https://doi.org/10.1007/s11245-021-09770-1.

Menary, R. (2007). *Cognitive integration: Mind and cognition unbounded*. Palgrave Macmillan.

Menary, R. (2018). Cognitive integration: How culture transforms us and extends our cognitive capabilities. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford handbook of 4E cognition* (pp. 187–215). Oxford University Press

Morgan, K. P. (1991). Women and the knife: Cosmetic surgery and the colonization of women's bodies. *Hypatia*, 6(3), 25–53.

Mullen, M. K., & Soonhyung, Y. (1995). The cultural context of talk about the past: Implications for the development of autobiographical memory. *Cognitive Development*, *10*(3), 407–419. https://doi.org/10.1016/0885-2014(95)90004-7

Newen, A., De Bruin, L., & Gallagher, S. (2018). *The Oxford handbook of 4E cognition*. Oxford: Oxford University Press.

Nussbaum, Martha C. (2001). Symposium on Amartya Sen's philosophy: 5 adaptive preferences and women's options. *Economics and Philosophy, 17*(1), 67–88.

O'Regan, J. Kevin, & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(5), 883–917.

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica, 144*(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Superson, A., & Brennan, S. (2005). Feminist philosophy in the analytic tradition. *Hypatia, 20*(1), 1–9. https://doi.org/10.1111/j.1527-2001.2005.tb00532.x

Sutton, J. (2010). Exograms and Interdisciplinarity: History, the extended mind, and the civilizing process. In Menary, R. (Ed.), *The extended mind* (pp. 189–225). Cambridge: MIT Press.

Williams, J. (2018). *Stand out of our light: Freedom and resistance in the attention economy*. Cambridge University Press. https://doi.org/10.1017/9781108453004

Young, I. M. (2009). Five faces of oppression. In G. L. Henderson & M. Waterstone (Eds.), *Geographic thought: A praxis perspective* (pp. 55–71). Routledge.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Index

affordance 7, 8, 25, 48, 50–71, 119, 120, 134–139, 211, 212

Brandom, Robert B. 100–112, 212–220

cognitive science 4, 6, 8, 40, 43, 47, 52, 64, 77, 81–95, 119–120, 135, 143–149, 181–184, 226–228, 248–249
computation 4–7, 35, 39, 43, 46, 54–58, 63, 71, 85–90, 94, 114, 117–121, 144, 154, 181, 227–231, 238, 248
consciousness 3, 13, 21, 27, 30, 58, 116, 212, 226

enactivism 7, 8, 35, 37, 77–82, 88, 93–94, 117, 120, 129, 133–138, 145, 148, 176, 198, 211, 230, 237–238

Gibson, Eleanor J. 3, 9, 44, 48, 55, 57, 61, 62, 67
Gibson, James J. 3, 8, 9, 25, 26, 30–67, 117, 119, 134–135, 169

habit 13–14, 21–23, 62, 135–150, 231, 235

intentionality 3, 9, 27, 32, 39, 40, 83, 84, 101–111, 144–154, 162–169, 174–177, 211, 223

McDowell, John H. 99–111, 129–140

normativity 5, 9, 132, 144–154, 211–227, 234–235

representation 3–9, 15, 34–53, 61–79, 82–94, 114–129, 133–154, 167, 175–176, 181–191, 211, 216, 226–227, 230, 236, 238, 248–249
Ryle, Gilbert 5–9, 15, 44, 50, 63–64, 202, 233, 237

skills 14, 21, 60, 67, 68, 77, 164–165, 184, 211, 220, 222, 254

teleosemantics 83–84, 143–148, 151–155

Varela, Francisco 4, 7, 77, 79, 116, 117, 120, 145, 148, 149

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com