

Rivista Internazionale di Filosofia e Psicologia si propone come luogo di discussione e riflessione critica sui nessi che legano ricerca filosofica e ricerca psicologica. In un clima culturale caratterizzato da un crescente specialismo e da un'estrema parcellizzazione delle attività di ricerca, che rendono problematica la circolazione delle idee e lo scambio delle conoscenze, *Rivista Internazionale di Filosofia e Psicologia* promuove una pratica teorica aperta, critica, laica, antidogmatica, interdisciplinare e interculturale, che non considera filosofia e psicologia come due forme di sapere chiuse e ripiegate su se stesse, ma come due direzioni di ricerca reciprocamente intersecanti in virtù del loro comune oggetto d'indagine: la condizione umana, unitamente al complesso delle sue relazioni con il contesto ambientale e sociale.

Rivista Internazionale di Filosofia e Psicologia ospita contributi di natura critico-comparativa ed empirica aventi come oggetto la mente umana, con particolare attenzione ai processi cognitivi, emotivi e motivazionali; i fondamenti della teoria della conoscenza; la costituzione dell'intersoggettività; le dinamiche individuali e sociali delle relazioni etiche; il rapporto tra umano e non-umano; la genesi e la riproduzione delle formazioni culturali e le dinamiche interculturali; l'origine del campo della "anomalia" e del "patologico", anche alla luce delle recenti scoperte compiute nei campi delle neuroscienze, delle scienze cognitive, della biologia e della genetica. Sono incoraggiati studi diretti alla valorizzazione dei modelli interpretativi dell'umano e del non-umano storicamente considerati "marginali" rispetto alla loro capacità di creare un dialogo critico e costruttivo con le nuove scienze della mente, con particolare attenzione sia agli indirizzi filosofici "continentali", tra cui il trascendentalismo, l'idealismo, la fenomenologia e la Teoria Critica, sia a quelli derivanti dalla tradizione psicodinamica e psicoanalitica.



Rivista quadrimestrale

Registrazione presso il Tribunale di Milano n. 634 in data 26.11.2010 - Direttore responsabile: Aurelia Delfino
ISSN 2039-4667; E-ISSN 2239-2629
Proprietà: Associazione R.I.F.P.
Via Dante Alighieri, 262 - 70122 BARI

Editore

Mimesis Edizioni
Piazza Don Enrico Mapelli, 75 - 20099 Sesto San Giovanni (MI)
Tel. +39 02 24861657
E-mail: mimesis@mimesisedizioni.it
Sito web: www.mimesisedizioni.it

Stampa

Digital Team, via dei Platani, 4 - 61038 Fano (PU)

Modalità di invio dei contributi

Le proposte di pubblicazione vanno inviate secondo le modalità indicate sul sito <http://www.rifp.it>

Direzione e Redazione

Dipartimento di Scienze della Formazione, Psicologia, Comunicazione - Università degli Studi di Bari "Aldo Moro"
Palazzo Chiaia-Napolitano, via Crisanzio, 42 - 70121 Bari (I)
Tel. +39 080 5714540

■ Editor in Chief

Sandro Nannini - *Siena*

■ Managing Editor

Luigi Pastore - *Bari*

■ Editorial Board

Saulo de Freitas Araujo - *Juiz de Fora*

Pierluigi Barrotta - *Pisa*

Roberto Cubelli - *Trento*

Sara Dellantonio - *Trento*

Alfredo Ferrarin - *Pisa*

Remo Job - *Trento*

Claudio La Rocca - *Genova*

Andrea Lavazza - *Arezzo*

Enrico Moriconi - *Pisa*

Luigi Pastore - *Bari*

Achille Varzi - *New York*

■ Editorial Office

Magda Altman - *San Sebastian*

Giovanni Coppolino Billè - *Bari*

Antonio Di Oronzio - *Bari*

Paolo Petricca - *Chieti*

Giuseppe Saracino - *Bari*

■ Layout Editor

Mario Velluso - *Bari*

■ Scientific Advisory Board

Maria Armezzani - *Padova*

Bethania Assy - *Rio de Janeiro*

Lorenzo Bernini - *Verona*

Alberto Giovanni Biuso - *Catania*

Andrea Bosco - *Bari*

Marco Buzzoni - *Macerata*

Mario De Caro - *Roma*

Alessandra De Coro - *Roma*

Vincenzo Costa - *Milano*

Nevia Dolcini - *Macao*

Luciano Floridi - *Oxford*

Pim Haselager - *Nijmegen*

Barbara Held - *Bowdoin*

Uriah Kriegel - *Paris*

Alexei Krioukov - *Samarkand*

Ping Li - *Guangzhou*

Lorenzo Magnani - *Pavia*

Annette Mülberger - *Groningen*

Valerio Meattini - *Bari*

Cristina Meini - *Vercelli*

Carmela Morabito - *Roma*

Alberto Mario Mura - *Sassari*

Angelica Nuzzo - *New York*

Alberto Oliverio - *Roma*

Lisa Osbeck - *Carrollton*

Alfredo Paternoster - *Bergamo*

Martina Plümacher - *Berlin*

Athanassios Raftopoulos - *Nicosia*

Giacomo Romano - *Siena*

Giovanni Stanghellini - *Chieti*

Achim Stephan - *Osnabrück*

Thomas Sturm - *Barcelona*

Luca Vanzago - *Pavia*

Nicla Vassallo - *Genova*

Giuseppe Varnier - *Siena*

Andrea Zhok - *Milano*

**RIVISTA INTERNAZIONALE
DI FILOSOFIA E PSICOLOGIA**
Anno 14 • Numeri 1-2 • 2023

SOMMARIO

EDITORIALE

- 1 *Sara Dellantonio, Luigi Pastore*
What future for cognitive science(s)?

RICERCHE

- 11 *Lisa Osbeck & Saulo de Freitas Araujo*
The future of cognitive science is pluralistic, but what does that mean?
- 27 *Philip V. Kargopoulos*
What philosophy, if any, is needed for cognitive science?
- 40 *Max Coltheart*
Psychology is – and should be – central to cognitive science
- 59 *Aurora Alegiani, Massimo Marraffa & Tiziana Vistarini*
Two open questions in the reformist agenda of the philosophy of cognitive science
- 74 *Frank van der Velde*
From unified to specific theories of cognition
- 88 *William Bechtel*
Rethinking cognitive architecture: A heterarchical network of different types of information processors
- 103 *Daniel C. Burnston & Antonella Tramacere*
Distributed loci of control: Overcoming stale dichotomies in biology and cognitive science
- 118 *Sandro Nannini*
The mind-body problem between philosophy and the cognitive sciences
- 135 *Alan Costall*
"Cognition" - Let's forget it?

RECENSIONI

- 142 Irriducibile. La coscienza, la vita, i computer e la nostra natura/F. Faggin
(*Cristiano Cali*)
- 146 La mente allargata/R. Manzotti
(*Daniela Troiano*)



I contenuti di Rivista Internazionale di Filosofia e Psicologia
Sono soggetti alla licenza Creative Commons
Attribuzione-4.0 Internazionale

Per maggiori informazioni:
<http://creativecommons.org/licenses/by/4.0/>

EDITORIALE

What future for cognitive science(s)?

Sara Dellantonio^(α) & Luigi Pastore^(β)

Abstract In this introduction to the thematic issue on *the future of the cognitive science(s)*, we examine how challenges and uncertainties surrounding the past and present of this discipline make it difficult to chart its future. We focus on two main questions. The first is whether cognitive science is a single unified field or inherently pluralistic. This question can be asked at various levels: First, with respect to the disciplines that should be included in the cognitive hexagon and their reciprocal relationships: should we speak of cognitive science or of the cognitive sciences? Second, with regard to the conceptual and methodological changes (turns or revolutions) that have taken place within the cognitive project from its inception to the present day. Third, it pertains to cognitive psychology as a discipline. Before the emergence of cognitive science psychology was a fragmented discipline characterized by different traditions and approaches: has cognitive science been able to stem this fragmentation? Finally, we can question the unity of the cognitive architecture itself: is cognition produced by homogeneous or heterogeneous mechanisms for information processing? We show that the issue of unity is addressed by several of the papers included in this thematic issue. In the second part of this introduction, we query the role that each component discipline should play in the cognitive project and in particular which should lead the project going forward, and why. Again, we show how this issue has been tackled by several articles featured in this collection.

KEYWORDS: Future of Cognitive Science; Cognitive Psychology; Pluralism and Cognitive Science; Philosophy and Cognitive Science; Fragmentation of Psychology

Riassunto *Quale futuro per la scienza cognitiva?* - In questa introduzione al fascicolo tematico dedicato al *futuro della scienza cognitiva* prendiamo in considerazione sfide e incertezze che caratterizzano il passato e il presente di questa disciplina, rendendo difficile prevedere il suo futuro. Due le questioni principali su cui concentriamo l'attenzione. La prima: la scienza cognitiva è un ambito unitario o intrinsecamente pluralistico? Questo problema si manifesta a diversi livelli. In primo luogo, riguarda le discipline che dovrebbero comporre l'esagono cognitivo e le loro relazioni reciproche: dovremmo parlare di scienza cognitiva o di scienze cognitive? In secondo luogo, riguarda le trasformazioni (svolte o rivoluzioni) concettuali e metodologiche avvenute all'interno di questo progetto dalla sua nascita fino ai giorni nostri. In terzo luogo, riguarda la psicologia cognitiva. Prima della nascita della scienza cognitiva la psicologia era una disciplina frammentaria caratterizzata da diverse tradizioni e approcci: possiamo dire che la scienza cognitiva abbia posto rimedio a tale frammentazione? Infine, il problema dell'unità sorge anche in relazione all'architettura cognitiva stessa, considerando che la cognizione potrebbe o potrebbe non essere prodotta da meccanismi omogenei di elaborazione delle informazioni. Nella presentazione che segue si cerca di mostrare come l'unità, in tutte queste varianti, sia una questione affrontata da diversi articoli presenti in questo fascicolo tematico. Nella seconda parte di questa introduzione prendiamo in considerazione il problema del ruolo che ciascuna disciplina componente dovrebbe svolgere nel progetto cognitivo e, in particolare, quale dovrebbe guidare il progetto e perché, ponendo in evidenza come anche questa sia una questione centrale, tematizzata da diversi lavori presentati in questo fascicolo.

PAROLE CHIAVE: Futuro della scienza cognitiva; Psicologia cognitiva; Pluralismo e scienza cognitiva; Filosofia e scienza cognitiva; Frammentazione della psicologia

^(α)Dipartimento di Psicologia e Scienze Cognitive, Università degli Studi di Trento, Palazzo Fedrigotti, Corso Bettini, 31 – 38068 Rovereto (IT)

^(β)Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università degli Studi di Bari "Aldo Moro", Palazzo Chiaia-Napolitano, via Crisanzio, 42 – 70121 Bari (IT)

E-mail: sara.dellantonio@unitn.it (✉); luigi.pastore@uniba.it (✉)



WHAT FUTURE FOR COGNITIVE SCIENCE(S)? Charting the future is a complex task that requires a clear and detailed understanding of the present and the past. When it comes to cognitive science the challenge is made even more complex by the many uncertainties that surround its present and past. First and foremost, we can question the identity of cognitive science as a scientific discipline. Should we speak of cognitive science in the singular, or are we dealing with an inherently and irreducibly plural field of study? How many and which disciplines can rightfully be considered integral parts of cognitive science? Has cognitive science as a discipline developed in a unified and coherent manner, or one characterized by fractures and discontinuities?

Even today, when we think about cognitive science, it is difficult not to conjure up that classic image: the cognitive hexagon, proposed in 1978 in a report presented by Samuel Keiser, George A. Miller, and Edward Walker to the Sloan Foundation.¹ Undoubtedly, for many professionals in the field, as well as non-specialists, this is the most common and widely shared representation of cognitive science; it shows a unified and transdisciplinary field of research fueled by contributions from different disciplines. The image of the cognitive hexagon has a twofold effect. From a descriptive standpoint, the vertices define the disciplines included in the project: anthropology, philosophy, artificial intelligence, linguistics, psychology, and neuroscience. From a normative perspective, the connections between these vertices indicate the roles attributed to certain disciplines – and reveal that some (particularly anthropology and philosophy) are less “central” than others.

The image of the cognitive hexagon was the result of a process of systematization that unfolded over a couple of decades. It was the outcome of a complex journey accomplished in several stages that affected the number and structure of the fields of inquiry involved in research on cognition. Interrogating the identity of cognitive science thus involves reflecting on this long and intricate process and wondering whether cognitive science has finally stabilized or instead continues to evolve.

The idea that cognitive science can be understood as a unified field of research and conceived as a singular scientific discipline can be traced back to a five-year period before the preparation of the report for the Sloan Foundation.² More specifically, to 1973 when Hugh C. Longuet-Higgins, a physicist and chemist by training and a scholar actively engaged in creating artificial models for the study of intelligence and perception, wondered, «what science or sciences are likely to be enriched by artificial intelligence studies?». Longuet-Higgins’ answer is still relevant: «*all those sciences which are directly relevant to human thought and perception. These cognitive sciences*

may be roughly grouped under four main headings: *Mathematical* – including formal logic, the theory of programs and programming languages, the mathematical theory of classification and of complex data structures; *Linguistic* – including semantics, syntax, phonology and phonetics; *Psychological* – including the psychology of vision, hearing and touch; and *Physiological* – including sensory physiology and the detailed study of the various organs of the brain».³

According to Longuet-Higgins, cognitive science cannot be conceived as a set of well-defined disciplines and their relations, but rather as a field of research formed by the overlap of various domains which are so closely interconnected as to form a unique whole. This is the reason why he suggests using the term *cognitive science*: «perhaps *cognitive science* in the singular would be preferable to the plural form, in the view of the ultimate impossibility of viewing any of these subjects in isolation».⁴

Longuet-Higgins’ reflections have foundational value for understanding the construction of the theoretical and methodological identity that cognitive science assumed between its birth and the late 1970s, a period in which it found a more precise definition, culminating in the adoption of the cognitive hexagon as its institutionalized representation. The unity that emerged from Longuet-Higgins’ reflection was of a theoretical, methodological, and linguistic nature: those disciplines that make up cognitive science must somehow benefit from the theoretical and methodological developments in artificial intelligence. Computer science and cybernetics were research programs capable of providing the theory, methodology, and language that would produce unity and, therefore, a seminal form of disciplinary identity in an emerging field of study. How? Perception, attention, competent language use, memory, problem-solving, decision-making and other aspects of intelligent behavior were theorized and described as types of information processing: they result from the way in which an organism detects processes and transforms available information, thereby giving rise to organism internal entities such as symbols and representations; the latter can be used to provide a causal explanation of human intelligence and behavior.

As the name suggests, *cognitive science* specifically investigates *cognition*. This term has a long and very important history in the philosophical and psychological lexicon; but took on a changed meaning within the lemma *cognitive science*.⁵ As highlighted by Thomas Sturm and Hans Gundlach, the word *cognition* came to us via the Latin translation *cognitio* of older Greek terms (*gnosis*, *gignosko*) for human knowledge and the human ability to acquire knowledge. So, *cognition* originally referred to the acquisition, preservation, and (theoretical and practical) use of knowledge by the human

mind. This idea was emphatically summarized in 1732 in Christian Wolff's *Psychologia empirica methodo scientifica pertractata*: «cognitio est actio animae, qua notionem vel ideam rei sibi acquirit».⁶ An echo of this definition can still be found a couple of centuries later in the entry for “*cognition*” in George F. Stout and James M. Baldwin's *Dictionary of philosophy and psychology* (1918): «cognition is an ultimate mode of consciousness coordinate with conation and affection».⁷ Traditionally, the word cognition indicates the acquisition and use of knowledge and was intrinsically linked to consciousness, volition, and emotion.

And yet, this is not the sense in which the term cognition is used in the lexicon of cognitive science. The word *cognition* still designates an investigation of the processes by which individuals acquire, retain, and utilize knowledge, but from a very different perspective. *Cognition* is no longer used to indicate the *contents* (ideas) or *processes* (intellectual capacities or psychological faculties) of knowledge acquisition accessible from the first-person perspective but rather to describe the acquisition and use of mainly abstract knowledge separate from the *emotional* and *conative* spheres. The use of terms such as *cognition* or *cognitive* are strictly related to notions and concepts such as *information*, *information processing*, *computation*, *representation*, *symbol*, and *manipulation of symbols*.⁸ As Ulric Neisser wrote in his *Cognitive psychology*: «the term “cognition” refers to all the processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used. It is concerned with these processes even when they operate in the absence of relevant stimulation, as in images and hallucinations. Such terms as *sensation*, *perception*, *imagery*, *retention*, *recall*, *problem-solving*, and *thinking*, among many others, refer to hypothetical stages or aspects of cognition. Given such a sweeping definition, it is apparent that cognition is involved in everything a human being might possibly do; that every psychological phenomenon is a cognitive phenomenon».⁹

As Neisser's quotation shows, by 1967 the transformation of the meaning “cognition” was already complete, especially if we consider “sensory input” and “information” to be synonymous terms. Cognition as a process of continuous acquisition and transformation of information gradually became a *foundational conceptual metaphor that offered a new metatheoretical framework for investigating human intelligence and behavior*, one that quickly captured general interest, especially in the field of empirical and experimental psychology.¹⁰ The lexicon of computer science and cybernetics offered linguistic conventions, theoretical models, and conceptual metaphors which allowed psychology to go back to studying the mind. It could finally renew itself, breaking with the dominant theoretical models of the time, namely those pro-

vided by behaviorism.

In its attempt to make psychology part of the natural sciences, behaviorism reduced its object of inquiry to directly observable behavior, excluding any causal reference to entities and processes not directly observable. While this theoretical move had safeguarded psychology against the methodological and theoretical dangers of introspectionism, it also generated irreducible fractures and fragmentations in psychological knowledge. These fractures impacted, for instance, the relationship between experimental psychology and clinical practice, with implications for the identification, diagnosis, and treatment of psychopathologies, and led to an increasing diversity of opinions regarding the existence and role of “unconscious” processes in shaping individual overt behaviors.

According to George Mandler, psychologists enthusiastically embraced the “new” *cognitive* movement (which he believes emerged in the five-year period between 1955 and 1960) because of «the various tensions and inadequacies [of the psychology] of the first half of the twentieth century».¹¹

The new language of cognitive science *promised* to overcome the *fragmentation* Mandler was referring to.¹² Describing human mental activity as continuous information processing and transformation seemed to provide a rigorous and scientific framework for psychology. Conceived in this manner, all mental activity could be thought of and described as a sequence of rules and a set of mechanisms operating at different levels, most of them under the threshold of consciousness and therefore below the level of awareness. Indeed, this new approach could account for unconscious dimensions of human thought in terms of subpersonal information processing: not all the information available to individuals is known to them; they are not aware of how this information is processed before they use it; and not all uses of available information are consciously and deliberately controlled. Mentalistic elements were reincorporated in the lexicon of psychology and mental activities could be considered to play a causal role in the production of intelligent thought and behavior.

Nearly seventy years after the birth of this new discipline, we can take stock and consider what may lie ahead. In particular, we can assess the current status of the project and consider whether cognitive science has actually helped us overcome the tensions and inadequacies of past theories of mind. And we can make predictions regarding future developments, asking in what direction this project will or should progress.

When the cognitive science project launched its new methodological and theoretical approaches to the investigation of mind, these were considered to constitute a (cognitive) “*revolution*”.¹³ Some authors believe there have been other revolutions or – at the very least – *turns* and *shifts* that have radically modi-

fied the original project. More precisely, they point out that the original project was never fully abandoned, but continued to coexist alongside new lines of research, leading to the development of multiple school of thoughts within cognitive science.

Typically, the (first) cognitive revolution is linked to the *brain as computer* analogy, a functionalist conception of the relationship between mind and brain, and the idea that information processing systems can be analyzed at three levels:¹⁴ first, by describing the *computational* problems they must solve, i.e. the tasks it must accomplish; second, by identifying the *algorithms* that can solve these problems, i.e., the way in which system tasks can be accomplished; third, by determining how such algorithms can be *implemented* by the hardware, i.e., by the biological substrate. Two essential tenets of traditional cognitive science related to this idea are *internalism* and *representationalism*. As well summarized, for instance, by Mario Villalobos and David Silverman: «internalism, roughly, is the idea that the material realisers (or vehicles) of cognitive states and processes reside, always and entirely, inside the organism's boundaries (paradigmatically inside the head). Representationalism, on the other hand, is the idea that the intelligent behaviour of organisms is the result, at least partially, of the manipulation of content-bearing representations».¹⁵

Some authors argue that a second cognitive revolution, shift or turn occurred when this functional model of the mind was challenged by the development of *cognitive neuroscience*. Roughly speaking, the *cognitive neuroscience revolution* meant there could be no complete explanation of the mind independent of the brain; any psychological theory at the level of description begs investigation of its brain-level implementation. As Worth Boone and Gualtiero Piccinini state: «current neuroscience is empirically well grounded and should constrain our cognitive explanations».¹⁶

Other authors are inclined to consider the shift in attitudes towards culture and society as a second cognitive revolution, turn, or shift. In his seminal book on the history of the cognitive revolution, Howard Gardner notes that the classical version of the cognitive project included among its assumptions a «de-emphasis on affect, context, culture, and history»: «though mainstream cognitive scientists do not necessarily bear any animus against the affective realm, against the context that surrounds any action or thought, or against historical or cultural analyses, in practice they attempt to factor out these elements to the maximum extent possible».¹⁷ Over time, this attitude has changed to a greater or lesser extent.

In some cases, this change in attitude has led to the (relatively modest) acknowledgement that «culture and society constitute new functional systems and develop new human cognitive abili-

ties and functions».¹⁸ In other cases, however, developments of this idea have taken a more radical course. This is the case with so-called *4E cognition* (i.e., Embodied, Embedded, Extended, and Enactive cognition),¹⁹ an approach which highlights the importance of considering interactions between the body, environment, tools, and actions in understanding cognitive processes and even suggests that cognition *emerges* from these complex interactions. At least some forms of *4E cognition* also challenge the very tenets of traditional cognitive science, i.e., internalism and representationalism, instead embracing externalism and antirepresentationalism. These more radical forms of 4E cognition propose a description of the mind that appear to offer a genuine alternative to classic cognition, and thus can be considered to constitute a real «revolution».²⁰

The issue of cognitive revolutions/turns is tied to the previously mentioned issue of disciplinary fragmentation in psychological models, a fragmentation that the cognitive project had promised to overcome/address/resolve. In fact, in the first half of the twentieth century, psychology was characterized by different traditions or schools. According to Ardila these included structuralism, functionalism, reflexology, behaviourism, gestalt, topology, psychoanalysis, and existential psychology: «the creators of these schools considered that they had the explanation for all psychological phenomena, that they had adequate methodology, and that they could also ignore all that was done by other schools».²¹

At least originally, the cognitive project might have provided an opportunity to consolidate and improve a solid and integrated evidence-based cognitive psychology over other psychological traditions or schools.

Yet, the idea that cognitive science could lead to the development of a unitary scientific picture of cognition does not come without problems. First of all, as previously mentioned, the cognitive project consists in a constellation of disciplines and it is not obvious what we should expect from this plurality of perspectives and in particular how/whether this can produce a unitary science of mind. The classic question that has been raised since the beginning of this project is whether these disciplines will or should develop into a unitary field of study, i.e., a cognitive science (in the singular) or whether they will or should remain a multi-disciplinary domain described by the term cognitive sciences (in the plural). No final answer has been found to this question, even though at the moment, after almost 70 years of cognitive research, *de facto* the cognitive sciences have not reached any unity yet. Secondly, the goal of unifying cognitive psychology has traditionally been considered an extremely difficult if not impossible to achieve project – at least within such a short

time.²² In fact, the multiple turns in the cognitive research that characterizes the history of this project confirm that we have certainly not reached the stage of unified psychological description. As cognitive research has continued to evolve, psychological theories may have become even more fragmented, and the project of a unified description even more elusive.

Finally, the problem of unity might also be raised at the level of the cognitive system and its processes. In this respect, one might ask whether cognition is produced by homogeneous mechanisms of information processing or whether different capacities or forms of knowledge are realized through different mechanisms and processes. If cognition is produced by a homogeneous set of mechanisms, we can at least strive to develop a unitary model of cognition and cognitive architecture. But this effort is futile if cognition is produced by different mechanisms and processes: in this case, some level of fragmentation or inhomogeneity would be an intrinsic characteristic of cognition.

Unity across all these variations is a question addressed by several of the papers included in this thematic issue. In this thematic issue on *The future of the cognitive science(s)*, Lisa Osbeck and Saulo de Freitas Araujo trace the radical transformation of cognitive science from its inception to the present day. In their paper, *The future of cognitive science is pluralistic, but what does that mean?*, they illustrate how the original project of cognitive science has undergone several quite radical, if little debated transformations. Osbeck and Araujo ask how we should conceptualize such transformations: do these new views still form part of cognitive science; are they still scientific approaches to the mind; has cognitive science become pluralistic? They argue that cognitive science already includes a range of different perspectives as well as different scientific approaches and methods and will continue to be pluralistic in the future. Pluralism, they point out, is not the same as interdisciplinarity: conducting interdisciplinary research has always been part of the cognitive project and yet, in its original instantiation, interdisciplinary approaches shared common assumptions. Neither does pluralism entail relativism; in the authors' view, pluralism does not make it impossible to establish whether one perspective is better than another. The point is rather that our knowledge is necessarily limited and always in progress; only a multi-method approach can help us grasp this complexity.

This issue is also addressed – at least from a certain perspective – in Aurora Alegiani, Massimo Marraffa, and Tiziana Vistarini's paper entitled *Two open questions in the reformist agenda of the philosophy of cognitive science* and in Daniel C. Burnston and Antonella Tramacere's work, *Distributed loci of control: Overcoming stale dichotomies in biology and cognitive science*. Both of these

works consider seemingly irreconcilable positions within the cognitive science and show that/how it is possible to find coherent syntheses that mend fractures by identifying new forms of unity.

Alegiani, Marraffa and Vistarini put forward a reformist agenda for cognitive science and argue that the classic computational-representational framework must be integrated with the 4E approach to cognition (i.e., with the idea that cognition is embodied, embedded, enacted, and/or extended) to achieve the twofold aim: ending the anti-biologism and radical internalism of classical cognitive science and preserving the notions of representation and computation which still appear crucial for an explanation of the mind. Burnston and Tramacere consider the seemingly irreconcilable dichotomy between the radically internalist perspective of classical cognitive science and the radically externalist perspective of 4E cognition: «“internalists” argue that the causal source of the phenomenon of interest is internal to the organism, and “externalists” that it is external to it» (cf. this issue, *infra*, p. 104). The authors suggest that we need to reach a minimal agreement on a distributed causal source (they call it “locus of control”). This would allow us to keep (working on) the notion of mental representation even though it would have to be radically revisited. Moreover, from a methodological point of view, we would move away from the idea of a holistic system (in which individual components do not play any causal role, but only the relationships among the parts can be causally relevant) and we would be encouraged to decompose the system into internal and external factors to investigate how their separate contributions and their interactions combine to bring about fine-grained changes in behavior.

Other papers address the question of unity from the point of view of cognitive architecture and the mechanisms of information processing. Frank van der Velde's work, *From unified to specific theories of cognition*, considers whether we should expect future cognitive theories to reveal a single set of mechanisms for the explanation of different cognitive capacities or whether we should presume that different capacities will be explained on the basis of different mechanisms. The classic computational architecture theorized primarily by Newell, Fodor and Pylyshyn who aimed to guarantee the main features of human level cognition – i.e., productivity, compositionality and systematicity – was meant to describe any aspect of cognition. All cognitive processes were considered to be a form of symbol manipulation and thus to be explicable by a computational theory of cognition. Yet, in van der Velde's view «the aim for unified theories of cognition may be out of reach» (cf. this issue, *infra*, p. 75). Apparently, different kinds of cognition are possible and this opens the door to the idea that cognitive mecha-

nisms might not be unitary (or homogeneous) either: «certain forms of implementation could be selective for certain forms of information processing» (cf. this issue, *infra*, p. 79). In the paper, van der Velde examines the example of the connection structures between nouns and verbs suggested by Hebb. These are alternative to the von Neuman architecture (information is not stored by symbols) and yet they can still ensure full combinatorial productivity. This offers further corroboration for the view that «a “single set of mechanisms” might not exist» (cf. this issue, *infra*, p. 83). From this point of view, the future of cognitive science might be naturally fragmented: if our cognitive capacities are implemented by several mechanisms, we should search for distinct domain-specific cognitive architectures and their unique implementations.

William Bechtel's *Rethinking cognitive architecture: A heterarchical network of different types of information processors* goes in an analogous direction. Bechtel challenges the classic view developed in cognitive science that cognitive architecture is characterized by a homogeneous set of capacities and is organized hierarchically. The idea of a homogeneous set of capacities was suggested with enthusiastic tones and as a possibly great achievement of psychological research, e.g., by Allan Newell in 1990: «psychology has arrived at the possibility of unified theories of cognition – theories that gain their power by posing a single system of mechanisms that operate together to produce the full range of human cognition».²³

Moreover, a hierarchical design presupposes that «multiple lower-level units report to a single higher-level unit, and ultimately one top-level unit oversees the whole system. The different units are organized into a pyramid» (this issue, *infra*, p. 98). That's the way the brain is commonly described. The idea is basically that input information is processed first and then transmitted to higher processing centers to reach its ultimate destination: centralized executive control, located in the prefrontal cortex. Bechtel defends a different perspective in which the vertebrate brain is characterized by several different information processing systems. These are integrated with each other but not on the basis of a hierarchical structure. Their organization is rather heterarchical: «heterarchical organization can differ from hierarchy in multiple ways – there can be more units at higher levels than at lower levels, inverting the pyramid, and when units exercise control over others, multiple ones may control the same component and controlled components can, in some respects, exercise control over those taken to be in control» (cf. this issue, *infra*, p. 100, footnote 42). In this structure, the components are organized in a network and they can influence each other. The behavior of the organism does not result from

commands from a higher processing center which controls and coordinates all the information provided by lower-level systems but is rather the product of complex coordination between the components of the network. The aim of future research carried out by cognitive science should be to investigate the different types of information processed by the brain of the vertebrates as well as the diverse processing architectures they exhibit and the way in which their components coordinate to generate a coherent behavior. Thus, in Bechtel's view, the cognitive science of the future should overcome instead of pursuing the ideal that we might develop unified theories of human cognition based on homogeneous mechanisms and guided ultimately by high-level abstract cognition.

An analogous issue is discussed by Aurora Alegiani, Massimo Marraffa, and Tiziana Vistarini. They focus on Clark's active externalism, in particular, on the idea that the classical notion of mental representation must be replaced with a continuum of representational genera. In their view, this implies that we cannot adopt a single explanatory strategy but – depending on the specifics of the case – «must respectively appeal to the dynamical non-representational or the representational explanation» (cf. this issue, *infra*, p. 62). Organisms are constantly generating and updating internal models that predict incoming sensory information based on prior knowledge and expectations. These predictions are compared with actual sensory inputs, and the discrepancies (i.e., the prediction errors) are used to update and refine internal models. The authors discuss PP and compare Clark's “radical predictive processing” (rPP) in to the “conservative” view (cPP). Both cPP and rPP are reformist proposals that aim to reconcile the pragmatic turn in cognitive science with the desire to preserve a representational approach to cognition. However, while cPP is entirely representational (even though it does not conceive of representations in the traditional, “linguaformal” way), Clark's rPP describes a hybrid model. In Clark's view, cognition operates on the basis of both representational and nonrepresentational processes: low level processing is non-representational and model free while high level processing is representational. Yet, Alegiani, Marraffa and Vistarini question this hybrid architecture: «why resort to a hybrid notion if we do have a well-grounded notion of representation (that of cPP) that does not neglect the role of action, but rather emphasizes its prominence in our cognitive processes?» (cf. this issue, *infra*, p. 69). Of course, in the light of the previous works in this volume, we might wonder whether hybrid architectures should still be considered epistemologically weaker than homologous architectures.

A further major problem with the cognitive project addressed by several papers in the themat-

ic issue regards the reciprocal relationships among its constituent disciplines and their theories; in particular, the role each should play, which should lead the project, and why. Considering that cognitive science is directly engaged in studying *human cognition*, i.e., according to the definition offered by Gardner, «the nature of knowledge, its components, its sources, its development, and its deployment»,²⁴ cognitive psychology might seem the best qualified discipline to lead this project. Indeed, the title of Gardner's book (*The mind's new science*) presupposes that we already had a science of mind before the cognitive revolution; that old science of mind is empirical psychology.

However, it is not so obvious that guidance of the project necessarily belongs to psychology. As Gardner himself acknowledged, the topics addressed by cognitive science are long-standing epistemological questions that were traditionally investigated by philosophy. For this reason, philosophy could be the discipline to lead the cognitive project, while the latter could itself be seen as an improvement of previous philosophical research. In fact, the relationships between philosophy and psychology in the development of new theories of mind are quite complex and have always been highly debated. The reason why most descriptions of the cognitive science consider psychology instead of philosophy as the discipline appointed to develop new theories on cognition and cognitive processes is that – although not without difficulties or ambiguities – psychology employs the scientific method and this is a precondition for developing a *science* of mind.

And yet there is at least a third candidate that could provide leadership in this project. We might agree that linguistics, anthropology, and artificial intelligence cannot head up the cognitive project because: (i) linguistics focusses uniquely on a very specific aspect of cognition, i.e., language; (ii) anthropology describes the cultural diversity of human beings, and the aim of the cognitive science is exactly to show that this is only apparent or superficial and that it can be explained on the basis of universal mechanisms; and (iii) artificial intelligence serves as a (man-made) model to understand human cognition.

But neuroscience cannot be automatically considered as an ancillary discipline. In fact, one might wonder whether psychology depends on or can be reduced to neuroscience. In this case, the psychological agenda should be set by (cognitive) neuroscience while the study of cognition would be limited to the investigation of how knowledge is implemented by the brain. As a matter of fact, this idea would be perfectly compatible with the cognitive neuroscience revolution proposed by Worth Boone and Gualtiero Piccinini.

Several papers in this thematic issue question what role the various disciplines should play with-

in the cognitive project and which discipline should head up the cognitive project and why.

Philip Kargopoulos' paper, *What philosophy, if any, is needed for cognitive science?*, focusses on the role played by philosophy and on the changes that philosophy and philosophizing have undergone and will necessarily continue to undergo by participating in this project. First of all, Kargopoulos argues that philosophy has no "natural right" to be included in cognitive science just because the term "mind" was introduced by philosophers and because the mind was originally investigated by this discipline. It is also a philosophical illusion «that philosophy is needed to provide the ontology of the endeavor and to ensure some ontological orthodoxy or some methodological or other kind of correctness and oversight» (cf. this issue, *infra*, p. 29). In his view, we must justify why philosophy has something to offer for the advancement of the work of the other disciplines belonging to the cognitive project and we must clarify the relationship that philosophy entertains with psychology and their respective roles in cognitive science since both disciplines cultivate the ambition of elaborating theories about the mind. Kargopoulos argues that philosophy revolves around the notion of representation which is a tremendously complex topic that must be addressed by philosophical research. Indeed, this is why philosophy as a discipline has still a major role to play in the cognitive project. Moreover, philosophy has already shown that it is able "to propose small or grand hypotheses and theories" and to collaborate, especially with psychology and linguistics, in evaluating them: the modularity thesis and the language of thought are two prominent examples in this direction. From this perspective, at least some philosophical problems are not ordained to be everlasting, but may find specific solutions through (empirical and theoretical) research.

Max Coltheart's *Psychology is – and should be – central to cognitive science* claims that psychology should take the lead in the cognitive project. In Coltheart's view, the cognitive project has not become a cohesive interdisciplinary field over time, but this is not a failure since unity was never the intention of its founding fathers. To understand the very nature of this multidisciplinary field we must determine when a study conducted by two or more constituent disciplines counts as cognitive science. In fact, several lines of multidisciplinary research in these disciplines are not meant to investigate how the mind works but to solve other kind of problems: e.g., research on face recognition systems is not aimed at comprehending how humans recognize faces but at solving the practical problem of making artificial systems more efficient. Coltheart's thesis is that a multidisciplinary study conducted by two or more constituent disciplines can count as cognitive science only if cogni-

tive psychology is involved and this is the reason why «(cognitive) psychology is central to cognitive science» (cf. this issue, *infra*, p. 43). Ultimately, cognitive science is about mental processes and mental processes are the specific subject of cognitive psychology. One could argue that philosophy – which has not been taken into consideration by Coltheart – is another potential candidate to head such a coalition. The answer to this objection implied in Coltheart's reflection is that the study of mental processes should be based on observations and experiments and driven by the models developed by cognitive psychologists which offer effective explanations for phenomena such as cognitive impairments, recognition, performance of cognitive tasks, reasoning, etc.

Sandro Nannini's paper *The mind-body problem in philosophy and the cognitive sciences* offers an argument supporting the position that neuroscience should lead the cognitive project. In Nannini's view, especially after the so-called "second cognitive turn" in the 1990s – i.e., the cognitive neuroscience revolution we mentioned above – neuroscientific research became crucial for understanding the nature of the mind and cognitive processes, while the mind-body problem turned into a central issue of the cognitive project. In this paper, Nannini reviews the main solutions philosophers have proposed for the mind-body problem throughout history and considers how cognitive scientists have accepted or criticized these solutions. Eliminativism (specifically, "soft physicalistic eliminativism") is presented as the most convincing and suitable ontological-epistemological framework for promoting interdisciplinary and multidisciplinary studies among the various cognitive sciences. This "soft physicalistic eliminativism" builds on the work of the Churchlands and it assumes that folk psychology, cognitive psychology, and cognitive neuroscience operate within distinct "discourse universes" with different ontological commitments: folk psychology speaks of mental states, cognitive psychology describes psycho-functional states, and cognitive neuroscience focuses on neurological states. Psycho-functional states are idealized and partial redescriptions of mental states in the language of scientific psychology. Mental states are supposed to be reducible to psycho-functional states and psycho-functional states are, in turn, considered to be implemented by physico-chemical processes and identical to a higher-order property of brain dynamics. In this approach, «philosophers of mind suggest an ontological-epistemological framework capable of favouring interdisciplinary research in the field of "the science of mind" while the cognitive scientists evaluate whether this philosophical suggestion is of any use to them» (cf. this issue, *infra*, p. 128). Even though this framework reflects a policy of "divide and rule" by recognizing that philosophy, psychology and

cognitive neuroscience play a role in the cognitive project, in practice it endorses the absolute primacy of neuroscience in the cognitive project.

At least implicitly/tangentially, other contributions in this issue also take stand on the question of leadership. William Bechtel's *Rethinking cognitive architecture: A heterarchical network of different types of information processors* suggests that we should rely on brain structures to understand how the cognition of vertebrates works and this seems to be in line with the idea that neuroscience should play a central role in the cognitive project. Daniel C. Burnston and Antonella Tramacere's work on *Distributed loci of control: Overcoming stale dichotomies in biology and cognitive science* appears to go in a similar direction to that of Bechtel but to substantiate an even more radical position in which biology forms an integral part of the cognitive project. Biology is the discipline that reveals the deep structures from which we must draw inspiration to understand all levels of cognition.

Up until now, we have taken for granted that – beyond revolutions, plurality, and fragmentation – it still makes sense to speak of cognitive science or at least of the cognitive sciences and that the idea of joining forces for the scientific study of cognition is meaningful and deserves a future. However, even this minimalist view is not unanimous. In the final lines of his 2003 article titled *The cognitive revolution: A historical perspective*, referring to the founders of the cognitive program, Miller himself states that: «some veterans of those days question whether the program was successful, and whether there really is something now that we can call "cognitive science"».²⁵

In this thematic issue, in his "Cognition" – *Let's forget it?* Alan Costall takes the skeptical position that cognitive science was a project doomed to failure even before its inception. He challenges the very notion of cognition, claiming it is not only problematic to define but possibly also misleading and unscientific. As we have already seen the term cognition has a long history, reaching back to scholastic philosophy. Yet, until the 60s its use in psychology was «sporadic and selective» (cf. this issue, *infra*, p. 136). By the 80s the term was already widespread in psychological discussions and became essential for defining the very subject of psychological research. In Costall's view, the spread of this notion was not accompanied by any clarification of its meaning, which has remained obscure. In particular, it is not clear whether "cognitive" refers to particular psychological processes or a (meta)theoretical approach. This is one of the reasons why the cognitive revolution has not found unanimous consensus among psychologists. Rather than constituting a revolution, according to Costall, the cognitive turn represents a *re-turn* to the past and, in particular, to a form of a *modified behaviorism*. By the mid-50s it became clear

that «defining psychology as the science of behavior was like defining physics as the science of meter reading». The psychological lexicon could no longer avoid mentalistic terms. However, «we were still reluctant to use such terms as “mentalism” to describe what was needed, so we talked about cognition instead» (cf. this issue, *infra*, p. 139). Thus, according to Costall, the very notion of cognition is ambiguous and *empty*. Researchers in psychology mostly hung on to this notion because this it was at the core of their discipline and to abandon it seemed impossible. As he asks, «is it just possible that we ourselves don't really know what we are talking about when we talk about “cognition”. If so, let's forget it and find some better words» (cf. this issue, *infra*, p. 140).

What clearly emerges from the collection of these contributions is that – whatever it may be – the future of cognitive sciences is not easy to decipher or straightforward to predict.

Notes

¹ Cf. S.J. KEYSER, G.A. MILLER, E. WALKER, *Cognitive science – 1978: Report of the state of the art committee to the advisers of the Alfred P. Sloan Foundation*. On Sloan Foundation report(s) and the so-called “cognitive hexagon” cf., i.e., H. GARDNER, *The mind's new science. A history of the cognitive revolution*; W. BECHTEL, A. ABRAHAMSEN, G. GRAHAM, *The life of cognitive science*; M.A. BODEN, *Mind as machine. A history of cognitive science*; C. BAUM, *Stabilizing cognition: An STS approach to the Sloan Foundation report*.

² S. WALTER, A. STEPHAN, *Einleitung*, p. 2.

³ H.C. LONGUET-HIGGINS, *Comment by professor H.C. Longuet-Higgins on Lighthill report*.

⁴ *Ibidem*.

⁵ Cf., i.e., C.D. GREEN, *Where did the word “cognitive” come from anyway*.

⁶ Wolff's text is quoted according to T. STURM, H. GUNDLACH, *Zur Geschichte und Geschichtsschreibung der >kognitiven Revolution< - eine Reflexion*, p. 8: «cognitio(n) is an action of the soul by which it acquires a notion or idea of a thing for itself» (English translation by S.D. and L.P.).

⁷ Stout and Baldwin's entry is quoted according to T. STURM, H. GUNDLACH, *Zur Geschichte und Geschichtsschreibung der >kognitiven Revolution< - eine Reflexion*, p. 8.

⁸ Cf. S. WALTER, A. STEPHAN, *Einleitung*, pp. 1-2. Cf. also B.J. BAARS, *The cognitive revolution in psychology*, pp. 5-6; U. NEISSER, *Cognitive psychology*, pp. 9-12; H. GARDNER, *The mind's new science*, pp. 41-42; J. BRUNER, *Notes on the cognitive revolution*, p. 7.

⁹ U. NEISSER, *Cognitive psychology*, p. 4.

¹⁰ B.J. BAARS, *The cognitive revolution in psychology*, p. 6ff and p. 409ff.

¹¹ G. MANDLER, *What is cognitive psychology? What isn't?*, p. 9 (quoted according to H. GARDNER, *The mind's new science*, p. 29). Till the mid-1950s, psychology, especially experimental psychology, was still dominated by behaviorism.

¹² On the fragmentation and lack of unity in psychology cf., i.e., L.A.W. STAATS, *Psychology's crisis of disunity:*

Philosophy and method for a unified science; L.A.W. STAATS, *Disunity's prisoners, blind to a new approach to unification*. On *cognitive theory* as a new theoretical framework for overcoming the issue of lack of unity in psychology cf. Baars' reply to Staats' criticism.

¹³ Whether cognitive theories truly represented scientific revolutions in comparison with previous approaches is still an open issue. In general, cognitive scientists and psychologists view these cognitive theories as radically new compared to past theories, especially the behavioristic approach – cf., i.e., D.S. PALERMO, *Is a scientific revolution taking place in psychology?*. A more cautious, if not outright skeptical, attitude has been adopted by epistemologists and historians of science (cf., i.e., W. O'DONOHUE, K.E. FERGUSON, A.E. NAUGLE, *The structure of the cognitive revolution: An examination from the philosophy of science*; C.D. GREENWOOD, *Understanding the cognitive revolution in psychology*; L.B. BRISKMAN, *Is a Kuhnian analysis applicable to psychology?*).

¹⁴ Cf. D. MARR, *Vision. A computational investigation into the human representation and processing of visual information*. Cf. also A. NEWELL, *The knowledge model*; Z. PLYSHYN, *Computation and cognition. Toward a foundation for cognitive science*.

¹⁵ M. VILLALOBOS, D. SILVERMAN, *Extended functionalism, radical enactivism, and the autopoietic theory of cognition: prospects for a full revolution in cognitive science*, p. 720. Cf. also C. BUCKNER, E. FRIDLAND, *What is cognition? Angsty monism, permissive pluralism(s), and the future of cognitive science*; D. WILLIAMS, L. COLLING, *From symbols to icons: The return of resemblance in the cognitive neuroscience revolution*.

¹⁶ W. BOONE, G. PICCININI, *The cognitive neuroscience revolution*, p. 1510.

¹⁷ H. GARDNER, *The mind's new science*, p. 41.

¹⁸ J. BOBRYK, *The social construction of mind and the future of cognitive science*, p. 494.

¹⁹ Cf., i.e., A.M. GLEMBERG, J.K. WITT, J. METCALFE, *From revolution to embodiment: 25 years of cognitive psychology*.

²⁰ M. VILLALOBOS, D. SILVERMAN, *Extended functionalism, radical enactivism, and the autopoietic theory of cognition: Prospects for a full revolution in cognitive science*; M. VILLALOBOS, *Enactive cognitive science: Revisionism or revolution?*

²¹ R. ARDILA, *Toward unity in psychology*, p. 302.

²² Cf. K. STANOVICH, *How to think straight about psychology*; G. HENRIQUES, *The tree of knowledge system and the theoretical unification in psychology*; M.R. ROSENZWEIG, *Unity and diversity of psychology*; F.J. WERTZ, *Multiple methods in psychology: Epistemological grounding and the possibility of unity*; G. HENRIQUES, *Psychology defined*; D.L. KRANTZ, *Psychology's search for the unity*.

²³ A. NEWELL, *Unified theories of cognition*, p. 1.

²⁴ H. GARDNER, *The mind's new science*, p. 6.

²⁵ G.A. MILLER, *The cognitive revolution: A historical perspective*, p. 144.

Literature

ARDILA, R. (1992). *Toward unity in psychology: The experimental synthesis of behavior*. In: «International Journal of Psychology», vol. XXVII, n. 5, pp. 299-310.

BAARS, B.J. (1986). *The cognitive revolution in psychology*.

- gy, The Guilford Press, New York/London.
- BAARS, J.B. (1985). *The logic of unification*. In: «Contemporary Psychology», vol. XXX, n. 4, p. 340.
- BAUM, C. (2016). *Stabilizing cognition: An STS approach to the Sloan Foundation report*. In: «Theory and Psychology», vol. XXVI, n. 6, pp. 773-787.
- BECHTEL, W., ABRAHAMSEN, A., GRAHAM, G. (1998). *The life of cognitive science*. In: W. BECHTEL, P. GRAHAM (eds.), *A companion to cognitive science*, Blackwell, Malden/Oxford, pp. 1-104.
- BOBRYK, J. (2002). *The social construction of mind and the future of cognitive science*. In: «Foundations of Science», vol. VII, pp. 481-495.
- BODEN, M.A. (2006). *Mind as machine. A history of cognitive science*, vol. I, Oxford University Press, Oxford/New York.
- BOONE, W., PICCININI, G. (2016). *The cognitive neuroscience revolution*. In: «Synthese», vol. CXCIII, n. 5, pp. 1509-1534.
- BRISKMAN, L.B. (1972). *Is a Kuhnian analysis applicable to psychology?*. In: «Science Studies», vol. II, n. 1, pp. 87-97.
- BRUNER, J. (1984). *Notes on the cognitive revolution*. In: «Interchanges», vol. XV, n. 1, pp. 1-8.
- BUCKNER, C., FRIDLAND, E. (2017). *What is cognition? Angsty monism, permissive pluralism(s), and the future of cognitive science*. In: «Synthese», vol. CXCIV, n. 4, pp. 4191-4195.
- GARDNER, H. (1985). *The mind's new science. A history of the cognitive revolution*, Basic Books, New York.
- GLEMBERG, A.M., WITT, J.K., METCALFE, J. (2013). *From revolution to embodiment: 25 years of cognitive psychology*. In: «Perspective on Psychological Science», vol. VIII, n. 5, pp. 573-585.
- GREEN, C.D. (1996). *Where did the word "cognitive" come from anyway?*. In: «Canadian Psychology», vol. XXXVII, n. 1, pp. 31-39.
- GREENWOOD, C.D. (1999). *Understanding the cognitive revolution in psychology*. In: «Journal of the History of the Behavioral Sciences», vol. XXXV, n. 1, pp. 1-22.
- HENRIQUES, G. (2003). *The tree of knowledge system and the theoretical unification in psychology*. In: «Review of General Psychology», vol. VII, n. 2, pp. 150-182.
- HENRIQUES, G. (2004). *Psychology defined*. In: «Journal of Clinical Psychology», vol. LX, n. 12, pp. 1207-1221.
- KEYSER, S.J., MILLER, G.A., WALKER, E. (1978). *Cognitive science – 1978: Report of the state of the art committee to the advisers of the Alfred P. Sloan Foundation*.
- KRANTZ, D.L. (1987). *Psychology's search for the unity*. In: «New Ideas in Psychology», vol. V, n. 3, pp. 329-339.
- LONGUET-HIGGINS, H.C. (1973). *Comment by professor H.C. Longuet-Higgins on Lighthill report*. In: *Artificial intelligence. Lighthill report 1973: A paper symposium*, document available at URL: https://rodsmith.nz/wp-content/uploads/Lighthill_1973_Report.pdf
- MARR, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information*, Freeman & co., San Francisco (CA).
- MILLER, G.A. (2003). *The cognitive revolution: A historical perspective*. In: «Trends in Cognitive Sciences», vol. VII, n. 3, pp. 141-144.
- NEISSER, U. (1967). *Cognitive psychology*, Appleton Century Crofts, New York.
- NEWELL, A. (1982). *The knowledge model*. In: «Artificial Intelligence», vol. XVIII, n. 1, pp. 87-127.
- NEWELL, A. (1990). *Unified theories of cognition*, Harvard University Press.
- O'DONOHUE, W., FERGUSON, K.E., NAUGLE, A.E. (2003). *The structure of the cognitive revolution: An examination from the philosophy of science*. In: «The Behavior Analyst», vol. XXVI, n. 1, pp. 85-110.
- PALERMO, D.S. (1971). *Is a scientific revolution taking place in psychology?*. In: «Science Studies», vol. I, n. 1, pp. 135-155.
- PYLYSHYN, Z. (1984). *Computation and cognition. Toward a foundation for cognitive science*, MIT Press, Cambridge (MA).
- ROSENZWEIG, M.R. (1992). *Unity and diversity of psychology*. In: «International Journal of Psychology», vol. XXVII, n. 5, pp. 283-290.
- STAATS, L.A.W. (1983). *Psychology's crisis of disunity: Philosophy and method for a unified science*, Praeger, New York.
- STAATS, L.A.W. (1985). *Disunity's prisoners, blind to a new approach to unification*. In: «Contemporary Psychology», vol. XXX, n. 4, pp. 339-340.
- STANOVICH, K. (2001). *How to think straight about psychology*, Allyn and Bacon, Boston, 6th edition.
- STURM, T., GUNDLACH, H. (2013). *Zur Geschichte und Geschichtsschreibung der >kognitiven Revolution<*. In: A. STEPHAN, S. WALTER (Hrsgs.), *Handbuch Kognitionswissenschaft*, J.B. Metzler Verlag, Stuttgart/Weimar, pp. 7-21.
- VILLALOBOS, M. (2013). *Enactive cognitive science: Revisionism or revolution?*. In: «Adaptive Behavior», vol. XXI, n. 3, pp. 159-167.
- VILLALOBOS, M., SILVERMAN, D. (2018). *Extended functionalism, radical enactivism, and the autopoietic theory of cognition: prospects for a full revolution in cognitive science*. In: «Phenomenology and the Cognitive Sciences», vol. XVII, n. 4, pp. 719-739.
- WALTER, S., STEPHAN, A. (2013). *Einleitung*. In: A. STEPHAN, S. WALTER (Hrsg.), *Handbuch Kognitionswissenschaft*, J.B. Metzler Verlag, Stuttgart/Weimar, pp. 1-5.
- WERTZ, F.J. (1999). *Multiple methods in psychology: Epistemological grounding and the possibility of unity*. In: «Journal of Theoretical and Philosophical Psychology», vol. XIX, n. 2, pp. 131-166.
- WILLIAMS, D., COLLING, L. (2018). *From symbols to icons: The return of resemblance in the cognitive neuroscience revolution*. In: «Synthese», vol. CXCIV, n. 5, pp. 1941-1967.

RICERCHE

The future of cognitive science is pluralistic, but what does that mean?

Lisa M. Osbeck^(α) & Saulo de Freitas Araujo^(β)

Ricevuto: 2 aprile 2022; accettato: 8 dicembre 2022

Abstract We imagine the future of cognitive science by first considering its past, which shows remarkable transformation from a field that, although interdisciplinary, was initially marked by a narrow set of assumptions concerning its subject matter. In the last decades, multiple alternative frameworks with radically different ontological and epistemic commitments (e.g., situated cognition, embodied cognition, extended mind) found broad support. We address the question of how to understand these changes, noting as logical alternatives that (1) newer approaches are not properly *cognitive*; (2) that newer approaches are cognitive but not *science*; and (3) that cognitive science has become pluralistic. We endorse the third position and venture to guess that the future of cognitive science is also pluralistic. We are left, however, with the question of what this *means*. After noting the polysemous nature of the term “pluralism”, we attempt to add clarity by distinguishing three forms: ontological, epistemic, and ethical. We then consider what each form might imply for the future of cognitive science.

KEYWORDS: Cognitive Science; Pluralism; Relativism; History of Science; Philosophy of Science

Riassunto *Il futuro della scienza cognitiva è pluralista, ma che vuol dire?* – Pensiamo al futuro della scienza cognitiva in primo luogo considerando il suo passato, il quale è notevolmente cambiato rispetto al presente. Per quanto si proponesse come un campo di studi interdisciplinari, la scienza cognitiva delle origini era caratterizzata da un insieme ristretto di assunzioni riguardanti il proprio oggetto. Negli ultimi decenni hanno trovato supporto diverse cornici teoriche in reciproca competizione e con impegni ontologici ed epistemologici radicalmente differenti (si pensi, per esempio, alla cognizione situata, alla cognizione incarnata, alla mente estesa). Proveremo a dare risposta alla domanda su come intendere questi cambiamenti, prendendo atto che ci troviamo di fronte a una serie di posizioni che sono logicamente alternative fra loro: (1) gli approcci più recenti non sono propriamente *cognitivi*; (2) oppure che gli approcci più recenti sono cognitivi, ma non *scientifici*; (3) la scienza cognitiva è diventata pluralista. Noi crediamo che la terza posizione sia corretta e scommettiamo su un futuro della scienza cognitiva che sia anche pluralista. Resta aperto, tuttavia, il problema di cosa questo *significhi*. Dopo aver preso atto della natura polisemica del termine “pluralismo”, cercheremo di far chiarezza distinguendo tre forme di pluralismo: ontologico, epistemico, ed etico. Considereremo quindi ciò che ciascuna può comportare per il futuro della scienza cognitiva.

PAROLE CHIAVE: Scienza cognitiva; Pluralismo; Relativismo; Storia della scienza; Filosofia della scienza

^(α)Department of Anthropology, Psychology, and Sociology, University of West Georgia, 1601 Maple Street - 30118 Carrollton, GA (USA)

^(β)Departamento de Psicologia, Universidade Federal de Juiz de Fora, Campus Universitario s/n - 36036-900 Juiz de Fora, MG (Brazil)

E-mail: losbeck@westga.edu (✉); saulo.araujo@ufjf.br



IN *THE MIND'S NEW SCIENCE*, Howard Gardner identifies the key features, core assumptions, or “symptoms” of cognitive science as he found it in the mid-1970’s, the period during which he first encountered the term.¹ These features include, (1) analysis of *representations*, a distinct level of analysis between input and output focusing on relations between symbolic/syntactic entities.² The important idea is that this symbolic realm contains not only «acceptable scientific constructs» but is the appropriate focus for a science of mind; (2) a central relationship to the *computer*, both symbolically, as a model of human mentation, and procedurally, as a basis for designing and testing simulations of human mental processes; (3) in an attempt to limit analysis to what is clear and practical (i.e., not hopelessly complex), an *avoidance of “murky” concepts*, or more specifically a «de-emphasis on affect, context, culture, and history»;³ (4) commitment to and faith in the investigatory advantages of *interdisciplinary* cooperation, with the idea that «more powerful insights» are thus attained; (5) “rootedness” in philosophy, or rather, «classical *philosophical* problems» inherited from classical Greek and Enlightenment philosophers.⁴ By this last point Gardner means not that cognitive scientists actively engage in writing philosophical works but that long-standing philosophical questions serve as a «logical point of departure for investigations in cognitive science».⁵

We note that this set of core assumptions is descriptive only, as indicated by Gardner’s project in the first two sections of his book of offering a historical account of cognitive science in the latter half of the 20th century (to the mid-1980s). Moreover, he qualifies several of the assumptions he includes (e.g., by noting that cognitive scientists have nothing against context or affect in principle, or that the philosophical grounding of cognitive science is a contentious matter for some). In doing so Gardner implies that he offers only his own analysis, his view of what is essential to the emerging discipline, as a response to the fact that he found nothing systematic written on cognitive science prior to that time. Nevertheless, the core assumptions reflect his professional association with cognitive scientists of many stripes and his deep familiarity with the literature, thus it is reasonable to treat his postulates of cognitive science as reflecting real features of the “new science”,⁶ even if only strong trends.

Moreover, these trends and features are easily identified in the literature of the time. Although an abundance of exemplary references might be given, Fodor’s *Language of thought* and Newell and Simon’s physical symbol system hypothesis⁷ are among the best-known exemplars of the essential or definitive understanding of cognitive science within the period in question, illustrative of the received view Gardner conveys. Fodor was clear in his effort to develop the appropriate conceptual substrate for the rapidly expanding cognitive science, and alt-

hough his original *Language of thought* proposal (LOT) included the additional controversial inclusion of nativism, the central, most coherent argument is for the computational metaphor and stipulation that the computation in question takes place over a vast syntactical system of content according to rules. Thus, although Gardner’s project was historical, Fodor’s is conceptually foundational, with the features of cognitive science not regarded as arbitrary. On the contrary, the features appear to be binding and constitutive of the subject matter, most famously expressed in Fodor’s depiction of computation over representation as «the only game in town»,⁸ followed by the even bolder assertion that «there aren’t any alternatives which seem to be even remotely plausible».⁹

Of course, alternatives have since been forwarded and regarded by their proponents as highly plausible. By the later 1980s the emergence of “cognitive neuroscience” all but eviscerated the neglect of bodily processes in cutting edge cognitive science.¹⁰ Yet one fallout of the shift in focus to the cognitive neurosciences was diminishing collaboration with some of the efforts that were originally part of the broadly conceived interdisciplinary project of cognitive science (e.g., linguistics, cognitive anthropology, learning and education). Simultaneously, opposition to traditional representationalism and the related neglect of context (bodily and environmental) provoked a resistance movement at least in some such cognate disciplines. There we find a set of interrelated approaches that maintain a primary focus on cognition but offer models of cognition as blatantly interpenetrated with context – bodily, social, and environmental. With this difference in emphasis, epistemic priorities are also shifted, for example from concepts to activity. As one example, «A theory of situated cognition suggests that activity and perception are importantly and epistemologically prior – at a nonconceptual level – to conceptualization».¹¹ By the end of the 20th century, one can see only an expansion of alternatives united in opposition to *Good Old Fashioned Artificial Intelligence* (GOFAI): emerging, to some extent overlapping frameworks including situated cognition,¹² distributed cognition,¹³ embodied cognition,¹⁴ enactivism,¹⁵ extended cognition,¹⁶ neurophenomenology,¹⁷ and ecological psychology,¹⁸ among others.¹⁹ (Let us call these for convenience, the “dissenters,” because they depart from an orthodox cognitive science discernable in Gardner’s key features and Fodor’s swagger). As Buckner and Fridland put it: «virtually every aspect of this initial sketch of the nature of cognition and cognitive explanation has now been challenged».²⁰

In addition to posing challenges, new frameworks also introduce new concepts, such as “affordance”²¹ which has itself undergone a series of important changes over time.²² They also make possible an extension of the range of phenomena

considered to be within the province of cognitive science.²³ That the alternative frameworks also continue to evolve in response to academic challenges and technological advance is clear. For example, Gallagher describes three waves of development of the extended mind hypothesis since its first appearance at the end of the 20th century.²⁴ Important too is that some approaches appear to offer a bridge or middle ground between orthodox and newer perspectives on cognition. Some retain a principal role for representations and mental models while recognizing the participation of social, historical, or cultural processes in the broader systems within which cognition is accomplished,²⁵ or maintain that computation can be important in newer frameworks.²⁶ Thus, it is not a matter of one orthodox alternative currently positioned against one radical form of dissent; it is rather a broad array of options offered in attempt to address age old questions concerning means by which knowledge is acquired, problems are solved, and new innovations achieved. These problems are, of course, at once philosophical and psychological, concerning the nature of knowledge and the human processes by which it is accomplished. We may add that the (in)compatibility between divergent approaches is a matter of ongoing discussion and debate, though it is in large part beyond the scope of our present task.²⁷

Instead, the problem that orients this paper is this: how best to conceptualize the transformation from a single plausible alternative (computation over representations) to a range of options that do not share its most fundamental assumptions? We offer three logical possibilities for making sense of such a development in relation to cognitive science. The first is that the newer approaches are not part of *cognitive science*; the second that they are cognitive but not representative of cognitive *science*; the third is that *cognitive science* itself is *pluralistic*. We then defend the third alternative and discuss its broader implications in terms of the focus of the special issue – the future of cognitive science.

1 Ways of understanding transformations in cognitive science

There are certainly different ways to understand or to account for the transformations in cognitive science that have occurred in the past decades. Here, we focus on those views that seem to represent the best logical alternatives.

1.1 Alternative/dissenting approaches are not part of cognitive science

A claim to the effect that dissenting approaches are not part of cognitive science proper might be based on demonstration that “cognitive” has a narrow range of references and that these provide a standard for censure. The idea that the dissent-

ers are not part of cognitive psychology or cognitive science because they are not properly cognitive is promoted both explicitly and indirectly in various contexts, both on the part of the cognitive science orthodoxy and among the dissenters themselves.²⁸ An example from the orthodoxy is Adams and Aizawa,²⁹ who seek in part «to keep cognitive psychology on track»³⁰ by limiting its scope to genuinely, exclusively cognitive processes, a cause which in their view requires proposal of a distinguishing “*mark of the cognitive*”. Their question, in other words, concerns «what regions of spacetime contain cognitive processing?»³¹ Their answer advances two claims, the first that the cognitive is marked by the presence of at least *some* non-derived (non-normative) content. Derived content is that with social meaning: «traffic lights, gas gauges, and flags are paradigm cases of items bearing derived content. Thoughts, experiences, and perceptions are paradigm cases of items bearing non-derived content», but only if they include non-derived content as «original representations»,³² which they view as needed to enable explanation of a hypothetical “lone thinker” deprived of social interaction. A human born and raised with no social interaction and in a world with no socially determined symbols would still have thought (assuming we would call the person human); that thought would be representational like any other, but the representations in question would be not derived from social meanings. Because language is inherently social, non-derived content would be non-linguistic thought, or at least non-linguistic aspects of linguistic thought. Adams and Aizawa further identify “the cognitive” with the «many forms of information processing that manipulate and transform information in ways unlike those found in processes spanning the brain, body and environment»,³³ with vision, memory, and attention serving as examples of processes of this kind.

Adams and Aizawa are explicitly critical of dissenting approaches, especially the extended mind theses, their primary target. The extended mind thesis offers a direct challenge to individualistic frameworks, replacing it with a strong version of externalism that conceives of mind and environment as not merely coupled but *co-constituted*, and distinguished from philosophical externalism by the additional claim that the environment is actively involved in processing on an ongoing basis. It is, therefore, part of the very processing system that produces knowledge.³⁴ Adams and Aizawa’s main goal is to eliminate from the cognitive science orthodoxy extended mind frameworks that regard the cognitive system as inclusive of the material and social environment. Yet we find definitions of “cognitive” that would exclude dissenting frameworks also among the dissenters themselves. As one example, in a reply to an author he

accuses of making claims that answer the wrong questions, James Greeno identifies the situative approach as a perspective that contrasts with the cognitive, differing at the level of both concepts and explanatory schema. He notes:

The cognitive and situative perspectives differ in this. The cognitive perspective's basic concepts and explanatory schemata are about processes and structures that are assumed to function at the level of individual agents. These processes and structures include knowledge as well as perception, memory, inference, decision, and so on. The situative perspective adopts a different primary focus of analysis. Situativity focuses primarily at the level of interactive systems that include individuals as participants, interacting with each other and with material and representational systems.³⁵

Note that Greeno's contrast between cognitive and situative perspectives includes a difference in the unit of analysis "only".³⁶ He does not suggest that environmental or bodily factors are unimportant to cognition. Neither does the situative approach ignore representation, but the primary attention is to systems: persons in interaction with other persons and meanings they co-create and propagate.

Although Greeno focuses on the individual unit of analysis as essential to the cognitive perspective, for Adams and Aizawa "the cognitive" is defined more restrictively. Accordingly, in one case (for Greeno) "cognitive" is implicitly problematized as narrow in focus; in the other case (for Adams and Aizawa), the narrow focus of "cognitive" is affirmed and deemed necessary. We see, however, that even with these two initial examples representing two very different starting points, there are divergent views on the meaning or necessary parameters of "cognitive", and by extension cognitive science. Adams and Aizawa's effort to exclude extended mind is based in part on appeal to what has been traditionally regarded as "cognitive" or included in our understanding of "cognition". But this is less straightforward than might be imagined from their project. We are hardly the first to point out the inherent difficulty of narrowly defining "the cognitive" or restricting it to a slender range of processes. Indeed, the effort to define cognition and specify its referents is a problem that dates at least to early efforts to organize a distinct branch of cognitive study, at least two decades before the emerge of cognitive science as Gardner describes it.³⁷

We consult as an example an essay by Egon Brunswik titled *Scope and aspects of the cognitive problem*, published in 1957 and prepared for a *Cognition Symposium* held at the University of Colorado in May of 1955: «the first one in the field of psychology to concern itself with the important

and difficult area of cognition». ³⁸ Brunswik defines the problem of cognition simply as «the problem of the acquisition of knowledge», ³⁹ with perception of space and perception of personality characteristics (so called "social perception") as representative subproblems within this broad and enduring focus. His conclusion stresses that

only by detailed analysis of ecological [environmental] textures can the cognitive problem be restored from mere utilization problems to its full scope of achievement problems and thus again become the key to the core question of psychology, that of the adjustment of the organism to a complex environment.⁴⁰

We see here in an early effort to define the scope of cognitive psychology, at least, an emphasis on the environment, in stark contrast to the *de*-emphasis on environment Gardner had underscored as a key feature of the emergent cognitive science of the mid-1970s, cognitive science proper, he might say. Commenting on Brunswik's remarks and other papers in the symposium, Osgood acknowledges that although the authors are concerned with «the common problem of cognitive process», they exhibit «quite different emphases and with quite different levels of analysis in the constructs employed». ⁴¹ Perhaps most telling is that despite the core features he had discerned among its representatives, Gardner himself defines cognitive science as «a contemporary, empirically based effort to answer long-standing questions, particularly those concerned with the nature of knowledge, its components, its sources, its development, and its deployment». ⁴² Assuming Gardner's depiction of the field in the mid-1970s is accurate (which we do), it is perhaps most fitting to say that the definition of "cognitive" in psychology was originally broad, then narrowed with the advent of "cognitive science" as an interdisciplinary specialty, becoming more explicitly associated with on the processing of information and computational metaphor. However, by the early decades of the 21st century, the meaning appears to be broadening again. We find this to be affirmed elsewhere: «over time, the understanding of what it is to be cognitive has expanded, diversified, and become more contentious». ⁴³

Additionally worth noting is that there is an important distinction recognized between "cognitive" and "cognitivism," the latter implying a set of assumptions which range from the individual level of analysis Greeno cited and the doctrine of computation over representations alleged by an insistent Fodor. Various early critics of cognitivism in psychology make an implicit distinction between "cognitive" and cognitivism by emphasizing that the target of criticism is the *ism* itself – the doctrinaire stance, the dogmatic commitment to a

focus on internal representation and information processing, along with the assumption that all psychological processes should be understood from such a grounding.⁴⁴ Wertz, a methodologist and phenomenological psychologist, criticizes what he calls the “narrowly homogenous” group relying on cybernetics as a basis for psychology.⁴⁵ Moreover, Bruner, one of the participants of the cognitive revolution, denounces the reductionist framework of traditional cognitive science, and calls for the rescue of meaning and culture in cognitive psychology.⁴⁶ To these examples we may add the very long line of thinkers who preceded or kept pace with behaviorism and focused on a broad range of mental processes including thinking, attention, perception, and imagining.⁴⁷ We see nothing sound or productive about questioning whether such thinkers are properly cognitive in orientation or focus.

■ 1.2 *Alternative/dissenting approaches are not cognitive science*

A second possible strategy is to exclude alternative or dissenting approaches to cognitive science from the science aspect of cognitive science. That is, it is possible to grant that even though all approaches that focus on reasoning and problem-solving are at least broadly concerned with *cognitive* processes, their ways of *investigating* it are not properly scientific. Evidence in favor of this argument might point to the fact that some of the alternative approaches (e.g., extended mind theory) began as philosophical frameworks, offering conceptual models for cognition that introduce new concepts and procedures out of step with the practices of orthodox cognitive scientists or cognitive psychologists.

Such a view might be levied either from the standpoint of a scientific purism that defends the epistemic priority of a narrow conception of science *or* from the perspective of those who disparage science as inadequate to the task. For example, Gardner notes that some critics of cognitivism «hold that factors like affect, history, or context will never be explicable by science: they are inherently humanistic or aesthetic dimensions».⁴⁸ Gardner also acknowledges that other critics of cognitivism target a limited conception of science that would exclude context and affect, an «anti-septic cognitive science» as he puts it.⁴⁹ In either case, the concern is not only with definitions of science but with methods and epistemic validation, of epistemology rather than ontology.

We find the possibility that dissenting positions are not part of cognitive *science* even less viable than the first. This point does not deserve a long discussion; thus, we note only a few obvious reasons here. Leaving aside the fact that there is no settled opinion on the nature and limits of science itself, the rootedness of alternative or dissenting perspectives in philosophical frameworks can

hardly disqualify them as contributing to the science, either by precedent or logic. As Gardner makes clear, cognitive science is if anything a contemporary extension of philosophical problems of ancient origin. The whole enterprise may be conceived of as inherently and ultimately philosophical. Of course, as Gardner also notes, an empirical approach to these long-standing questions, the reliance on empirical data of various forms has also characterized the field. Thus, we might legitimately question the relation of the foundational philosophical argument to the collection or possibility of collection of empirical data to test the implications of the theoretical model.

But philosophical sophistication can hardly be used as grounds for questioning the empirical basis of a dissenting framework. To remain with the example of Clark,⁵⁰ he does not “collect data” in a sense traditional to psychology or cognitive science yet relies on evidence through thought experiments (e.g., the “Otto and Inga” problem) to draw the reader into appreciation of the theoretical affordances of the extended mind view and its superior fit for the complexities of a contemporary, technology dependent society. Indeed, we might characterize his approach as guided by the assumption that cognitive science frameworks require an upgrade precisely because they do not adequately fit the data generated by thought experiments.

Importantly, in seeking to theorize the nature of scientific reasoning, other cognitive scientists use case-based historical analysis.⁵¹ Practice-based approaches make use of ethnography.⁵² Thus, the second reason to avoid questioning the scientific legitimacy of these approaches is that although they make primary use of insight-based interpretive analysis rather than calculation, alternative models of cognition rely on data of different forms but are no less empirically grounded for so doing. As a final, not trivial point we may look at the self-identifications of those who advance alternative perspectives, not so much at specific disciplinary identity or job title (e.g., philosopher) but at the aims of the overall project to which they are contributing. For example, Greeno characterizes the situative perspective as «a broad collection of scientific work being developed by many people».⁵³ If the broader goal is to offer systematic analysis of any of the total range of psychological and interactive processes that contribute to phenomena we would label cognitive, we find it difficult to take seriously any claim that would position these efforts outside of science.

■ 1.3 *Cognitive science is pluralistic*

Rejecting the first two options, we are left with the problem that several ways of understanding learning and reasoning and problem solving have been refashioned into something almost unrecognizable from the field Gardner described as emerg-

ing in the mid-1970s. Cognitive science is altered ontologically, now including views that construe the cognitive processing system itself as broadened beyond the individual mind or brain, that regard cognition as an interactive phenomenon inclusive of the social and cultural worlds. It is also open to a new range of investigatory procedures, requiring a movement from principal reliance on computer simulation to field observation, interviews, case studies, thought experiment, and other modes of inquiry. Our position is that we should understand cognitive science as increasingly pluralistic over the past several decades, and further, that we see no reason to believe that this situation is likely to change in any future we can foresee. Hence our thesis is present in the first part of our title: the future of cognitive science is pluralistic.⁵⁴

The second part of our title sets the agenda for the remainder of the paper. That is, having made the claim that the future of cognitive science is pluralistic, what does that mean? What are the various ways in which this pluralism might be understood? What are the implications both for making sense of the dissenting and alternative frameworks within cognitive science over the past several decades and any additional alternatives likely to appear or evolve over decades to come? We begin to answer this question by first providing a brief historical overview. We will then describe several versions or forms of pluralism, including epistemic, ontological, and ethical, and discuss the ways and extent to which these forms are present in cognitive science in its current form and as we dare to predict its future. We will then attempt to add further clarity by discussing what pluralism is not. Doing so will include two main distinctions: (1) A distinction of pluralism and interdisciplinarity, and (2) a distinction of pluralism from relativism. We will then offer a tentative conclusion pointed to the future of cognitive science.

2 Meanings of pluralism: A brief history and overview

“Pluralism” is a polysemic term. Over the years it has evolved in so many directions that it is quite infeasible to present a single historical account of it. Nonetheless, it is possible to indicate some of its major features.⁵⁵

The common roots of all forms of pluralism can be traced back to the beginnings of Greek philosophy when the question about “the one and the many” was raised as a philosophical problem. Pre-Socratic philosophers disputed over the underlying unity or multiplicity of the world, although the term “pluralism” did not appear in this context.⁵⁶

As a philosophical term, it first appeared in the eighteenth-century German tradition, especially in the writings of Wolff and Kant.⁵⁷ For Wolff, pluralism is a metaphysical position, which admits

the existence of several beings in the world: «The idealists either admit more than one being or hold themselves to be the one real being. The former are called *pluralists*; the latter, *egoists*».⁵⁸ Kant, on the other hand, understands pluralism as an ethical thesis: «The opposite of egoism can only be *pluralism*, that is, the way of thinking in which one is not concerned with oneself as the whole world, but rather regards and conducts oneself as a mere citizen of the world».⁵⁹ From the beginning, then, we see that the term points to distinct concepts.

In the nineteenth century, the concept of pluralism gained new contours. It was extended beyond the purely metaphysical or ethical domain to grasp also the psychological, cosmological, and theological realms, resulting in a tripartite classification of pluralism.⁶⁰ In the second half of that century, debates over pluralism appeared in Britain and the United States. In this context, the great enemy to be faced was monism, after the influence of the German metaphysical tradition:

Coleridge, following Schelling sang of the “miraculous all,” the “omnific” spirit, the unity apprehended by reason and imagination, the human soul annihilated in God. De Quincey had studied Kant and Herder. Later on, Carlyle advised British youths to “open their Goethe” and offered for their silent admiration the indivisible nature of German philosophy.⁶¹

It happens that the battle over monism was not only a metaphysical, but also a theological, ethical, and political one, which led to the dissemination of pluralism as a philosophical label in religion and politics.⁶² Not surprisingly, in the first decades of the twentieth century, distinct concepts of pluralism were widespread in Europe and America.⁶³

After some decades of lying dormant, pluralism has reemerged over the past several decades. However, it remains poorly understood for the most part, largely because it emerges in very divergent contexts and in connection with a wide variety of phenomena in and out of psychology and philosophy.⁶⁴

By appealing to this brief overview, we just want to indicate that pluralism is not a fixed and static concept, or one that can be easily understood independently of the context in which it appears. On the contrary, the history of the concept shows that both unity and multiplicity have been conceptualized in different ways, giving rise to different concepts of monism and pluralism, which invite some kind of classification or systematization.

3 Forms of pluralism

Out of several ways in which a field may invoke pluralism, we want to highlight three special forms: ontological, epistemic, and ethical. We consider each

of these in a general way before turning to the question of what each form would mean for cognitive science, present and future, and thus what is entailed in our suggestion that the future of cognitive science is pluralistic in each of these ways.

■ 3.1 *Ontological pluralism*

Traditionally, ontological pluralism is associated with the assertion that the world is not an organic unity; instead, there are in a sense multiple fields that may or may not interact at any given time. According to William James, pluralism means that «the substance of reality may never get totally collected, that some of it may remain outside of the largest combination of it ever made». ⁶⁵ Thus, he continues, «things are “with” one another in many ways, but nothing includes everything, or dominates everything [...] Something always escapes». ⁶⁶ In this sense, ontological pluralism implicates not only ontological diversity, but also dynamism and transformation. ⁶⁷

We noted the variety of positions represented by contemporary frameworks that dissent from the orthodoxy of the computation over representation (GOFAI) model (e.g., situated, embodied, embedded, enactive, extended cognition, neurophenomenology). Yet despite the fact that cognitive science displays a set of ontologically diverse positions, ⁶⁸ explicit philosophical commitments to ontological pluralism as specifically relevant to cognitive phenomena are not common.

Contemporary descriptions of ontological pluralism are more prominent in philosophy of biology, and from these we may draw out its relevant features and reflect on how they apply to cognition in the broad sense we advocate. One of the most comprehensive recent articulations of pluralism in biology is Sandra Mitchell's *Unsimple truths*. ⁶⁹ Mitchell begins her argument with an appeal to complexity as a primary characteristic of the natural and human worlds. Although Mitchell focuses principally on epistemic pluralism, which we discuss in the next section, we can glean from her discussion that a contemporary version of ontological pluralism is expressed in any philosophical position that emphasizes the nature of reality as a vast array of complex *systems*: material, social, and temporal in structure and composition.

Mitchell provides examples of natural phenomena (a honeybee colony, an eukaryotic slime mold that exhibits two different states of existence, psychiatric disorders) that exhibit different forms of complexity: «compositional, dynamic, and evolved». ⁷⁰ These forms, in other words, reflect (respectively) complexity in organization, and complexity in causal interactions, both internal to the system and with context variation, the latter representing what she calls «evolved contingency». ⁷¹ This latter form of complexity refers

to the «uniqueness as a characteristic of complex biological systems» that distinguish them from the complexity of physical structures unaffected by evolutionary mechanisms. ⁷² Thus, important to Mitchell's analysis is that we must understand complexity as «variegated in kinds». ⁷³ Complexity of organization alone is further variegated by forms of organization evidenced at different levels, such as genetic and biochemical.

■ 3.1.1 *Ontological pluralism and cognitive science*

Given the dynamic nature and neurological substrate of cognitive phenomena, biological complexity is a more fitting analogy for the subject matter of cognitive science than is physics. The analogous framework provided by philosophy of biology enables us to analyze what pluralism as the ontological foundation of cognitive science would seem to imply and require.

First, it would foreground the complexity of cognitive phenomena by virtue of the natural and social systems within which cognitive phenomena emerge, and, drawing from Mitchell's distinctions, it would entail differentiation of various kinds of complexity in cognitive systems, including different levels of organization and differences between systems owing to evolutionary diversity (e.g., between minds of humans and other animals, or between minds of different representatives of a species). The important element in this assertion is “foreground”, how the relation between cognition and context is theorized, to what extent context is regarded as figure or ground, or whether a figure-ground organization is even applicable. We find very few, if any, contemporary positions that would deny that context is important to cognition, even those that promote more traditional approaches. For example, Adams and Aizawa, ⁷⁴ although seeking to limit the definition of cognition, assert that it is now uncontroversial as to whether cognition is impacted by bodily or social context, and, as noted, Gardner acknowledged the acceptance of this fact on the part of cognitive scientists even in the mid-1970s. In the case of Adams and Aizawa, the question is not whether cognition is impacted by context but instead whether there is at least in principle some aspect of cognition *not* impacted, something pure and original to an individual mind, without social influence of any kind.

It is important to acknowledge that there are two different ways of understanding pluralism as a foundation for cognitive science that arise at this juncture. The first concerns the nature of cognition itself, emphasizing the inherent complexity of the phenomena, the variegated nature of the complexity itself.

The second concerns the theoretical frameworks that in some sense compete to organize our

understanding of cognitive phenomena, the pluralism of the science that endeavors to make sense of the phenomena. Thus, the second sense concerns cognitive science as a field. We bring this up in part to underscore that pluralism can be understood in different ways, complicating the task of offering an account of the meaning and implications of pluralism in cognitive science. However, that these senses overlap is also indisputable. For example, an important question that arises is whether pluralism in the first sense (pertaining to the subject matter or referent) can accommodate pluralism in the second sense, that of acknowledging that different frameworks make sense of cognitive science in different ways. For example, the dynamical approach to cognition is among those which suggest that it cannot. In this case the focus is on the irreducibly temporal structuring of cognitive systems, for which reason computational models are inadequate:

Cognitive processes and their contexts unfold continuously and simultaneously in real time. Computational models specify a discrete sequence of static internal states in arbitrary “step” time (t_1 , t_2 , etc.). Imposing the latter onto the former is like wearing shoes on your hands.⁷⁵

In short, an ontology emphasizing complexity and interacting systems at the level of the subject matter (ontological pluralism) also has unavoidable epistemic (and relatedly, methodological) implications, to which we turn next.

3.2 Epistemic pluralism

Acknowledging the inordinate complexity of the world does not require the world is unknowable, only that the task of knowing it is itself complex. Epistemic pluralism is thus an attempt to deal with such complexity, by accepting that there are several ways to make sense of it, and that none is capable alone to do the job. According to James, «when we reach more complex facts, the number of ways in which we may regard them is literally countless. They are perfect well-springs of properties, which are only little by little developed to our knowledge».⁷⁶ Epistemic pluralism thus rejects the idea of unification of science around a single explanatory model.⁷⁷

For Mitchell, a complex world begets complex representations of it, which we may understand in part as representations (models, theories) aimed at different levels or different aspects of the phenomenon of interest. Important to note, however, is that the existence of divergent models of a phenomenon, and correspondingly, different schools of thought in relation to a discipline does not itself constitute epistemic pluralism. The broad history of philosophy and psychology, including cognitive

science, may be seen as a panoply of alternative conceptions of the mind and world and their relations, both over time and at any given period. Epistemological pluralism concerns the *stance* we should take toward the alternatives, more specifically, «the appropriate stance we should take toward the models, theories, and explanations proffered by scientists».⁷⁸ In turn, there are closely related methodological implications, concerning the type and range of methods accepted as appropriate to the level of analysis or framework employed. Epistemic pluralism also implies that any given starting point for making knowledge claims is not absolute, that is, it is limited by inherent constraints, by which the vantage point offers a necessarily limited view. A central question seems to concern whether the relationship between alternative models/theories/explanations/methods must be one of competition or whether they may be fruitfully regarded as complementary. Epistemic pluralism suggests complementarity is possible, but such an assumption leaves certain questions unanswered, such as whether alternative frameworks can be sufficiently “commensurable” to enable true complementarity.⁷⁹ To this question we will return in relation to cognitive science and its future. First, we consider some of the statements of epistemic pluralism that have appeared in recent decades in psychological literature.

3.2.1 Epistemic pluralism and cognitive science

In cognitive science, the term “explanatory pluralism” is more common than “epistemic pluralism,” but the meaning is essentially same:

[w]e use many theories to understand the universe. This basic idea, often called explanatory pluralism, derives from the diverse levels of organization in the universe, and the equally diverse explanatory goals of human beings.⁸⁰

Like Mitchell, however, we would note that pluralism traditionally includes more than the observation that different humans use different explanations and includes the further assumption that this multiplicity is beneficial in the pursuit of knowledge. For example, a special section of *Theory & Psychology* concerns the relevance of explanatory pluralism to psychology. The editor’s introduction presents explanatory pluralism as «an alternative to reduction and eliminativism, on the one hand, and methodological dualism on the other».⁸¹ The key to this alternative is to regard theories as not mere rivals in competition for dominance but as aimed at different levels, different units of analysis of the same complex phenomena. Explanatory pluralism «holds that theories at different levels of description, like psychology and neuroscience, can co-evolve and mutually influence

each other, without the higher-level theory being replaced by, or reduced to, the lower level one».⁸² Similarly, McCauley and Bechtel find in explanatory pluralism «a middle ground between the theoretical and ontological parsimoniousness of reductionists, on the one hand, and the metaphysical extravagances of antireductionists, on the other».⁸³

Does such a middle ground fit the challenges of cognitive science's history of orthodoxy and dissent and its and contemporary debates and alternative frameworks? Although no longer regarded as the only game in town, reductive approaches to cognitive science continue in a position of some prominence, now more frequently in relation to brain-base processes and neurological explanations. Can one reconcile reductionist approaches with those that foreground context, culture, and temporality? As earlier noted, both insistent individualist and reductionist frameworks⁸⁴ and those that oppose computational models⁸⁵ seem to find no room for compromise between the extremes. Yet elsewhere one finds positive statements that such reconciliation is possible, both in Mitchell's account and in relation to cognitive science specifically. Mitchell claims that reductive strategies have a place in pluralism as she conceives it. That is, some of the world's causal structures may afford description in terms of reductive accounts and offering such accounts may serve certain purposes to advance knowledge.

For cognitive science, we may draw as an implication from Mitchell's allowance that there is room for reductive models of some mental phenomena, including the conception of "non-derived" cognition, for example, however far removed from the complex, context-sensitive models more fitting to cognition "in the wild".⁸⁶ Fabry for example, considers recent predictive processing (PP) accounts (which emphasize «continuous attempt to minimize prediction error»)⁸⁷ and seek to limit explanation to internal, neural-based processes at the exclusion of the kinds of processes theorized in 4E approaches (that theorize cognition as cognition is embedded, extended, embodied, and enactive). Obviously the ontological and epistemological assumptions characteristic of internalist and 4E frameworks display radical differences. They entail divergent models for understanding cognition, have radically different theoretical starting points, and invite very different modes of inquiry. But Fabry believes a form of reconciliation is possible through explanatory pluralism, understood as «careful combination of these distinct scientific explanations» to move in the direction of ever more complete models of cognitive phenomena, which in turn increases explanatory potential.⁸⁸ Similarly, McCauley regards pluralism as a more comprehensive approach:

[p]luralism underscores how the sciences inte-

grate information about patterns that systems exhibit not just with that available at lower levels about those systems' parts but also with that at higher levels. Inquiry at those higher levels takes up factors influencing those parts' organization and workings and examines both the settings in which a system may be situated and the various external factors that constrain its shape and inputs.⁸⁹

However, reduction becomes problematic if used exclusively and reflexively, independent of specific goal or for phenomena that do not lend themselves to reductive models. The important question for cognitive science is to determine what aspects of cognition are best suited to reductive models and which are not, relating always to a specific goal rather than as a binding framework imposed on all its phenomena and sanctioning its modes of inquiring.

We note, however, that such complimentary and co-participatory harmony in the name of progress may not be embraced by those who forward reductive strategies, or perhaps by those who oppose them. Can we assume that those who currently forward reductive frameworks will invite relegation to the status of offering a limited, purpose-tied role in explanation of cognitive phenomena? Will concession be made to context in such a way that new models and methods are openly embraced by reductionists? What evidence suggests that this is likely? And if not, is tolerance for reductionistic accounts on the part of those who emphasize complexity likely to be reciprocal, and what are the epistemic consequences if it is not?

Another way of posing the question is that if divergent approaches are complementary, what such complementarity might mean and how it might be achieved. This has not yet been settled within cognitive science; that must remain the task of its future. Mitchell's answer is that it involves overall progress in knowledge, and she favors what she identifies as a pragmatist framework conceptualizing the goal of pluralistic inquiry. A great many implicit assumptions are raised by an appeal to *pragmatism* as a framing for the pluralism most relevant to cognitive science and its future. As one example, classical articulations of pluralism⁹⁰ incorporate an emphasis on constraints on knowledge imposed by the limitations of the vantage point or perspective from which one views the world, a vantage point that is impacted by many things including disposition and *values*, both on an individual and collective level. Moreover, an appeal to *progress* requires us to define not only our starting points but also our eventual endpoint, especially when the phenomenon to be understood concerns the nature and functioning of human thought. How is progress to be conceptualized, and what are to be used as markers against which it may be measured?⁹¹ We

cannot address such questions without reference to broader goals and to the values that inflect the very goal of epistemic progress. Therefore, to invoke pragmatism and progress it is also necessary to conceptualize ontological and epistemic pluralism as overlapping with a third form: ethical pluralism.

3.3 Ethical pluralism

Ethical pluralism is the view that there are several valid moral values and, accordingly, several ways to act rightly in the world. For example, Isaiah Berlin claims that «there are many objective ends, ultimate values, some incompatible with others, pursued by different societies at various times, or by different groups in the same society, by entire classes or churches or races, or by particular individuals within them».⁹²

Although implicated by explanatory pluralism for the reasons noted, explicit discussions of ethical pluralism relevant to science are not easy to locate. Therefore, we will say less on this form of pluralism than we have said on the others. Even in our analogous study of philosophy of biology we must search a bit for the proper context. Mitchell does not openly extol ethical pluralism but she devotes a chapter to discussion of policy, equates policy and decision making with concern for “how we act in the world” as well as the consequences of actions and the values that guide appraisal of those consequences.⁹³ The question of “how we act in the world” (or should act) brings us into the realm of ethics; by linking this issue to complexity, by extension we come close to an implication of ethical pluralism. The point is that if we understand our subject matter as ontologically and epistemically pluralistic, there are ethical implications of this view. Thus, Mitchell identifies such a pursuit (policy, understood as relevant to action in the world) as «perhaps the most important way in which an understanding of complexity may revise our thinking about the world».⁹⁴ On her account, attunement to ontological complexity requires revisiting standard strategies for policymaking by more centrally encompassing an emphasis on «eliminable risk»,⁹⁵ emphasizing the challenges that accompany such an emphasis.

3.3.1 Ethical pluralism and cognitive science

Relevant policy making dilemmas confronting cognitive science may seem more remote than those applying to biology, but this can hardly be the case on closer examination, wherein we recognize cognitive science as involving knowledge production itself, and indirectly its seemingly unbounded set of possible applications. More directly, the overlap of cognitive science with value-laden pursuits such as education and climate science brings a focus on values glaringly into view.

That is, the way we understand cognition and cognitive systems and how this knowledge might be applied to policy implicates values. Values are expressed in and through policy, that is. A specific set of values that should structure and limit the applications of cognitive science is beyond the scope of our analysis. We hope only to make clear that ontological and epistemic pluralism ultimately overlap with questions concerning ethical pluralism, and we hope to bring emphasis to the importance of this implication. At a minimum, to acknowledge that the future of cognitive science is pluralistic calls for more open attention to and discussion of value - cultural, moral, personal, and political - in all the ways these impact our understanding and application of the subject matter. We find this to be in keeping with comprehensive recent analyses of the place of values in relation to science and policymaking more generally, such as the excellent work by Heather Douglas.⁹⁶ We also consider ethical pluralism in cognitive science to encourage a value of epistemic humility or modesty, given the inherent limitations and constraints on all representations (models and theories) of what may be an infinitely complex world, and thus to be clearer as to the limited aims each model might support.

4 What pluralism is not

After offering our assessment that the future of cognitive science is pluralistic, we endeavored to provide a broad account of what that might entail and imply, that is, to address the question raised in our paper’s title, “but what does that mean”? We hope that in so doing, we have added some clarity, but also perhaps underscored that the meaning of pluralism is itself complex, that it is not one thing but many. In the interests of attempting to clarify pluralism’s meanings, one final task is to acknowledge two things it is not. That is, we briefly discuss two concepts with which pluralism may be conflated.

4.1 Pluralism vs. interdisciplinarity

Cognitive science has always been an interdisciplinary effort, but it has not always been pluralistic. This obviously implies an important difference between the two concepts, which we will attempt to describe here.

As we have seen, interdisciplinarity is a defining feature of cognitive science.⁹⁷ In saying that cognitive science was originally interdisciplinary, Gardner’s main point was that it involved collaboration and conceptual input from various fields from its earliest stages of development. Yet his analysis of the core features of cognitive science (again, as found in the 1970s), is rather essentializing, highlighting a set of shared ontological and epistemological assumptions upon which the original disciplinary starting points eventually con-

verged. The implication is that the various disciplines contributing to cognitive science, or rather, in his term, “cognitive sciences” – philosophy, linguistics, anthropology, neuroscience, artificial intelligence, and psychology⁹⁸ – share the set of core assumptions or features named. Whether or not this assumption holds, we make this point only to note that interdisciplinarity alone is no guarantee of pluralism. Similarly, a single discipline may exhibit ontological and epistemic pluralism without explicitly promoting interdisciplinarity, as Mitchell so powerfully demonstrates for biology, and has been claimed in relation to the discipline of psychology.⁹⁹

In recent years research and theory on interdisciplinarity has exploded, along with the cognates of multidisciplinary and transdisciplinary,¹⁰⁰ so ongoing conversation about *both* overlap and distinctions between interdisciplinarity and pluralism gains new significance. This topic is thus too large to analyze in detail here. We only note that pluralism as an ontological and epistemic stance is conceptually distinct from disciplinary practices and configurations, whatever their aims.

4.2 Pluralism vs. relativism

Finally, we must point out, as does Mitchell,¹⁰¹ that pluralism is properly distinguished from relativism and should not be understood as entailing or requiring relativism. We acknowledge, first, that relativism remains for many a dirty word, yet it has appeared in varied forms historically and with different historical starting points,¹⁰² which complicates any attempt at a universally valid distinction between pluralism and relativism. However, this does not mean that we should not even attempt it.¹⁰³ For example, a first cut may be made between a form of relativism that rejects the privilege of *any* vantage point, leading to skeptical conclusions, and a form of relativism based on an assertion that *all* vantage points (and their knowledge claims) are equally valid – “anything goes” or passes as valid knowledge. In this sense, neither of these forms is required or implied by pluralism.¹⁰⁴

Instead, pluralism underscores the inherently limited nature of all claims to knowledge and a commitment to viewing knowledge as in eternally in progress. As Mitchell emphasizes, the world’s complexity does not mean that it cannot be understood in principle, or to greater degrees, or that we lack the tools to understand it in practice.¹⁰⁵ Similarly, it does not imply that one claim to knowledge is no more valid than another. Instead, pluralism entails the much less problematic assumption that our knowledge is necessarily limited, that our models of its nature should be complex, tentative, and qualified, but more optimistically, that our knowledge of the complex world is ever evolving, empowered by multiple methods of inquiry and levels of analysis.

With regards to cognitive science specifically, Colin Alen uses the label “relaxed pluralism” to describe his “stance” toward the field, implicitly drawing a contrast to at least extreme forms of relativism:

It is pluralistic in the sense that it tolerates different ways of selecting which natural phenomena are appropriate targets for investigation within the science, even when they make incompatible judgments about cases. And while relaxed it is not “lazy” – that is, not just anything goes. Rather, the point is that enquiry should not be stifled by a conservatism about terms and their meanings that insists on stipulating what we are studying before we study it, especially when this conservatism is coupled with an introspectively-based claim to knowledge of the subject of enquiry that is highly resistant to empirical adjustments. Nevertheless, the relaxed but industrious pluralist must offer an explanation for why these apparently incompatible ways of carving up the phenomena do cohere, rather than driving the discipline toward disintegration.¹⁰⁶

Admittedly, a wide variety of definitions and classifications of both pluralism and relativism add complexity to their relationship and require finer-grained analysis of the stances or arguments defended than we are able to offer here. Our principal point is that facile conflation of pluralism with relativism is problematic.

5 Summary and conclusions

In contemplating the future of cognitive science, we are drawn to consideration of its past. Fritz Heider, assigned the role of discussant for the papers comprising what was framed as the first symposium on cognition in contemporary psychology in 1955,¹⁰⁷ was asked to offer comment on the “whence, what and whither” of cognition:

The first question – whence? – concerns the history of the problem of cognition; the second – what? – refers to some present orientations toward the problem; and the third – whither – is asked in order to stimulate some considerations about its possible future fate.¹⁰⁸

Although he offered analysis of the current state of the study of cognition (the “what”) and pronounced it in good form, he refrained from making predictions about its future, calling such a task one of «making prophecies».¹⁰⁹

We have been more willing than was Heider to make prophecies in this paper, in asserting our view that the future of cognitive science is pluralistic. We do so, however, with the benefit of the

“whence”, that is, with a view of the historical path from the computational framework as the only robust alternative to a contemporary landscape of alternatives with different ontological and epistemic foundations. In seeing no signs of diminishing appreciation for the complexity of the matter, we see no reason to regard the future as anything but pluralistic in turn.

In addition to making our pronouncement that the “what and whither” is pluralistic, we also used the opportunity of this paper to offer a bit of clarification as to what this might mean. We considered ontological and epistemic pluralisms in turn. Although we acknowledged that explicit discussions of ethical pluralism as relevant to cognitive science are in short supply, we were able to draw implications from analogous discussions of pluralism in other fields, especially philosophy of biology. We hope to be forgiven for preaching that the time for such reflection was seldom more apparent. With this emphasis, the intent of our paper is only to encourage further reflection, not to offer a final word on pluralism or more precisely focused analysis of the future, which remains open.

Notes

¹ H. GARDNER, *The mind's new science*, p. 38.

² Note that there are several variations of representational theory positing different relations to that which is represented (e.g., D. PITT, *Mental representation*; N. SHEA, *Representation in cognitive science*; B. VON ECK-ARDT, *The representational theory of mind*).

³ H. GARDNER, *The mind's new science*, p. 41.

⁴ *Ibid.*, p. 42.

⁵ *Ibid.*, p. 43.

⁶ H. GARDNER, *The mind's new science*, p. xiii.

⁷ Cf. J.A. FODOR, *The language of thought*; A. NEWELL, *Physical symbol systems*; H.A. SIMON, *Studying human intelligence by creating artificial intelligence*.

⁸ J.A. FODOR, *The language of thought*, p. 27.

⁹ J.A. FODOR, *Methodological solipsism considered as a research strategy in cognitive science*, p. 63.

¹⁰ Cf., e.g. R.P. COOPER, T. SHALLICE, *Cognitive neuroscience: Origins and promise*; D. STANDAGE, T. TRAPPENBERG, *Cognitive neuroscience*.

¹¹ J.S. BROWN, A. COLLINS, P. DUGUID, *Situated cognition and the culture of learning*, here p. 41.

¹² Cf. J. LAVE, E. WENGER, *Situated learning: Legitimate peripheral participation*.

¹³ Cf. E. HUTCHINS, *Cognition in the wild*.

¹⁴ Cf. G. LAKOFF, M. JOHNSON, *Philosophy in the flesh*; F.J. VARELA, E. THOMPSON, E. ROSCH, *The embodied mind: Cognitive science and human experience*.

¹⁵ Cf. C. BAERVELDT, T. VERHEGGEN, *Enactivism*.

¹⁶ Cf. A. CLARK, D. CHALMERS, *The extended mind*.

¹⁷ Cf. F.J. VARELA, *Neurophenomenology: A methodological remedy for the hard problem*.

¹⁸ Cf. J.J. GIBSON, *The ecological approach to visual perception*.

¹⁹ For more detail, cf. L. OSBECK, *Transformations in cognitive science*.

²⁰ C. BUCKNER, E. FRIDLAND, *What is cognition?*, p.

4191.

²¹ Cf. J.J. GIBSON, *The theory of affordances*; J.J. GIBSON, *The ecological approach to visual perception*.

²² Cf. A. CHERO, *An outline of a theory of affordances*; E. COSTA, *Affordances-in-practice: An ethnographic critique of social media logic and context collapse*; R. WITHAGEN, H.J. DE POEL, D. ARAÚJO, G.J. PEPPING, *Affordances can invite behavior: Reconsidering the relationship between affordances and agency*.

²³ For example, a forthcoming edited volume on the topic of ignorance proceeds from an embodied, extended mind framework (S. ARFINI, L. MAGNANI, *Embodied, extended, ignorant minds. New studies on the nature of not-knowing*).

²⁴ Cf. S. GALLAGHER, *The extended mind: State of the question*.

²⁵ Cf., e.g., A. CLARK, D. CHALMERS, *The extended mind*; E. HUTCHINS, *Cognition in the wild*; N.J. NERSESSIAN, *Interdisciplinarity in the making: Models and methods in frontier science*.

²⁶ Cf. M. VILLALOBOS, J. DEWHURST, *Why post-cognitivism does not (necessarily) entail anti-computationalism*.

²⁷ Cf. M. MAIESE, *Can the mind be embodied, enactive, affective, and extended?*; A. SCARINZI, (2020). *4Es are too many: Why enactive world making does not need the extended mind thesis*.

²⁸ We should note that the orthodoxy has evolved from the mid-1970s, most notably by incorporation of neuroscience but also to theorize the contribution of the environment. Adams and Aizawa acknowledge these developments yet still seek a principled basis upon which to identify genuinely cognitive phenomena, and to portray the mind as interacting with the environment but not “constituted” by it (cf. F. ADAMS, K. AIZAWA, *The bounds of cognition*).

²⁹ Cf. F. ADAMS, K. AIZAWA, *The bounds of cognition*; F. ADAMS, K. AIZAWA, *Why the mind is still in the head*; F. ADAMS, K. AIZAWA, *Defending the bounds of cognition*.

³⁰ F. ADAMS, K. AIZAWA, *The bounds of cognition*, p. ix.

³¹ *Ibid.*, p. 6.

³² *Ibid.*, p. 32.

³³ *Ibid.*, p. 31.

³⁴ Cf. A. CLARK, *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*; A. CLARK, *Intrinsic content, active memory and the extended mind*; A. CLARK, D. CHALMERS, *The extended mind*; R. MENARY, *Introduction: The extended mind in focus*.

³⁵ J.G. GREENO, *On claims that answer the wrong questions*, p. 7.

³⁶ We say “only” in quotation marks because this is a fundamental and transformative difference between the two perspectives.

³⁷ We could say that recent debates illustrate not only the complexity of the problem, but also the growing disagreements about how to solve it (see, e.g., A. JANSSEN, C. KLEIN, M. SLORS, *What is a cognitive ontology, anyway?*; R. PAIN, *Phenomenology and cognitive neuroscience: Can a process ontology help resolve the impasse?*; C.J. PRICE, K.J. FRISTON, *Functional ontologies for cognition: The systematic definition of structure and function*).

³⁸ K. MUENZINGER, *Introduction*, p. 1.

³⁹ E. BRUNSWIK, *Scope and aspects of the cognitive problem*, p. 6.

⁴⁰ *Ibid.*, p. 30.

⁴¹ C.E. OSGOOD, *Discussion of paper by Egon Brunswick*, p. 33.

⁴² H. GARDNER, *The mind's new science*, p. 6.

⁴³ A. ABRAHAMSEN, W. BECHTEL, *History and core themes*, p. 9.

⁴⁴ Cf., e.g., B. DE GELDER, *The cognitivist conjuring trick or how development vanished*; R. MONK, *Cognitivism and cognitive psychology*; A. STILL, A. COSTALL, *Introduction: In place of cognitivism*, pp. 1-11; A. STILL, A. COSTALL (eds.) *Against cognitivism: Alternative foundations for cognitive psychology*.

⁴⁵ F. WERTZ, *Cognitive psychology and the understanding of perception*, p. 104.

⁴⁶ Cf. J. BRUNER, *Acts of meaning*.

⁴⁷ Cf. W. WUNDT, *Erlebtes und Erkanntes*; J. DEWEY, *How we think*; W. JAMES, *The principles of psychology*; M. WERTHEIMER, *Produktives Denken*.

⁴⁸ H. GARDNER, *The mind's new science*, p. 42.

⁴⁹ *Ibid.*

⁵⁰ Cf. A. CLARK, *Intrinsic content, active memory and the extended mind*; A. CLARK, D. CHALMERS, *The extended mind*.

⁵¹ Cf., e.g., R.D. TWENEY, *Faraday's discovery of induction: A cognitive approach*; N.J. NERSESSIAN, *Faraday's field concept*; N.J. NERSESSIAN, *Creating scientific concepts*.

⁵² Cf., e.g., J. LAVE, *Cognition in practice: Mind, mathematics, and culture in everyday life*; E. HUTCHINS, *Cognition in the wild*; N.J. NERSESSIAN, *Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions*; N.J. NERSESSIAN, *Interdisciplinary in the making: Models and methods in frontier science*.

⁵³ J.G. GREENO, *On claims that answer the wrong questions*, p. 6.

⁵⁴ Note that a similar conclusion is reached by Allen, though his focus is the problem of narrow conceptions of cognition, and whether a focus on defining cognition is even a worthy project (cf. C. ALLEN, *On (not) defining cognition*).

⁵⁵ A more systematic account of pluralism is to be found in S.F. ARAUJO, L.M. OSBECK, *Ever not quite. Pluralism in William James and contemporary psychology*.

⁵⁶ Cf., e.g., G.S. KIRK, J.E. RAVEN, M. SCHOFIELD (eds.), *The presocratic philosophers: A critical history with a selection of texts*.

⁵⁷ Cf. W. KERBER, *Pluralismus*.

⁵⁸ C. WOLFF, *Vernünfftige Gedanken von Gott, der Welt, und der Seele des Menschen*, pp. 19-20 – translation from German by S.A.

⁵⁹ I. KANT, *Anthropology from a pragmatic point of view*, p. 241-242.

⁶⁰ T.W. KRUG, *Allgemeines Handwörterbuch der philosophischen Wissenschaften*, p. 278.

⁶¹ J. WAHL, *The pluralist philosophies from England and America*, p. 1. It should be noted that Hegel and Hegelianism played a central role in this context, as many of the pluralists' opposers were Hegelians (cf. also W.J. MANDER, *British Idealism: A history*).

⁶² Cf., e.g., M. BEVIR (ed.), *Modern pluralism*; C. LABORDE, *Pluralist thought and the State in Britain and France, 1900-1925*; J. WAHL, *The pluralist philosophies from England and America*.

⁶³ Cf. J.-H. BOEX-BOREL, *Le pluralisme. Essai sur la discontinuité et l'hétérogénéité phénomènes*; W. JAMES, *A pluralistic universe*; P. LANER, *Pluralismus oder Monismus*; J. WARD, *The realm of ends or pluralism and theism*.

⁶⁴ For more focused discussion on pluralism's contem-

porary contexts of use and associated meanings, cf. S.F. ARAUJO, L.M. OSBECK, *Ever not quite. Pluralism in William James and contemporary psychology* (forthcoming).

⁶⁵ W. JAMES, *A pluralistic universe*, p. 20. James is perhaps the best-known pluralist in the American tradition. Besides offering a series of lectures on pluralism, he discussed its meanings and implications throughout his work. For a detailed account on James's pluralism, cf. S.F. ARAUJO, L.M. OSBECK, *Ever not quite. Pluralism in William James and contemporary psychology*.

⁶⁶ *Ibid.*, p. 145.

⁶⁷ It is also possible to develop pluralism as an ontological alternative to dualism and physicalism, as Ludwig claims (cf. D. LUDWIG, *New wave pluralism*).

⁶⁸ Cf., e.g., A. JANSSEN, C. KLEIN, M. SLORS, *What is a cognitive ontology, anyway?*; R. PAIN, *Phenomenology and cognitive neuroscience*; C.J. PRICE, K.J. FRISTON, *Functional ontologies for cognition*.

⁶⁹ Cf. S.D. MITCHELL, *Unsimple truths*.

⁷⁰ *Ibid.*, p. 20.

⁷¹ *Ibid.*, p. 71.

⁷² *Ibid.*, p. 122.

⁷³ *Ibid.*, p. 21.

⁷⁴ Cf. F. ADAMS, K. AIZAWA, *The bounds of cognition*.

⁷⁵ T. VAN GELDER, R.F. PORT, *Mind as motion*, p. 2.

⁷⁶ W. JAMES, *Brute and human intellect*, p. 12.

⁷⁷ For different developments of epistemic pluralism and its particularities, cf. A. COLIVA, N.J. LEE, N. PEDERSEN (eds.), *Epistemic pluralism*; S. RUPHY, *Scientific pluralism reconsidered*; N. ZANGWILL, *The missing link and the ambitions of epistemology*.

⁷⁸ S.D. MITCHELL, *Unsimple truths*, p. 14.

⁷⁹ The issue of the potential incommensurability between epistemic models, as raised by Kuhn and others, is of utmost importance for the establishment of epistemological pluralism in cognitive science (cf., e.g., T. KUHN, *The structure of scientific revolutions*). However, it is beyond our scope in the present paper.

⁸⁰ R. DALE, E. DIETRICH, A. CHEMERO, *Explanatory pluralism in cognitive science*, p. 739.

⁸¹ H.L. DE JONG, *Introduction: A symposium on explanatory pluralism*, p. 731.

⁸² *Ibidem*.

⁸³ R.N. MCCAULEY, W. BECHTEL, *Explanatory pluralism and heuristic identity theory*, p. 736.

⁸⁴ Cf. F. ADAMS, K. AIZAWA, *The bounds of cognition*.

⁸⁵ Cf., e.g., T. VAN GELDER, R.F. PORT, *Mind as motion*.

⁸⁶ Cf. E. HUTCHINS, *Cognition in the wild*.

⁸⁷ R.E. FABRY, *Transcending the evidentiary boundary*, p. 392.

⁸⁸ *Ibid.*, p. 404.

⁸⁹ R.N. MCCAULEY, *Time is of the essence*, pp. 613-614.

⁹⁰ Cf., e.g., W. JAMES, *A pluralistic universe*.

⁹¹ For an excellent recent discussion of the concept of progress in psychological science, cf. U. FEEST, *Progress in psychology*.

⁹² I. BERLIN, *The crooked timber of humanity*, pp. 79-80. Ethical pluralism, however, can develop in different directions with several particularities within each approach (cf., e.g., I. LIU, *Ethical pluralism and the appeal to human nature*; R. MADSEN, T. STRONG (eds.), *The many and the one*; B.C. POSTOW, *Toward honest ethical pluralism*).

⁹³ S.D. MITCHELL, *Unsimple truths*, p. 85.

⁹⁴ *Ibidem*.

⁹⁵ *Ibid.*, p. 87.

⁹⁶ Cf. H.E. DOUGLAS, *Science, policy, and the value-free*

ideal; H.E. DOUGLAS, *Values in science*.

⁹⁷ Cf., e.g., A. ABRAHAMSEN, W. BECHTEL, *History and core themes*; H. GARDNER, *The mind's new science*.

⁹⁸ Cf. H. GARDNER, *The mind's new science*, p. 37.

⁹⁹ Cf. S.F. ARAUJO, L.M. OSBECK, *Ever not quite. Pluralism in William James and contemporary psychology*; J.R. GOERTZEN, *Dialectical pluralism*.

¹⁰⁰ Cf., e.g., R. FRODEMAN, *Sustainable knowledge*; R. FRODEMAN, *The future of interdisciplinarity*; J.T. KLEIN, *Interdisciplinarity: History, theory, and practice*; J.T. KLEIN, *Typologies of interdisciplinarity*; N.J. NERSESSIAN, *Interdisciplinarity in the making: Models and methods in frontier science*.

¹⁰¹ Cf. S.D. MITCHELL, *Biological complexity and integrative pluralism*.

¹⁰² Cf. S. HALES (ed.), *A companion to relativism*; R. HARRÉ, M. KRAUSZ, *Varieties of relativism*; M. KUSCH, *Relativism in the philosophy of science*.

¹⁰³ For a recent attempt, cf. S.J. VEIGL, *Notes on a complicated relationship*.

¹⁰⁴ Cf. H. CHANG, *Is water H₂O? Evidence, realism and pluralism*; S. KELLERT, H. LONGINO, C.K. WATERS, *The pluralist stance*; S.D. MITCHELL, *Biological complexity and integrative pluralism*.

¹⁰⁵ Cf. S.D. MITCHELL, *Biological complexity and integrative pluralism*.

¹⁰⁶ C. ALLEN, *On (not) defining cognition*, p. 4241.

¹⁰⁷ Cf. K. MUENZINGER, *Introduction*.

¹⁰⁸ F. HEIDER, *Trends in cognitive theory*, p. 201.

¹⁰⁹ *Ibid.*, p. 209.

Literature

- ABRAHAMSEN, A., BECHTEL, W. (2012). *History and core themes*. In: K. FRANKISH, W. RAMSEY (eds), *The Cambridge handbook of cognitive science*, Cambridge University Press, Cambridge, pp. 9-28.
- ADAMS, F., AIZAWA, K. (2008). *The bounds of cognition*, Blackwell, London/New York.
- ADAMS, F., AIZAWA, K. (2009). *Why the mind is still in the head*. In: M. AYDEDE, P. ROBBINS (eds.), *The Cambridge handbook of situated cognition*, Cambridge University Press, Cambridge, pp. 78-95.
- ADAMS, F., AIZAWA, K. (2010). *Defending the bounds of cognition*. In: R. MENARY (ed.), *The extended mind*, MIT Press, Cambridge (MA), pp. 67-80.
- ALLEN, C. (2017). *On (not) defining cognition*. In: «Synthese», vol. CXCIV, n. 11, pp. 4233-4249.
- ARAUJO, S.F., OSBECK, L.M. (forthcoming). *Ever not quite. Pluralism in William James and contemporary psychology*, Cambridge University Press, Cambridge.
- ARFINI, S., MAGNANI, L. (forthcoming). *Embodied, extended, ignorant minds. New studies on the nature of not-knowing*, Springer International Publishing, Cham.
- BAERVELDT, C., VERHEGGEN, T. (2012). *Enactivism*. In: J. VALSINER (ed.), *Oxford handbook of culture and psychology*, Oxford University Press, Oxford, pp. 165-190.
- BERLIN, I. (1990). *The crooked timber of humanity*, John Murray, London.
- BEVIR, M. (ed.) (2012). *Modern pluralism*, Cambridge University Press, Cambridge.
- BOEX-BOREL, J.-H. (1909). *Le pluralisme. Essai sur la discontinuité et l'hétérogénéité phénomènes*, F. Alcan, Paris.
- BROWN, J.S., COLLINS, A., DUGUID, P. (1989). *Situated cognition and the culture of learning*. In: «Educational Researcher», vol. XVIII, n. 1, pp. 32-42.
- BRUNER, J. (1990). *Acts of meaning*, Harvard University Press, Cambridge (MA).
- BRUNSWIK, E. (1957). *Scope and aspects of the cognitive problem*. In: H. GRUBER, K. HAMMOND, R. JESSOR (eds.), *Contemporary approaches to cognition*, Harvard University Press, Cambridge (MA), pp. 5-31.
- BUCKNER, C., FRIDLAND, E. (2017). *What is cognition? Angry monism, permissive pluralism(s), and the future of cognitive science*. In: «Synthese», vol. CXCIV, n. 11, pp. 4191-4195.
- CHANG, H. (2012). *Is water H₂O? Evidence, realism and pluralism*, Springer, Berlin/New York/Boston.
- CHEMERO, A. (2003). *An outline of a theory of affordances*. In: «Ecological Psychology», vol. XV, n. 2, pp. 181-195.
- CLARK, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*, Oxford University Press, Oxford.
- CLARK, A. (2005). *Intrinsic content, active memory and the extended mind*. In: «Analysis», vol. LXV, n. 1, pp. 1-11.
- CLARK, A., CHALMERS, D. (1998). *The extended mind*. In: «Analysis», vol. LVIII, n. 7, pp. 7-19.
- COLIVA, A., LEE, N.J., PEDERSEN, N. (eds.) (2017), *Epistemic pluralism*, Palgrave-Macmillan, London/New York.
- COOPER, R.P., SHALLICE, T. (2010). *Cognitive neuroscience: The troubled marriage of cognitive science and neuroscience*. In: «Topics in Cognitive Science», vol. II, n. 3, pp. 398-406.
- COSTA, E. (2018). *Affordances-in-practice: An ethnographic critique of social media logic and context collapse*. In: «New Media and Society» vol. XX, n. 10, pp. 3641-3656.
- DALE, R., DIETRICH, E., CHEMERO, A. (2009). *Explanatory pluralism in cognitive science*. In: «Cognitive Science», XXXIII, n. 5, pp. 739-742.
- DE GELDER, B. (1985). *The cognitivist conjuring trick or how development vanished*. In: C. BAILEY, R. HARRIS (eds.), *Developmental mechanisms of language*, Pergamon, New York, pp. 149-166.
- DE JONG, H.L. (2001). *Introduction: A symposium on explanatory pluralism*. In: «Theory and Psychology», vol. XI, n. 6, pp. 731-735.
- DEWEY, J. (1910). *How we think*, D.C. Heath, Boston/New York/Chicago.
- DOUGLAS, H.E. (2009). *Science, policy, and the value-free ideal*, University of Pittsburgh Press, Pittsburgh.
- DOUGLAS, H.E. (2016). *Values in science*. In: P. HUMPHRIES (ed.), *The Oxford handbook of philosophy of science*, Oxford University Press, Oxford, pp. 609-630.
- FABRY, R.E. (2017). *Transcending the evidentiary boundary: Prediction error minimization, embodied interaction, and explanatory pluralism*. In: «Philosophical Psychology», vol. XXX, n. 4, pp. 391-414.
- FEEST, U. (forthcoming). *Progress in psychology*. In: Y. SHAN (ed.), *New philosophical perspectives on scientific progress*, Routledge, London/New York.
- FODOR, J.A. (1975). *The language of thought*, Crowell Press, New York.

- FODOR, J.A. (1980). *Methodological solipsism considered as a research strategy in cognitive science*. In: «Behavioral and Brain Sciences», vol. III, n. 1, pp. 63-109.
- FRODEMAN, R. (2013). *Sustainable knowledge: A theory of interdisciplinarity*, Springer, Berlin/New York/Boston.
- FRODEMAN, R. (2017). *The future of interdisciplinarity: An introduction to the 2nd edition*. In: R. FRODEMAN (ed.), *The Oxford handbook of interdisciplinarity*, Oxford University Press, Oxford/New York, pp. 3-8, 2nd edition.
- GALLAGHER, S. (2018). *The extended mind: State of the question*. In: «The Southern Journal of Philosophy», vol. LVI, n. 4, pp. pp. 421-447.
- GARDNER, H. (1987). *The mind's new science*, Basic Books, New York.
- GIBSON, J.J. (1977). *The theory of affordances*. In: R. SHAW, J. BRANSFORD (Ed.), *Perceiving, acting, and knowing*, Lawrence Erlbaum, Hillsdale (NJ), pp. 67-82.
- GIBSON, J.J. (1979). *The ecological approach to visual perception*, Houghton Mifflin, Boston.
- GOERTZEN, J.R. (2010). *Dialectical pluralism: A theoretical conceptualization of pluralism in psychology*. In: «New Ideas in Psychology», vol. XXVIII, n. 2, pp. 201-209.
- GREENO, J.G. (1997). *On claims that answer the wrong questions*. In: «Educational Researcher», vol. XXVI, n. 1, pp. 5-17.
- HALES, S. (ed.) (2011). *A companion to relativism*, Wiley-Blackwell, London/New York.
- HARRÉ, R. KRAUSZ, M. (1996). *Varieties of relativism*, Wiley-Blackwell, London/New York.
- HEIDER, F. (1957). *Trends in cognitive theory*. In: H. GRUBER, K. HAMMOND, R. JESSOR (eds.), *Contemporary approaches to cognition*, Harvard University Press, Cambridge (MA), pp. 201-210.
- HUTCHINS, E. (1995). *Cognition in the wild*, MIT Press, Cambridge (MA).
- JAMES, W. (1890). *The principles of psychology*, vol. I, Henry Holt and Co., New York.
- JAMES, W. (1977). *A pluralistic universe* (1909), Harvard University Press, Cambridge (MA).
- JAMES, W. (1983). *Brute and human intellect* (1878). In: W. JAMES (ed.), *Essays in psychology*, Harvard University Press, Cambridge (MA), pp. 1-37.
- JANSSEN, A., KLEIN, C., SLORS, M. (2017). *What is a cognitive ontology, anyway?*. In: «Philosophical Explorations», vol. XX, n. 2, pp. 123-128.
- KANT, I. (2009). *Anthropology from a pragmatic point of view* (1798). In: G. ZÖLLER, R. LOUDEN (eds.), *The Cambridge edition of the works of Immanuel Kant* (Anthropology, History, and Education), Cambridge University Press, Cambridge, pp. 227-429.
- KELLERT, S., LONGINO, H., WATERS, C.K. (2006). *The pluralist stance*. In: S. KELLERT, H. LONGINO, C.K. WATERS (eds.), *Scientific pluralism*, University of Minnesota Press, Minneapolis, pp. vii-xxix.
- KERBER, W. (1989). *Pluralismus*. In: J. RITTER, K. GRÜNDER (Hrsg.), *Historisches Wörterbuch der Philosophie*, Bd. VII, Schwabe & Co., Basel, pp. 988-993.
- KIRK, G.S., RAVEN, J.E., SCHOFIELD, M. (eds.) (2013). *The presocratic philosophers: A critical history with a selection of texts*, Cambridge University Press, Cambridge, 2nd edition.
- KLEIN, J.T. (1990). *Interdisciplinarity: History, theory, and practice*, Wayne State University Press.
- KLEIN, J.T. (2017). *Typologies of interdisciplinarity*. In: R. FRODEMAN (ed.), *The Oxford handbook of interdisciplinarity*, Oxford University Press, Oxford/New York, pp. 21-34.
- KRUG, T.W. (1833). *Allgemeines Handwörterbuch der philosophischen Wissenschaften*, Bd. III, Brockhaus, 2nd edition.
- KUHN, T. (1970). *The structure of scientific revolutions*, University of Chicago Press, Chicago, 2nd edition.
- KUSCH, M. (2021). *Relativism in the philosophy of science*, Cambridge University Press, Cambridge.
- LABORDE, C. (2000). *Pluralist thought and the State in Britain and France, 1900-1925*, Palgrave Macmillan, London/New York.
- LAKOFF, G., JOHNSON, M. (1999). *Philosophy in the flesh. The embodied mind and its challenge to western thought*, Basic Books, New York.
- LANER, P. (1905). *Pluralismus oder Monismus*, Albert Kohler, Berlin.
- LAVE, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*, Cambridge University Press, Cambridge.
- LAVE, J., WENGER, E. (1991). *Situated learning: Legitimate peripheral participation*, Cambridge University Press, Cambridge.
- LIU, I. (2018). *Ethical pluralism and the appeal to human nature*. In: «European Journal of Philosophy», vol. XXVI, n. 3, pp. 1103-1119.
- LUDWIG, D. (2013). *New wave pluralism*. In: «Dialectica», vol. LXVII, n. 4, pp. 545-560.
- MADSEN, R., STRONG, T. (eds.) (2003), *The many and the one*, Princeton University Press, Princeton.
- MAIESE, M. (2018). *Can the mind be embodied, enactive, affective, and extended?*. In: «Phenomenology and the Cognitive Sciences», vol. XVII, n. 2, pp. 343-361.
- MANDER, W.J. (2011). *British Idealism: A history*, Oxford University Press, Oxford.
- MCCAULEY, R.N. (2009). *Time is of the essence: Explanatory pluralism and accommodating theories about long-term processes*. In: «Philosophical Psychology», vol. XXII, n. 5, pp. 611-635.
- MCCAULEY, R.N., BECHTEL, W. (2001). *Explanatory pluralism and heuristic identity theory*. In: «Theory and Psychology», vol. XI, n. 6, pp. 736-760.
- MENARY, R. (2010). *Introduction: The extended mind in focus*. In: R. MENARY (ed.), *The extended mind*, MIT Press, Cambridge (MA), pp. 1-25.
- MITCHELL, S.D. (2003). *Biological complexity and integrative pluralism*, Oxford University Press, Oxford.
- MITCHELL, S.D. (2009). *Unsimple truths*, University of Chicago Press, Chicago.
- MONK, R. (1978). *Cognitivism and cognitive psychology*. In: «Behavioral and Brain Sciences», vol. I, n. 2, pp. 242-243.
- MUENZINGER, K. (1957). *Introduction*. In: H. GRUBER, K. HAMMOND, R. JESSOR (eds.), *Contemporary approaches to cognition*, Harvard University Press, Cambridge (MA), pp. 1-4.
- NERSESSIAN, N.J. (1985). *Faraday's field concept*. In: D.C. GOODING, F.A.J.L. JAMES (eds.), *Faraday rediscovered: Essays on the life & work of Michael Faraday, 1791-1867*, Stockton/Macmillan, London/New York, pp. 377-406.
- NERSESSIAN, N.J. (2005). *Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions*. In: M. GORMAN, R. TWEN-

- EY, D. GOODING, A. KINCANNON (eds.), *Scientific and technological thinking*, Erlbaum, pp. 17-56.
- NERSESSIAN, N.J. (2008). *Creating scientific concepts*, MIT Press, Cambridge (MA).
- NERSESSIAN, N.J. (2022). *Interdisciplinarity in the making: Models and methods in frontier science*, MIT Press, Cambridge (MA).
- NEWELL, A. (1980). *Physical symbol systems*. In: «Cognitive Science», vol. IV, n. 2, pp. 135-183.
- OSBECK, L.M. (2009). *Transformations in cognitive science: Implications and issues posed*. In: «Journal of Theoretical and Philosophical Psychology», vol. XXIX, n. 1, pp. 16-33.
- OSGOOD, C.E. (1957). *Discussion of paper by Egon Brunswick*. In: H. GRUBER, K. HAMMOND, R. JESSOR (eds.), *Contemporary approaches to cognition*, Harvard University Press, Cambridge (MA), pp. 33-40.
- PAIN, R. (2018). *Phenomenology and cognitive neuroscience: Can a process ontology help resolve the impasse?*. In: «Australasian Philosophical Review», vol. II, n. 2, pp. 204-208.
- PITT, D. (2020). *Mental representation*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Spring Edition, URL: <https://plato.stanford.edu/archives/spr2020/entries/mental-representation/>.
- POSNER, M.I., DI GIROLAMO, G.J. (2000). *Cognitive neuroscience: Origins and promise*. In: «Psychological Bulletin», vol. CXXVI, n. 6, pp. 873-889.
- POSTOW, B.C. (2007). *Toward honest ethical pluralism*. In: «Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition», vol. CXXXII, n. 2, pp. 191-210.
- PRICE, C.J., FRISTON, K.J. (2005). *Functional ontologies for cognition: The systematic definition of structure and function*. In: «Cognitive Neuropsychology», vol. XXII, n. 3, pp. 262-275.
- RUPHY, S. (2017). *Scientific pluralism reconsidered*, University of Pittsburgh Press, Pittsburgh.
- SCARINZI, A. (2020). *4Es are too many: Why enactive world making does not need the extended mind thesis*. In: «JoLMA», vol. I, n. 2, pp. 237-254.
- SHEA, N. (2018). *Representation in cognitive science*, Oxford University Press, Oxford.
- SIMON, H.A. (1981). *Studying human intelligence by creating artificial intelligence: When considered as a physical symbol system, the human brain can be fruitfully studied by computer simulation of its processes*. In: «American Scientist», vol. LXIX, n. 3, pp. 300-309.
- STANDAGE, D., TRAPPENBERG, T. (2012). *Cognitive neuroscience*. In: K. FRANKISH, W. RAMSEY (eds), *The Cambridge handbook of cognitive science*, Cambridge University Press, Cambridge, pp. 235-256.
- STILL, A., COSTALL, A. (1987). *Introduction: In place of cognitivism*. In: A. COSTALL, A. STILL (eds.), *Cognitive psychology in question*, St. Martin's Press, New York, pp. 1-11.
- STILL, A., COSTALL, A. (eds.) (1991). *Against cognitivism: Alternative foundations for cognitive psychology*, Harvester Wheatsheaf, Hemel Hempstead/London.
- TWENEY, R.D. (1985). *Faraday's discovery of induction: A cognitive approach*. In: D.C. GOODING, F.A.J.L. JAMES (eds.), *Faraday rediscovered: Essays on the life & work of Michael Faraday, 1791-1867*, Stockton/Macmillan, London/New York, pp. 189-209.
- VAN GELDER, T., PORT, R.F. (1995). *Mind as motion: Explorations in the dynamics of cognition*, MIT Press, Cambridge (MA).
- VARELA, F.J., THOMPSON, E., ROSCH, E. (1991). *The embodied mind: Cognitive science and human experience*, MIT Press, Cambridge (MA).
- VARELA, F.J. (1996). *Neurophenomenology: A methodological remedy for the hard problem*. In: «Journal of Consciousness Studies», vol. III, n. 4, pp. 330-349.
- VEIGL, S.J. (2021). *Notes on a complicated relationship: Scientific pluralism, epistemic relativism, and stances*. In: «Synthese», vol. CXCIX, n. 1-2, pp. 3485-3503.
- VILLALOBOS, M., DEWHURST, J. (2017). *Why post-cognitivism does not (necessarily) entail anti-computationalism*. In: «Adaptive Behavior», vol. XXV, n. 3, pp. 117-128.
- VON ECKARDT, B. (2012). *The representational theory of mind*. In: K. FRANKISH, W. RAMSEY (eds), *The Cambridge handbook of cognitive science*, Cambridge University Press, Cambridge, pp. 29-49.
- WAHL, J. (1925). *The pluralist philosophies from England and America*, translated by F. ROTHWELL, The Open Court Company, London.
- WARD, J. (1911). *The realm of ends or pluralism and theism*, Cambridge University Press, Cambridge.
- WERTHEIMER, M. (1945). *Produktives Denken*, Walde-mar Kramer, Frankfurt a.M.
- WERTZ, F. (1987). *Cognitive psychology and the understanding of perception*. In: «Phenomenological Psychology», vol. XVIII, n. 1, pp. 103-142.
- WITHAGEN, R., DE POEL, H.J., ARAÚJO, D., PEPPING, G.J. (2012). *Affordances can invite behavior: Reconsidering the relationship between affordances and agency*. In: «New Ideas in Psychology», vol. XXX, n. 2, pp. 250-258.
- WOLFF, C. (1722). *Vernünfftige Gedanken von Gott, der Welt, und der Seele des Menschen*, Renger, 2nd edition, pp. 19-20.
- WUNDT, W. (1920). *Erlebtes und Erkanntes*, Alfred Kröner, Stuttgart.
- ZANGWILL, N. (2020). *The missing link and the ambitions of epistemology*. In: «Metaphilosophy», vol. LI, n. 4, pp. 485-498.

RICERCHE

What philosophy, if any, is needed for cognitive science?

Philip V. Kargopoulos^(a)

Ricevuto: 31 marzo 2022 ; accettato: 20 marzo 2023

Abstract The question posed in the title serves as a springboard to examine the interdisciplinary nature of cognitive science and the role philosophy should play. I will argue that philosophy has a clearly defined role to play over and above the contributions made by philosophies specific to the various disciplines engaged in cognitive science. I also point out that by engaging with cognitive science – an endeavor defined by a clear scientific goal – philosophy itself will have to change in ways that are bound to affect the future of philosophizing. In the first part of this paper, I play devil’s advocate, arguing that philosophy should not be amongst the disciplines engaged in cognitive science. In the second part, I discuss how the relationship between philosophy and psychology has been changed by their common participation in cognitive science: psychology is now admittedly philosophical and philosophy has become naturalized. In the third part, I suggest that philosophy is better equipped than psychology to handle the thorny question of representation in cognitive science. In the fourth part and conclusion, I argue that cognitive science will influence both the theory and practice of philosophy: the futures of cognitive science and philosophy are tied together.

KEYWORDS: Philosophy; Cognitive Science; Cognitive Psychology; Naturalism; Epistemology

Riassunto *Quale filosofia – se una ce ne vuole – per la scienza cognitiva?* – L’interrogativo posto nel titolo funge da punto di partenza per analizzare la natura interdisciplinare della scienza cognitiva e il ruolo che la filosofia dovrebbe avere al suo interno. Sosterrò che alla filosofia spetti un ruolo ben definito, al di là dei contributi forniti dai particolari indirizzi filosofici coinvolti nella scienza cognitiva. Sottolineerò anche che, confrontandosi con la scienza cognitiva – un ambito caratterizzato da un chiaro obiettivo scientifico – la filosofia stessa dovrà cambiare, assumendo modalità che influenzeranno il futuro del filosofare. Nella prima parte di questo articolo metterò le vesti dell’avvocato del diavolo, sostenendo come la filosofia non dovrebbe essere tra le discipline coinvolte nella scienza cognitiva. Nella seconda parte discuterò come il rapporto tra filosofia e psicologia sia mutato per effetto della comune partecipazione al progetto della scienza cognitiva: ora la psicologia è apertamente filosofica e la filosofia è filosofia naturalizzata. Nella terza parte mostrerò come la filosofia sia meglio equipaggiata della psicologia per fronteggiare l’ostico problema della rappresentazione nella scienza cognitiva. Nella quarta parte e nelle conclusioni sosterrò che la scienza cognitiva potrà influenzare la teoria e la pratica del filosofare: il futuro della scienza cognitiva e della filosofia sono reciprocamente legati.

PAROLE CHIAVE: Filosofia; Scienza cognitiva; Psicologia cognitiva; Naturalismo; Epistemologia

^(a)School of Psychology, Faculty of Philosophy, Aristotle University of Thessaloniki, AUTH Campus - 54124 Thessaloniki (GR)

E-mail: kargop@psy.auth.gr (✉)



PHILOSOPHERS ARE ALWAYS BOUND TO metaphilosophize because philosophy is both a “reflective” and a “reflexive” enterprise: “reflective” like any theory-seeking discipline; “reflexive” in that it applies to itself whatever it detects in the reality which it professes to study. Moreover, in spite of academically imposed specializations, the various branches of philosophy, with the exception of most subfields of logic (which have escaped into the mature formal sciences), cannot afford to be overspecialized into watertight compartments, either at its ontological center or along its value-infused periphery. How can we do ethics, aesthetics, and political theory without taking into account human nature and the problems involved in studying it? By the same token, how can we do ontology (a theory of what it is to be) without taking into account the powers, conditions, and limitations of those who attempt it?

The problem has become sharper since philosophy (traditionally a discipline within the Humanities) joined forces with four sciences in an interdisciplinary attempt to solve one of the last enigmas in our worldview, that of the human mind. Cognitive science makes a bold promise: to provide a final theory that explains mentality, with extensions into applications to be tested on a constant and, hopefully, progressive basis. As a result, philosophy, when involved in cognitive science, cannot be limited to some high-handed “criticism-from-above” (i.e., some kind of “armchair cognitive science”), but has to demonstrate serious contributions to this endeavor. What then can we expect of philosophy as a partner in this interdisciplinary campaign?

1 Is philosophy an unsuitable or superfluous partner in cognitive science?

Let us begin with the provocative claim that philosophy is not needed in cognitive science. We can classify the arguments for leaving philosophy out of cognitive science into two camps: In the first are those who might claim that philosophy as a discipline is not a science (with a core of commonly accepted doctrine) but rather an intellectual activity (i.e. philosophizing) that does not fit into a purely scientific joint venture. The second contains those who might claim that whatever philosophy could offer in this endeavor is already provided by the other four sciences involved (psychology, computer science, neuroscience, linguistics) or by their respective philosophies making philosophy superfluous in cognitive science. Let us begin with the first case and drum up the strongest defense for its radical claim.

In the first place, there is a serious dispute as to what kind of knowledge philosophy is supposed to offer. Some think it should be the Socratic tradition of critical enquiry aimed at conceptual analy-

sis and clarification. This would even include philosophical attempts to employ ordinary language analysis to dispel the problems (in fact, puzzles) generated by language in the first place. Others look to the Aristotelian tradition for a final and complete theory of reality with all its neatly fitted components, more recently represented by various types of endorsement of reductionism and scientific realism. Between these two approaches, a series of other combinations may be advanced (in accordance with what is considered fundamental and what is considered derivative), but the widely accepted common ground is that philosophy traditionally deals with problems and puzzles that cannot just be solved by looking more carefully into nature and that philosophers should be ready to tolerate pluralism in their proposed solutions to such problems. At this point, the aims of cognitive science seem to diverge from philosophy. Cognitive science is an inter-scientific endeavor promising a testable and applicable theory of mentality. It cannot tolerate endless discussion about what “the mind” is, or what methods are appropriate for investigating it.

In the second place, it is hard to imagine what methodological contribution philosophy might contribute to cognitive science. The main benefit of an interdisciplinary approach to the mind is that the different sciences provide different methods for studying the mind. Psychology is an empirical experimental science that studies cognitive and behavioral phenomena. Linguistics is a formal science of symbolic systems that investigates the rules that govern the use of these symbols to convey information. It tests its theories on the basis of empirical intuitions as to what native speakers of the various languages would accept as grammatically correct and cognitively significant. Neuroscience is a biological science that deals with the physiology and anatomy of central nervous systems and their relations to animal and human behavior. Computer Science is a technological science (i.e., a kind of engineering), which leads to the construction and successful testing of artifacts (such as software programs). Given the above, what role could philosophy take on from a methodological viewpoint? What part of the mental remains to be systematically examined and how could philosophy provide a method for conducting this analysis?

But even assuming that philosophy does become involved in a dispute that has arisen within cognitive science and is prepared to advance a general thesis, like Fodor’s *modularity architecture*¹ or a specific hypothesis, like Fodor’s *language of thought*,² what kind of experiments or research projects could philosophy propose and successfully complete, in the way that the other cognitive sciences can? In general, what does philosophy have to propose that would advance the work of its partners?

It has to be admitted that although some of philosophy's celebrated negative philosophical arguments, such as Searle's *Chinese room thought experiment*,³ inspired considerable discussion, they did not significantly affect cognitive science research nor contribute much in terms of significant shifts in cognitive scientific theorizing or research programs.

A final point to consider is that to understand a philosopher's proposal, one has to come to terms with the whole or a large part of his/her theory and often engage in complex tasks of interpretation. Let us keep in mind that philosophy is traditionally classified as a field within the humanities, where the individual perspectives which give rise to a position matters for a proper understanding of that position. Such activities are not in the interests of working scientists and rightfully so. Science divides and conquers, allowing work to be carried out by different teams and individuals who recognize common problems and accept background knowledge that unifies their efforts. Philosophy traditionally does not operate in this manner, except within some narrowly defined schools of thought. Should we only allow that such schools of philosophy work on cognitive science, and if so, what would be their basic tenets?

In response to the above claims, let us begin by abandoning two philosophical illusions. The fact that the term "mind" was introduced by philosophers does not endorse philosophy's candidacy for inclusion in cognitive science. All sciences started as parts of philosophy, but then matured away from philosophy. Even terms like "energy" and "matter" originally appeared in Aristotle's *Physics*. Yet these terms are now of little philosophical interest, while any philosophers who would like to speak about energy and matter are well-advised to know their physics well or at least better than their philosophy. Other terms, like "memory" or "learning", instead began their career in ordinary parlance, initially had a small place in philosophy, fared well in psychology, and now form part of cognitive science. Still, other terms, like "consciousness" or "intentionality" began their career in philosophy and (finally) reached a respectable position in cognitive science. It is clear that philosophy's claim to be included in cognitive science as an equal partner cannot be based solely on its history of involvement with the mind. This means that by including philosophy in cognitive science, we move away from a traditional conception of philosophy as a discipline of the humanities, which has to focus on texts in a historical and hermeneutic way that involves understanding individual perspectives. It is of little value to cognitive science to investigate in what sense Hobbes was an early computationalist. Worrying about courting Cartesian solipsism does not call for the exegesis of "cogito", but rather involves recognizing a grounding problem.

The second philosophical illusion is the old Aristotelian idea that philosophy, in general, and metaphysics (i.e., "first philosophy" or ontology), in particular, is the "most architectonic" of the theoretical sciences, as other sciences deal with more specific kinds of being, while philosophy deals with being-in-itself and in its totality, or what is called "the deep picture" (underlying picture) and "the big picture" (complete picture) respectively. In cognitive science, this illusion could lead to the idea that philosophy is needed to provide the ontology of the endeavor and to ensure some ontological orthodoxy or some methodological or other kind of correctness and oversight.

In response to this idea, one can see from its history so far that cognitive science needs no commissars for ideology or clergy for orthodoxy. As far as the ontology of cognitive science is concerned, it is formed by the leading working hypotheses that originated in the sciences involved. In the first stage of cognitive science, it was AI that provided the ontology of the mental as computations on formal symbolic representations, while the appropriate methodology⁴ rested on reverse engineering, which led to some simulation tests, initially proposed by the first computational logician, Alan Turing.⁵ At the same time, the arbitrary *multiple realization principle*, which made good technological sense, placed AI at the center of cognitive science and neuroscience at the periphery. Paradigm shifts in cognitive science are not normally caused by philosophical arguments critical of the existing paradigm, but by anomalies accumulated from normal research, which lead to the adoption of a new ontology and a new accompanying methodology. The shift from the computational to the connectionist paradigm was ushered in by problems encountered in real-time calculations for simple perceptual judgments, in degraded input problems, in the non-plasticity of hardware and software architecture as compared with the plasticity of the brain, and a nagging feeling that human and animal rationality is more connected to inductive rather than deductive practices. The next step, that of neural nets and PDP, came as promising solutions to the aforementioned anomalies, not as a response to the philosophical critiques of Dreyfus⁶ and Searle,⁷ who hastily adapted the Chinese Room into the Chinese Gym Argument in order to cover connectionist alternatives endorsed by Patricia and Paul Churchland.⁸ As we are in the middle of a new paradigm in cognitive science that insists on "embodied" and "situated" mind, we witness a complete reversal, with neuroscience at the center of cognitive science proposing ontologies and methodologies and traditional AI at the periphery. Philosophical ideas, like Fodor's *Modularity* and *LOT*, and disputes, such as the dispute between Kosslyn⁹ and Pylyshyn¹⁰ concerning mental imag-

es, have contributed to the design of research in cognitive science. But shifts in the paradigm of cognitive science involve more than shifts in philosophical credo. Philosophy comes to the support of a cognitive science shift, normally after that shift happens, while at the same time, philosophers already involved in cognitive science often hold on to their old philosophical positions, even when cognitive science has moved away from them. On the one hand, one wonders if a philosopher working in cognitive science has to declare allegiance to eliminative materialism, identity theory, or epiphenomenalism. On the other hand, holding on to previously proposed approaches is understandable as older paradigms in cognitive science are not fully abandoned but continue to be employed and have special roles to fulfill. We may be far from the initial computationalism of the 60s and 70s, but we still hold on to principles advanced by computationalism, not just out of gratitude or nostalgia. We meaningfully remind ourselves, philosophers included, that “AI keeps us honest” and that unless we show the mental to be physicalistically constructible in principle, our work is not yet finished.

Another set of arguments against including philosophy in cognitive science is the idea that each of the other involved sciences already has its own philosophy (of AI, of linguistics, of psychology, even of neuroscience), so that a separate and equal philosophical partner would be superfluous. On the other hand, there are two ways in which this argument could be used to support the opposite conclusion. In the first place, if all the sciences of cognitive science come with their own philosophies in tow, one might argue that philosophy could serve as common ground where the four sciences meet. In this case, however, the metaphor “common ground” must be spelled out. In the second place, it is admitted that the philosophies of specific sciences normally deal with theoretical difficulties concerning principles upon which these sciences rest and philosophy is the rightful place where such principles should be discussed. In this case, the quasi-metaphorical term “principles” must be spelled out. We begin with the second task (principles) and hope to reach a determination of the first task (finding common ground) by the end of the paper.¹¹

A science normally turns to its own philosophy on the level of principles which may lead to significant differences in methods or approaches or generate more than one school of accepted scientific doctrine. We can divide these philosophical problems concerning the principles of sciences into two general kinds. Those that pertain to the truth of the theories (and thus to the reference of theoretical and “bridge” terms, to rules of inference, and in general to whatever affects the objectivity of the science involved) and those that pertain to the ex-

planation and adequacy of the proposed theories. In psychology, for example, behaviorism challenged not only the objectivity of mentalistic approaches, which have to rely on what cannot be intersubjectively observed and measured, but also their explanatory adequacy, since mentalistic explanations, according to Skinner,¹² simply repeat the description of the behavior of the animal on another mental “homuncular” level, thus leading to circular, rather than real, explanations.

In the most traditional axiomatic setting of theory, principles include definitions or lists of primitive terms, axioms or postulates, and rules of inference. These sciences refer primarily to abstract forms, which are ideal objects and thus carry with them a sense of absolute certainty as to truth which seems coextensive with proof. Yet even here, philosophical questions of principle arise out of difficulties generated by the problematic status of the entities referred to and paradoxical notions like infinity or self-reference. An archetypical example in that score is the most axiomatic of sciences, mathematics, where to doubt fundamental axioms involving infinity, like Euclid’s fifth postulate or Peano’s mathematical induction axiom, or the law of the excluded middle, leads to different theorems and radically different approaches resting on restrictions that affect even rules of inference.

Moving on from the formal sciences of proof to the sciences of confirmation and explanation, which have natural kinds as their objects, their principles include fundamental conceptions of space, time, and causation, which allow for a plurality of philosophical approaches. Next to these fundamental conceptions there are, in each science, explicitly stated or implicitly assumed, some fundamental guiding principles which often deserve philosophical scrutiny. Examples of these are the “No Action at a Distance” principle, the “Conservation of Mass” and the “Conservation of Energy”, “Least Action Principle”, “Entropy”, “Homeostasis” and even in psychology the well-known “Law of Effect”. Moreover, a similar guiding role is fulfilled by certain proposed models and metaphors, like the planetary model for the atom, the “trial and error” model of learning, the tripartite division of our inner world (from Plato’s appetite, sentiment and reason, to Freud’s id, ego, superego, or the subconscious-preconscious-conscious division, to reptilian, mammalian, rational brain). It made good sense to retain a questioning, philosophical attitude in the early stages of modern medicine when Claude Bernard proposed a revolutionary conception of illness as continuous with (and not antithetical to) the normal healthy state. In the present time, however, when most of the science of medicine is secured by way of biochemical and micromolecular explanations, there is no task for philosophy to perform in theoretical matters of mainstream medicine, except in areas of medicine still

trapped in immaturity, like psychiatry.¹³ Finally, and most importantly, in many sciences, principles include fundamental theoretical entities proposed to serve as the ontological basis of explanations. When the existence of such entities is established by experimental means, the need for keeping philosophical guard over that science is gradually eliminated and it proceeds to a mature stage.

Two points must be borne in mind. In the first place, as sciences mature and the reference of theoretical terms is established, there is less need for a philosophy of this science. Psychiatry and some parts of neurology still need philosophy, while the rest of the medical sciences can depend philosophically on the philosophy of biology and the philosophy of neuroscience. In the second place, a science that is reduced to another more basic science needs no philosophy, as the latter now carries the full weight of explanatory principles. There is no need for a philosophy of chemistry; there is a need for the philosophy of physics.

Sciences that are concerned with human beings, their activities, and their historically developing institutions, necessarily have an accompanying philosophy, even when they explicitly deny it. Depending on the definition of human nature they employ, and the place they assign to humans in the order of things, they avail themselves of different methods and principles. Like all sciences, the sciences of humans strive to be rigorous. Faced with the complexity of human phenomena, they often have easy recourse to truths by way of statistics, yet statistics will never provide the sought-after explanation. Even the ones that contain mathematized sets of laws that allow for predictions, like economics, offer explanations that rely on assumptions, such as rationality, whose nature is further investigated by sciences like psychology. Kahneman (who worked with Tversky) was awarded the 2002 Nobel prize in Economics for what was in fact a psychological investigation of reasoning practices, biases and errors.¹⁴ Indeed, most, if not all, of the social sciences (following Thucydides as our first master) end up relying on psychology for the ultimate explanation of their principles. Psychology is in a position to undertake this role as it shares common ground with the natural sciences (specifically the biological ones) and borders on the social sciences (the branch of social psychology) as well as the humanities, in that, unlike most sciences, it focuses on the individual or personal element (rather than the universal) but, like a genuine science, feels obliged to attempt general theories that can handle the individual (the psychometry of individual differences and theories of personality).

From the above, it is clear why cognitive science cannot include any of the social sciences (not even anthropology), even though they all involve human mentality (with its normal rationality and

occasional irrationality – or is it the other way round?), because these sciences depend for their explanations on psychology, which provides them with the only available means for reduction. They too are somehow “about the human mind” but make no attempt to explain it. At best they assume it in the way that psychology describes it in its best, worst, and in-between moments. If the sciences of the mind that are included in cognitive science could be reduced to psychology, the only philosophy needed in cognitive science would be the philosophy of psychology. But it was not destined to be so.

2 Psychology and philosophy (naturalized) cohabiting in cognitive science

In cognitive science there is yet no accepted final order of reduction; if there were, it would not be an interdisciplinary endeavor. Psychology in particular, which serves as the basis of reduction in the human-focused sciences is by no means the basis for reduction in cognitive science. Indeed, historically speaking, it is just the opposite, because cognitive science in general, with AI at its forefront, has provided a way for psychology to bypass its main foundational problem, that of psychophysical causation. Psychophysical interaction is primarily a problem of explanation that has precluded the successful reduction of psychology to the already mature natural sciences.¹⁵ The model of the computer provided the idea (or the hope) that such a solution was possible and ushered in a new approach in psychology, that of cognitivism. Later, another approach, using inductive machine learning by parallel distributed processes, ushered in another approach in psychology and cognitive science, that of connectionism. However, this mentalistic courting of associationism did not eliminate cognitivism, which survived in the basic tenet that the mental is composed of representations that carry syntactic structure as a prerequisite of the compositionality of the mental.

We are now running through a period of embodied and situated intelligence, which at times questions even the very basic concept of representation,¹⁶ and proposes a subsumption architecture of controlling and controlled mechanisms promising a very long road that starts with basic animal intelligence (the one that enables the fulfillment of the 4Fs) and reaches all the way up to rational animals of the homo sapiens species with their symbolic linguistic representations coming at the very end. It is clear that the obvious path forward, reducing psychology, to neuroscience, would not be an easy task.

Given the above limitations on psychological theorizing, what kind of role is psychology entitled to play within cognitive science, especially given the real possibility of its reduction to neuroscience

in some distant future? Psychology is the science that has dealt with the mind in an experimental way for over 140 years, while no other science has dared to investigate it systematically. Not only has it collected myriads of established major and minor truths concerning the mental, but in order to do so, it has divided up and organized the field in detailed ways: it has found, proposed, and diligently investigated many parts and kinds of attention, sensation, perception, memory, learning, thinking, knowing, reasoning, deciding, controlling, acting. Psychology is the meeting point for all the sciences in cognitive science because it provides the *explanandum* for all of them. In addition to psycholinguistics, which is the accepted common ground of psychology and linguistics, two sciences that have occupied guiding roles in cognitive science, AI and neuroscience rely on psychology in order to investigate their subject matter. Artificial Intelligence has to start with what is, or what constitutes intelligence, what are its kinds, and what psychologists have found in their attempts to detect, test, and measure it. Neuroscience (which has become cognitive) investigates the structure and functions of the brain using as a preliminary guide the rough and, very likely, erroneous maps of areas of the mind that psychologists have drawn. It is equally important that what psychology has to offer is not just a list of investigated items, but also some significant preliminary organization of these finds, including by way of explanatory schemes that psychology has put forward in its premature theorizing. In fact, the ground on which the edifice of cognitive science is built was provided by psychology. This real estate comes with a lot of solid ground but has entailed some false and perhaps wrong or questionable starts, which have nevertheless contributed many valuable insights.

It would appear then, that as far as theory and explanation are concerned, one could argue that whatever philosophy is to be practiced in Cognitive Science should be adequately and systematically covered by psychology, and if there is a philosophical residue not covered by psychology, this could be covered by the philosophy of psychology.

Without wishing to enter into a turf dispute between philosophy of psychology, philosophy of mind, philosophy of action, philosophy of language, we can agree on the fact that psychology, admittedly from the start a decided science, still carries an undeniable philosophical component that is broader than that found in other philosophies of specific sciences. If the term did not have an air of paradox in it, one could claim that psychology, in addition to being a biological science,¹⁷ is also a science of the mental,¹⁸ a social science,¹⁹ a humanistic discipline,²⁰ and, yes, a “philosophical science” in that a philosophical perspective is necessary to handle all aspects of psychology, all three

levels of psychological explanation (behavioral, cognitive, neuroscientific), as well as the many recognized approaches to (or schools of) psychology that have produced valuable insights, research, or knowledge.²¹

Even though it has remained in many ways philosophical, psychology has been divorced from philosophy since 1879, and this was a much needed and very successful divorce for psychology, as it allowed it to proceed as a rigorous experimental science which has produced a significant number of truths. Still, the progress of psychology in amassing truths has not led to an undisputed theoretical synthesis and, as a result, philosophical concerns still accompany all facets of the field, including its practical applications. Although one can be a good mathematician, physicist, biologist or medical doctor, even if he or she has no inkling of the philosophy of mathematics, physics, biology or medicine respectively, the same cannot be said of psychology. It is not only the fact that “un-philosophical” psychologists will not have fully understood the articles and the books they have studied if they ignore the kind of philosophy that underlies these different texts that informed their education. It is also a matter of day-to-day use of psychological knowledge: even the practitioners of clinical psychology have to deal with problems of causal determination, control, freedom of choice, and rationality. They have to live and work on a day-to-day basis in a philosophical, all-too-human puzzle.

In the same light though, philosophy itself, even before the official birth of cognitive science, and prior to any commitment to cognitivism, had turned towards psychology, as philosophers realized that the justificatory foundational approach to epistemology (Cartesian doubt initiated modern philosophy and reached all the way into the 1950s) had no rational possibility of grounding knowledge on a secure self-evident foundation. As a first step, philosophy turned to language to investigate the limits of its ability to handle its problems. It then quickly realized that an integral part of its investigation would involve focusing on psychological results. A first naturalistic Darwinian step brought philosophy and psychology closer, by way of early functionalism, in the works of James and Dewey, but the idea of the mind as an organ that supports adaptation by way of learning led to behaviorism (influenced also by positivist restrictions on method). The love affair between analytic philosophy and behaviorism began with Russell’s endorsement of its methodology²² and continued with Wittgenstein’s private language argument²³ and Ryle’s explaining away of the mental as dispositions to behavior,²⁴ a line followed by Logical Behaviorism. The most systematic of the analytic philosophers, W. V. O. Quine, not only proposed and defended a naturalized epistemology²⁵ but even took many steps forward in that di-

rection, applying behavioristic approaches to the handling of meaning²⁶ and reference²⁷ in the philosophy of language. Philosophers who take the task of searching after truth seriously cannot ignore science on the basis of some all-purpose scepticist argument positing that science needs an epistemological/philosophical foundation if its results are to be accepted. In the same way that philosophers who want to find out about space and time have to take physics into account, philosophers investigating anything that involves human perception, thought, language, memory, or action, have to consult the sciences that have been systematically investigating these faculties for many decades. Since psychology provides the meeting point for all the cognitive sciences, as philosophy again mixed with psychology, it too became a part of cognitive science.

It appears then that, after a successful divorce, the two partners have come together but in a new arrangement for cohabitation. Serious philosophy and thoughtful, theoretically oriented psychology live under the same roof of cognitive science, collaborating on almost all problems, yet maintaining some independence where one of the two has some comparative advantage in handling certain problems. What are some of the rules that govern this new cohabitation and what specific role is philosophy to play in this new living arrangement under the roof of cognitive science?

3 Psychology and mental representations

We have argued that psychology provides the ground on which the edifice of cognitive science is built and that this contribution is primarily on the side of the *explanandum*: it has searched, found, and organized a vast number of truths (most minor, some major) concerning mental faculties. The search for these truths rested on grounds that invited philosophical scrutiny and had already generated philosophical controversy among psychologists themselves (even in the very beginnings of psychology, e.g., imageless thought, unconscious inference). Most of the truths uncovered by psychological research are not only of a statistical nature but also rely on ideas about the divisions of the inner world that are arbitrary or are based on what is called “folk psychology”. Most of these classifications took place without neuroscientific or neuropsychological knowledge and were influenced in many ways by old-fashioned metaphors: in the best case, the tripartite nature of the soul, in other cases acknowledged myths, such as the well-known association of feeling love with the heart and body temperature. Some even rested on bits of universal human experience, like the peculiar connection between smell and memory, celebrated by Proust.

Phenomena associated with universal human experience that can also be observed or detected in

the animal kingdom, such as attention, sensation, and perception, were thought to be sufficiently straightforward to be examined using experimental methods without recourse to serious philosophical scrutiny. But higher faculties such as sentiments and thoughts generate puzzles. It is difficult to identify sentiments beyond the most basic (whose expression is detectable in animal and infant reactions), especially when it comes to questions regarding their relations to thought and motivation. This indeterminacy even affects the experiencers themselves. Moreover, by focusing on the defense mechanisms of the Ego, the psychodynamic influence on psychology made even introspective examination of sentiments appear susceptible to bias and systematic distortion. Even beyond psychodynamic doubts, as emotions are subject to development and cultivation in the developing human, private edifices of emotional makeup, coupled with what is considered socially appropriate to feel or express, lead to an extremely confused picture, and the associated research has remained inconclusive. Nostalgia, to give just one example, made its official debut as a pathological disturbance that led to aberrant behaviors in persons forced to permanently live and work away from home, but was later recognized and even encouraged as a normal mature human feeling and even as a sought-after characteristic of artistic expression.

Moving from sentiments to thought, in the passage from sensations to higher cognitive functions memory (normally associated with learning) was an intermediate step considered accessible to psychological research. Here, with Thorndike showing the way, psychologists began by investigating learning in animals. They placed them in escape mazes and measured memory in terms of speed and errors in subsequent performances. Leaving aside the many kinds and aspects of memory, the field is an easily accessible experimental ground in which psychology reigns supreme. Still, it did introduce a fundamental notion that will remain a key source of philosophical puzzlement and involvement, the idea of representation (in the absence of any initial stimulus), which in spite of ontological and methodological doubts, remains with us to the present. This characteristic is readily extendable to the mental in its entirety. As Brentano has argued, intentionality specifically characterizes the mental, since it comprises acts that are aimed at objects which do not have to exist or exist in the same way as the actual intended object. The difficulties surrounding this notion usher philosophy into psychological theorizing in ways that cannot be ignored unless one decides to arbitrarily ban or ignore mental representations *tout court*. Even the Behaviorists' daring decision to study learning as the observable change of behavior in response to environmental stimuli, with no recourse to what is going on inside the skin or

“inside” the mind, finally led serious Behaviorists²⁸ to introduce representations by way of cognitive maps, readiness, pre-conditionings, predictions. Further violations of the *Classical and Operant Conditioning* models, in phenomena like “learned helplessness”, show that learning cannot be divorced from sentiments or replaced simply by a neutral (or neutralized), observable reinforcement concept.

Representation is the key notion that all the disciplines in cognitive science have employed explicitly or implicitly in all research or theorizing. Given the enormous complexity of this Ur-notion, philosophy has to be constantly on call (and on guard) to handle the specifics of its employment and the implications and limitations of any experimental findings or theoretical proposals. Philosophy needs to make multiple contributions here to handle the many aspects, facets, kinds, characteristics, levels, and dimensions of representation.

4 Causal and formal aspects of representation in cognitive science

Philosophy’s encounter and dealings with representation run all the way back to Plato’s cave parable and Aristotle’s brilliant insight that the mind, in its various cognitive functions, handles abstracted forms (not the actual matter) of things. Today, we see how cognitive science struggles with and employs the idea of representation or may explicitly call for assistance from philosophy as in the case of *Knowledge Representation*²⁹, which resurrected the old problem of categories. Our past and present philosophical involvement with representation not only runs deep but also wide, as wide as Nelson Goodman’s³⁰ masterful treatment in *Languages of Art*, where depiction, description, exemplification, and expression are examined as types of symbolic representation in the arts (and not only in the arts, but also in other “ways of worldmaking”).

From an ontological point of view, representation must be classified with relations. All the puzzling relations (causality and similarity are the *par excellence* examples) that hold between real things (not ideal or artificial or conventional ones) encounter the difficulty of determining how many *relata* are involved. At the outset we may stipulate that it has to involve at the very least two *relata*, the representing and the represented, something which is easily discernible in the case of the conventional “symbol-object symbolized” relation. In the case of natural (not conventional) representation relations, what often makes the relation problematic is the well-known “process-product” ambiguity: the process of representation becomes the product representing, a difficulty that can be extended in the opposite direction as well, given the multiple stages of representation involved from

stimulus to mental image and, even further, to the word describing the mental experience. This problem also holds for conventional representations, although these are easier to disentangle.³¹ Another dimension of complication arises for the conventional employment of symbols from the pragmatics of representation, when we take into account the one who introduces the representing symbol and those to whom the representation is addressed (even those that are to be excluded from understanding the symbolism) and the conventions that are taken for granted in this situation.³² As we live in a world of mentality, surrounded mainly and often solely, by objects and symbols that presuppose intelligence of various kinds, learning the ways in which representations are to be interpreted, understood, and adopted, becomes the basis of our education and enculturation into responsible adulthood. Both difficulties outlined above require philosophy’s assistance in the ways explored below.

In a preliminary stage, philosophy is needed to avoid or accordingly clarify the process-product ambiguity contained in representation. This is not a once-for-all task; it has to be repeated in steps as previous processes become products in the next step or vice versa. The task is not a pedantic distinction concerning nuances of meaning, as processes and products differ in terms of the kind of intentionality they involve. In the second place with respect to conventional representations, we begin, following Searle,³³ by distinguishing real intentionality (what any cognitive function has as its object/content) with the “as-if intentionality” of the employed symbol which is parasitic on the real intentionality of the user and intended receiver of the symbolic representation. This “as-if intentionality” is not limited to the arts and discourses that are properly the objects of hermeneutic disciplines of interpretation (i.e., the Humanities), since it relates to a central system of conventional representation that has acquired natural status and directly affects mental and brain processes. I am talking here about language which is learned in conventional social settings that employ different languages, all of which are nevertheless grounded in specific biological structures and exhibit universal characteristics, such as universal grammar and, perhaps, emerge from the language of thought. There is evidence that once language is engaged, even simple pain alarm functions in the brain are altered.³⁴

There are two main aspects (or sides, if you prefer) of representation which are often intertwined so that careful philosophical work is needed to distinguish which part of which is involved in which part of the other. These two aspects of representation are the formal and the causal. The formal refers mainly to correspondences between various “representings” and “representeds” (and parts thereof) and it mainly detects similarities. It reaches deeper into questions of abstraction, ref-

erence, isomorphism, truth, and categoricity, but its main advantage is that it becomes the ground that allows for compositionality, which is one of the characteristics of the mental, at least in its human language-infected instantiation. The causal aspect of representation refers to processes by which elements and parts of the represented object or event and its inner representing counterpart are connected in actual temporal sequences that lead from the world “out there” to the brain inside the skull, and to the mind that has emerged and is experienced in the brain. This causal aspect is connected to objectivity, grounding, reference, and necessity but its greatest significance for cognitive science is that it allows for causal intervention and interference which, in turn, yield experimental, or at least observable, facts. These two aspects, the formal and the causal, are intimately connected in our search for the mental, so that even straightforward laboratory research needs philosophical assistance to disentangle them and propose further even crucial experiments. To mention just a few examples, consider Pylyshyn’s³⁵ cognitive penetrability experiments in mental imagery, Libet’s experiments on conscious decision,³⁶ or McKoon and Ratcliff’s experiments in the language of thought.³⁷

The causal aspect of representation has two roles to play here. The first is that it allows for experimental approaches adding, subtracting or modifying factors in the represented or the representing factors at each of the many stages in a process or processes. Ever since Fechner noted the existence of an afterimage when the stimulus was absent, there has clearly been space for the added dimension of the mental (an ‘in the mind’ in addition to what is “outside the skin” and “in the brain”). The renewal of solid cognitive research in the work of Sheppard³⁸ and Kosslyn³⁹ relied on the absence and recall of the absent stimulus as a mental image. But the converse was also employed in Logothetis’ experiments on binocular rivalry, where the stimulus is still present, but the final representation is absent from consciousness.⁴⁰ Neurological evidence for serious modifications to representations of the world due to lesions and disturbances of intermediate steps in processing abound in the work of O. Sacks.

The second important role of the causal aspect of representation is that it promises to be the way in which a reductive explanation of the mind could finally be achieved. There are good arguments in favor of the impossibility of a final and complete reduction of at least one important aspect of the mental, that is, the first-person subjective experience to the objectively third-person observable neurobiological substrate: T. Nagel’s bats,⁴¹ S. Kripke’s (and D. Chalmers’) zombies,⁴² the induced color blindness of F. Jackson’s Mary,⁴³ D. Chalmers’ hard problem,⁴⁴ all of which can be neatly subsumed un-

der J. Levine’s explanatory gap argument.⁴⁵ As is the case with other scientific endeavors which leave fundamental problems unsolved, there is good reason for cognitive science to be happy that the explanatory gap is constantly narrowing and will continue to do so. Even if we have to live with a small, albeit diminishing, explanatory gap, this is a gap that “runs through us”, it is ‘our’ explanatory gap, a devil we know and live with.

The two aspects of representation, together with their attending concepts and relations, have peculiar relations to logic. Even though they are foundational presuppositions of logic, they turn out to be recalcitrant to a full logical analysis. The formal aspect, resting on similarity, abstraction, and analysis, is a *sine qua non* presupposition for doing logic. At the same time, abstraction is a vague operation yielding different results regarding different aspects of form. Similarity is an equally, if not more, vague relation that any object has to itself and to anything else with some qualification, forcing us to always specify the respect in which something is similar to something else. When we compare it to other “logically tame” relations, like identity or equality, it is clear that it is not an equivalence relation: it lacks symmetry and transitivity, except in restricted mathematically defined contexts, like similar triangles.

In a similar manner, causality lacks a clear truth-functional profile and is also not an equivalence relation as it lacks both reflexivity (there is no “causa sui”) and symmetry (indeed it has to be asymmetrical to allow for the temporal succession element of causality), while its transitivity holds only in restricted contexts. On the other hand, however, though causality seems too temporal to be analyzed by our atemporal first-order logic, still the key logical connective of the conditional (which spells out the key logical relation of implication, tested by the tautology of the conditional) is surely related to what the organism causes to happen and what it predicts will happen, or not happen, as a result of its actions. An ontologically parsimonious characterization of the mind/brain as “doer and predictor” surely needs an “if-then” construction as a quasi-logical basis that straddles the present and the immediate future or the immediate past.

The problems of logically analyzing causality spills over to the problem mentioned at the outset, the indeterminacy in the number of the *relata* involved in causal attributions. The point is obvious: how many factors are to be considered as having influenced an event? How far back should the search go, and how wide should this search be?

Hume had shown that the attribution of causality requires an inductive leap from “constant conjunction in the past” to an “always” to be provided not by logic but by psychology. Yet even in the case of singular attribution of causality in the past (“a caused b”), what is called for is a strong

condition of necessity stating that the result “b” would not have happened, had the cause “a” not happened. This appeal to counterfactual conditionals is not only logically suspect but generates further hopeless indeterminacy when counterfactuals are connected to further counterfactuals. To cut through this *conundrum*, law courts appeal to common sense to measure, attribute and allocate blame.⁴⁶ In addition, even in the case of attributing causality in normal scientific (not legal) contexts, Hume’s third condition leads to lawlike generalizations. When we try to determine “lawlike”, we need recourse to counterfactuals, and when we try to establish counterfactuals, we need recourse to lawlike-ness. The indeterminacy this time turns out to be of the “chicken and egg” type.

Applying the same insight from causality to representation, even in clear-cut cases of conventional symbolic representations, the relation is never a two-place connection between representing and represented, but at least a five-place connection: symbol *s* represents object *o* for user *u* who intends *s* to be understood by receiver *r* under convention *c*. Even in the case of road signs that are to be understood by everybody (illiterates included), there is evidence that what the sign indicates often fails to convey the intended message to the vast majority of users. When we move away from conventions and consider mental representations, there is ample literature to show that meaning is fluid and indeterminate. Even in something as common as memory recall, E. Loftus⁴⁷ has shown that recall is not a reproduction of the past but a reconstruction. The plasticity of the brain and the intimate connection between concepts and induction indicate that even for an individual user the meaning of a mental representation remains unstable. This holds not only for “belief-constituted concepts” (to use A. MacIntyre’s term), but even for the simplest of qualia as to acceptable taste, which are subject to change, once one tastes something better, or something different, or learns something “distasteful” about what was ingested. Such changes can be radical, as the initial rejection of the taste of beer is replaced by enthusiasm, as in the celebrated example of D. Dennett. Although the notion of representation has generated opposition, for doubling our difficulties instead of helping with them,⁴⁸ there is hardly any doubt that it enjoys some kind of objective footing. Even by the most naturalistic and anti-mentalistic reckoning (that of B.F. Skinner), all that must be admitted is that there has to be some kind of entity or process inside the skin of the organism that corresponds or somehow relates to an event outside the skin that acts as a stimulus. Even in this minimal ground for representation, the two aspects of representation (causality and correspondence) come to the surface. The connection between the representation (the product) and

represented is a causal one, i.e., a process with many steps serving as products, or subprocesses, and not necessarily a one-to-one relation. We know, for example, that color perception involves more than one area of the brain. To make matters even more complicated, the direction of causality is never one way (from the represented to the representation), but runs both ways, often involving re-entries, even for something as basic and simple to comprehend as the alarm system of pain, which seems universal in the animal kingdom.

The task of clarifying the relations between the formal and the causal aspects of representation involves a host of concepts all of which are seen as part of philosophy’s involvement with our inductive practices and thought, which began as part of philosophical psychology in Hume’s thought, became part of epistemology after Kant, and continues to the present as part of the philosophy of science proper. To the extent that inductive practices are a main constituent of human intelligence, extending back to animal intelligence by way of learning, memory, discrimination, generalization, association, prediction, much of inductive thought has come under investigation by psychological research which has yielded important finds regarding confirmation bias (and other types of bias), in truths about concepts, and even theories of concepts (neatly separable into theories of similar features and theories of theory),⁴⁹ and even led Oaksford and Chater⁵⁰ to adopt a Bayesian interpretation of probability employed by the Personalist school of inductive logic, in the direction of a general approach to the mind as a predictor system⁵¹ in which truth, falsity, and uncertainty have an intimate connection to sentiment.⁵² We said earlier that psychology provides the ground (by way of providing the *explanandum*) on which the edifice of cognitive science is to be built, but it is the philosophy of science that offers the next step of evaluating and charting the ground and providing in this way a forum where cognitive scientific thinking can meet.

5 Conclusion: On the future of philosophy in cognitive science

Philosophy, then, has a major role to fulfill in cognitive science and will continue to do so as long as representation occupies center stage in cognitive scientific thinking. Specifically, philosophy allows for careful work clarifying the complex relations between the causal and the formal aspects of representation which are pertinent to both experimental and theoretical work. Its role however is not limited to providing ancillary distinctions for others to keep in mind while experimenting or theorizing.

Philosophy as part of cognitive science is in a position to propose small or grand hypotheses and theories, so it should have a part in testing and establishing such theoretical leaps in cooperation with the

other sciences, most often but not only psychology and linguistics. Fodor proposed at least two grand hypotheses: *Modularity Architecture* and *Language of Thought*. In his arguments in defense of his proposals, it is difficult to separate the philosopher, the linguist, and the psychologist. In an equally inspired way, Patricia Churchland coins the neologism “neuro-philosophy”⁵³ to characterize the attempt to solve problems in the philosophy of mind that are informed by new developments in neurobiological research. We expect that as the neurosciences uncover many new truths in the context of cognitive science, even common language is bound to change as a result of newer more accurate classifications of what is happening in our inner world.⁵⁴

When it collaborates on cognitive science, philosophy is no longer part of the history of ideas (offered for the appreciation and enrichment of life). Though philosophical problems appear to be eternal, certain advances in the sciences can and do solve such problems, not as an all-or-nothing triumph of one “-ism” over all other “-isms”, but as a piecemeal solution to parts or aspects of the problem.

The solution to Molyneux’s Problem provided by later scientific discoveries, should not be seen as a victory of nativism over empiricism or vice versa, but as a careful account of what can be considered innate in the relevant cognitive functions and what is the product of learning or development, at what stages and under what conditions and in which animals.⁵⁵ In the same light, the problem of free will does not lead to a resolution of the three-way combat between libertarianism, soft determinism, and hard determinism, but is resolved by understanding the many ways in which the nervous system exercises control over bodily function, behavior and even thought itself. There is no one Grand Riddle to be solved differently by the many grand “-isms” but many smaller riddles that can be solved with adequate empirical support. Such solutions can be provisionally accepted by most of those concerned.

In December 1879, Wilhelm Wundt lifted himself up from his philosophical armchair, rolled up his sleeves and entered his newly founded laboratory at the University of Leipzig. Before becoming, in this way, the first experimental psychologist to initiate the divorce of psychology from philosophy, he had been occupying the chair of Inductive Philosophy at various European academic institutions. Could some inductive philosophy be the kind of philosophy needed by cognitive science?

Notes

¹ Cf. J.A. FODOR, *The modularity of mind*.

² Cf. J.A. FODOR, *The language of thought*; J.A. FODOR, *LOT 2*.

³ Cf. J. SEARLE, *Minds, brains and programs*.

⁴ Cf. D. MARR, *Vision*; M. MINSKY, *The society of mind*.

⁵ Cf. A.M. TURING, *Computing machinery and intelli-*

gence.

⁶ Cf. H. DREYFUS, *What computers can’t do*; H. DREYFUS, *What computers still can’t do*.

⁷ Cf. J. SEARLE, *Could a machine think?*

⁸ Cf. P.M. CHURCHLAND, *The engine of reason the seat of the soul*; P. SMITH CHURCHLAND, P.M. CHURCHLAND, *Could a machine think?*

⁹ Cf. S.M. KOSSLYN, J.M. BALL, B.J. REISER, *Visual images preserve metric spatial information*; S.M. KOSSLYN, *Image and mind*.

¹⁰ Cf. Z.W. PYLYSHYN, *What the mind’s eye tells the mind’s brain*; Z.W. PYLYSHYN, *Imagery and artificial intelligence*; Z.W. PYLYSHYN, *The imagery debate*.

¹¹ Let us at the outset rule out one kind of philosophical involvement in the application of the sciences, especially those whose aims are more practical than theoretical: questions of moral application. This is not what we mean by “the philosophy of a science”, since it only relates to ethics or politics. Medical science has applied aspects as do all the technological sciences. Their philosophical involvement in such aspects have acquired names, such as “medical ethics”. Yet for medicine and for the science of AI, there are other theoretical aspects that deserve the name “philosophy of medicine” or “philosophy of AI”.

¹² Cf. B.F. SKINNER, *Behaviorism at fifty*.

¹³ In psychiatry, Claude Bernard’s idea is disputed: some maintain that mental illness is continuous with normal conditions, others insist on sharply distinguishing the disordered from the normal mind, while others opt for eliminating mental disorder altogether on sociological grounds.

¹⁴ Cf. D. KAHNEMAN, A. TVERSKY, *Subjective probability: A judgment of representativeness*; D. KAHNEMAN, A. TVERSKY, *On the psychology of prediction*; D. KAHNEMAN, *Who shall be the arbiter of our intuitions?*

¹⁵ Surviving even in the present discussion in J. Levine’s *Explanatory gap argument* (cf. J. LEVINE, *Materialism and qualia: The explanatory gap*; J. LEVINE, *Purple haze: The puzzle of consciousness*).

¹⁶ Cf. R.A. BROOKS, *Intelligence without representation*.

¹⁷ In three ways: neuroscientific, evolutionary, behaviorist.

¹⁸ A formal syntactic and, hopefully, semantic endeavor.

¹⁹ With a historical and ethnographical component.

²⁰ Aiming at self-understanding, not prediction or control.

²¹ From the most naturalistic to the most anti-naturalistic: neurobiological, evolutionary, behaviorist, psychodynamic, Gestalt, cognitivist, social, humanistic.

²² Cf. B. RUSSELL, *My philosophical development*.

²³ Cf. L. WITTGENSTEIN, *Philosophical investigations*.

²⁴ Cf. G. RYLE, *The concept of mind*.

²⁵ Cf. W.V.O. QUINE, *Ontological relativity and other essays*.

²⁶ Cf. W.V.O. QUINE, *Word and object*.

²⁷ Cf. W.V.O. QUINE, *The roots of reference*.

²⁸ Cf. E. TOLMAN, *Cognitive maps in rats and men*; J. GARCIA, R.A. KOELING, *Relation of cue to consequence in avoidance learning*; R.A. RESCORLA, R.L. SOLOMON, *Two-process learning theory*; N.J. MACINTOSH, *Conditioning and associative learning*.

²⁹ Cf. J.F. SOWA, *Knowledge representation: Logical, philosophical and computational foundations*; A.B. MARKMAN, *Knowledge representation*.

³⁰ Cf. N. GOODMAN, *Languages of art*; N. GOODMAN, *Problems and projects*; N. GOODMAN, *Ways of worldmaking*.

³¹ Compare, for example: (1) The representation (process) of the Last Supper occupied Da Vinci for 3 years, with (2) The representation (product) of the Last Supper has faded with the years, and with (3) The representation of the Last Supper follows St. John's Gospel (process or product?)

³² Drawing or gesturing the abstract shape of a fish in the early days of Christianity, for example, was a way for someone to declare their Christian faith to someone they encountered who, if Christian, would have understood that the fish symbol referred (by Greek acronym) to "Jesus Christ God's Son Savior", but if not would think the symbol was vaguely about fish.

³³ Cf. J. SEARLE, *Consciousness and language*.

³⁴ Cf. P. REINVILLE, G.M. DUNCAN, D.D. PRICE, B. CARRIER, M.C. BUSHNELL, *Pain affect encoded in human anterior cingulate but not somatosensory cortex*.

³⁵ Cf. Z.W. PYLYSHYN, *What the mind's eye tells the minds brain*; Z.W. PYLYSHYN, *The imagery debate*.

³⁶ Cf. B. LIBET, *Cortical activation in conscious and unconscious experience*; B. LIBET, *Mind time*.

³⁷ Cf. R. MCCOON, G. RATCLIFF, *Priming in episodic and semantic memory*; R. MCCOON, G. RATCLIFF, *Spreading activation versus compound cue accounts of priming*.

³⁸ Cf. R.N. SHEPARD, *The mental image*; R.N. SHEPARD, L.A. COOPER, *Mental images and their transformations*.

³⁹ Cf. S.M. KOSSLYN, J.M. BALL, B.J. REISER, *Visual images preserve metric spatial information*.

⁴⁰ Cf. N. LOGOTHETIS, J.D. SCHALL, *Neuronal correlates of subjective visual perception*.

⁴¹ Cf. T. NAGEL, *What is it like to be a bat?*.

⁴² Cf. D. CHALMERS, *The conscious mind*.

⁴³ Cf. F. JACKSON, *What Mary didn't know*.

⁴⁴ Cf. D. CHALMERS, *Facing up to the problem of consciousness*.

⁴⁵ Cf. J. LEVINE, *Materialism and qualia*; J. LEVINE, *Purple haze*.

⁴⁶ Appropriately enough, as "cause" in Greek derives from "aitia", i.e., blame.

⁴⁷ Cf. E. LOFTUS, *Leading questions and the eyewitness report*.

⁴⁸ Cf. R.A. BROOKS, *Intelligence without representation*; R.A. BROOKS, *Flesh and machines*.

⁴⁹ Cf. J.S. BRUNER, J.J. GOODNOW, G.A. AUSTIN, *A study of thinking*; M.I. POSNER, S.W. KEELE, *On the genesis of abstract ideas*; D.L. MEDIN, M.M. SHAFFER, *Context theory of classification learning*; E. ROSCH, *Principles of categorization*; S. ARMSTRONG, L. GLEITMAN, N. GLEITMAN, *What some concepts might not be*; L.W. BARSALOU, *Context-independent and context-dependent information in concepts*; L.W. BARSALOU, *Ad hoc categories*; L.W. BARSALOU, *The instability of graded structure*.

⁵⁰ Cf. M. OAKSFORD, N. CHATER, *Bayesian rationality*.

⁵¹ Developmental psychologists detect Bayesian inference even in 12-month old infants' cognitive dealings with the world, cf. E. TENGLAS, E. VUL, V. GIROTTO, M. GONZALEZ, J.B. TENEBBAUM, L.L. BONATTI, *Pure reasoning in 12-month-old infants as probabilistic inference*.

⁵² Cf. S. HARRIS, S.A. SCHETH, M.S. COHEN, *Functional neuroimaging of belief, disbelief and uncertainty*.

⁵³ Cf. P. SMITH CHURCHLAND, *Neurophilosophy*; P. SMITH CHURCHLAND, *Brain-wise*.

⁵⁴ While ordinary parlance has refused to change and replace "The sun rises" with its less poetic, scientifically correct equivalent, changes that affect our views on our 'inner world' are more readily adopted in common par-

lance. For example, most of the terms introduced by psychodynamic theories (repression, displacement, rationalization, oedipal, neurotic, etc.) have been adopted even though they were questioned by many psychologists and philosophers of science.

⁵⁵ A contemporary example of this old problem in a modern cognitive science setting in V. OCCELLI, *Molyneux's question: A window on cross-modal interplay in blindness*.

Literature

ARMSTRONG, S., GLEITMAN, L., GLEITMAN, N. (1983). *What some concepts might not be*. In: «Cognition», vol. XIII, n. 3, pp. 263-308.

BARSALOU, L.W. (1982). *Context-independent and context-dependent information in concepts*. In: «Memory and Cognition», vol. X, n. 1, pp. 82-93.

BARSALOU, L.W. (1983). *Ad hoc categories*. In: «Memory and Cognition», vol. XI, n. 3, pp. 211-227.

BARSALOU, L.W. (1987). *The instability of graded structure: Implications for the nature of concepts*. In: U. NEISSER (ed.), *Concepts and conceptual development*, Cambridge University Press, Cambridge, pp. 101-140.

BROOKS, R.A. (1991). *Intelligence without representation*. In: «Artificial Intelligence», vol. XLVII, n. 1-3, pp. 139-159.

BROOKS, R.A. (2002). *Flesh and machines*, Vintage, New York.

BRUNER, J.S., GOODNOW, J.J., AUSTIN, G.A. (1956). *A study of thinking*, Wiley, New York.

CHALMERS, D. (1995). *Facing up to the problem of consciousness*. In: «Journal of Consciousness Studies», vol. II, n. 3, pp. 200-219.

CHALMERS, D. (1996). *The conscious mind*, Oxford University Press, Oxford/New York.

CHURCHLAND, P.M. (1995). *The engine of reason the seat of the soul*, MIT Press, Cambridge (MA).

DREYFUS, H. (1979). *What computers can't do* (1972) Harper & Row, New York.

DREYFUS, H. (1992). *What computers still can't do*, MIT Press, Cambridge (MA).

FODOR, J.A. (1973). *The language of thought*, MIT Press, Cambridge (MA).

FODOR, J.A. (1983). *The modularity of mind*, MIT Press, Cambridge (MA).

FODOR, J.A. (2008). *LOT 2*, Oxford University Press, Oxford.

GARCIA, J., KOELING, R.A. (1966). *Relation of cue to consequence in avoidance learning*. In: «Psychonomic Science», vol. IV, n. 3, pp. 121-124.

GOODMAN, N. (1968). *Languages of art*, Bobbs and Merrill, New York.

GOODMAN, N. (1973). *Problems and projects*, Bobbs and Merrill, New York.

GOODMAN, N. (1978). *Ways of worldmaking*, Bobbs and Merrill, New York.

HARRIS, S., SCHETH, S.A., COHEN, M.S. (2008). *Functional neuroimaging of belief, disbelief and uncertainty*. In: «Annals of Neurology», vol. LXIII, pp. 141-147.

JACKSON, F. (1986). *What Mary didn't know*. In: «The Journal of Philosophy», vol. LXXXIII, n. 5, pp. 291-295.

KAHNEMAN, D. (1981). *Who shall be the arbiter of our intuitions?*. In: «Behavioral and Brain Sciences», vol. IV, n. 3, pp. 339-340.

KAHNEMAN, D., TVERSKY, A. (1972). *Subjective proba-*

- bility: *A judgment of representativeness*. In: «Cognitive Psychology», vol. III, n. 3, pp. 430-454.
- KAHNEMAN, D., TVERSKY, A. (1973). *On the psychology of prediction*. In: «Psychological Review», vol. LXXX, n. 4, pp. 237-251.
- KOSSLYN, S.M. (1980). *Image and mind*, Harvard University Press, Cambridge (MA).
- KOSSLYN, S.M., BALL, J.M., REISER, B.J. (1978). *Visual images preserve metric spatial information*. In: «Journal of Experimental Psychology: Human Perception», vol. IV, n. 1, pp. 47-60.
- LEVINE, J. (1983). *Materialism and qualia: The explanatory gap*. In: «Pacific Philosophical Quarterly», vol. LXIV, n. 4, 1983, pp. 354-361.
- LEVINE, J. (2001). *Purple haze: The puzzle of consciousness*, Oxford University Press, Oxford.
- LIBET, B. (1965). *Cortical activation in conscious and unconscious experience*. In: «Perspectives in Biology and Medicine», vol. IX, n. 1, pp. 77-86.
- LIBET, B. (2004). *Mind time*, Harvard University Press, Cambridge (MA).
- LOFTUS, E. (1975). *Leading questions and the eyewitness report*. In: «Cognitive Psychology», vol. VII, n. 4, pp. 560-572.
- LOGOTHETIS, N., SCHALL, J.D. (1989). *Neuronal correlates of subjective visual perception*. In: «Science», vol. CCXLV, n. 4919, pp. 761-763.
- MACINTOSH, N.J. (1983). *Conditioning and associative learning*, Oxford University Press, Oxford.
- MARKMAN, A.B. (1999). *Knowledge representation*, LEA, London.
- MARR, D. (1982). *Vision*, MIT Press, Cambridge (MA).
- MCCOON, R., RATCLIFF, G. (1979). *Priming in episodic and semantic memory*. In: «Journal of Verbal Learning and Verbal Behavior», vol. XVIII, n. 4, pp. 463-480.
- MCCOON, R., RATCLIFF, G. (1992). *Spreading activation versus compound cue accounts of priming*. In: «Journal of Experimental Psychology – Learning, Memory, and Cognition», vol. XVIII, n. 6, pp. 1155-1172.
- MEDIN, D.L., SHAFFER, M.M. (1978). *Context theory of classification learning*. In: «Psychological Review», vol. LXXXV, n. 3, pp. 207-238.
- MINSKY, M. (1987). *The society of mind*, Touchstone, New York.
- NAGEL, T. (1974). *What is it like to be a bat?*. In: «Philosophical Review», vol. LXXXIII, n. 4, pp. 435-450.
- OAKSFORD, M., CHATER, N. (2007). *Bayesian rationality*, Oxford University Press, Oxford.
- OCCELLI, V. (2014). *Molyneux's question: A window on cross-modal interplay in blindness*. In: «Rivista internazionale di Filosofia e di Psicologia», vol. V, n. 1, pp. 72-101.
- POSNER, M.I., KEELE, S.W. (1968). *On the genesis of abstract ideas*. In: «Journal of Experimental Psychology», vol. LXXVII, n. 3, pp. 353-363.
- PYLYSHYN, Z.W. (1973). *What the mind's eye tells the mind's brain*. In: «Psychological Bulletin», vol. LXXX, n. 1, pp. 1-24.
- PYLYSHYN, Z.W. (1978). *Imagery and artificial intelligence*. In: C. SAVAGE (ed.), *Perception and Cognition*, Minnesota University Press, Minneapolis, pp. 19-55.
- PYLYSHYN, Z.W. (1981). *The imagery debate*. In: «Psychological Review», vol. LXXXVIII, n. 1, pp. 16-45.
- QUINE, W.V.O. (1960). *Word and object*, MIT Press, Cambridge (MA).
- QUINE, W.V.O. (1969). *Ontological relativity and other essays*, Columbia University Press, New York.
- QUINE, W.V.O. (1974). *The roots of reference*, Open Court, La Salle.
- REINVILLE, P., DUNCAN, G.M., PRICE, D.D., CARRIER, B., BUSHNELL, M.C. (1997). *Pain affect encoded in human anterior cingulate but not somatosensory cortex*. In: «Science», vol. CCLXXVII, n. 5328, pp. 968-971.
- RESCORLA, R.A., SOLOMON, R.L. (1967). *Two-process learning theory: Relationships between Pavlovian and instrumental learning*. In: «Psychological Review», vol. LXXIV, n. 3, pp. 151-182.
- ROSCH, E. (1978). *Principles of categorization*. In: E. ROSCH, B.B. LLOYD (eds.), *Cognition and categorization*, Lawrence Erlbaum, Hillsdale (NJ), pp. 27-48.
- RUSSELL, B. (1959). *My philosophical development*, Allen and Unwin, London.
- RYLE, G. (1949). *The concept of mind*, Hutchinson House, London.
- SEARLE, J. (1980). *Minds, brains and programs*. In: «Behavioral and Brain Sciences», vol. III, n. 3, pp. 417-424.
- SEARLE, J. (1990). *Could a machine think?*. In: «Scientific American», vol. CCLXII, n. 1, pp. 20-25.
- SEARLE, J. (2002). *Consciousness and language*, Cambridge University Press, Cambridge.
- SHEPARD, R.N. (1968). *The mental image*. In: «American Psychologist», vol. XXXIII, n. 2, pp. 125-137.
- SHEPARD, R.N., COOPER, L.A. (1986). *Mental images and their transformations*, MIT Press, Cambridge (MA).
- SKINNER, B.F. (1963). *Behaviorism at fifty*. In: «Science», vol. CXL, n. 3570, pp. 951-958.
- SMITH CHURCHLAND, P. (1986). *Neurophilosophy*, MIT Press, Cambridge (MA).
- SMITH CHURCHLAND, P. (2002). *Brain-wise*, MIT Press, Cambridge (MA).
- SMITH CHURCHLAND, P., CHURCHLAND, P.M. (1990). *Could a machine think?*. In: «Scientific American», vol. CCLXII, n. 1, pp. 26-31.
- SOWA, J.F. (2000). *Knowledge representation: Logical, philosophical and computational foundations*, Brooks and Cole, Boston.
- TENGLAS, E., VUL, E., GIROTTO, V., GONZALEZ, M., TENENBAUM, J.B., BONATTI, L.L. (2011). *Pure reasoning in 12-month-old infants as probabilistic inference*. In: «Science», vol. CCCXXXII, n. 6033, pp. 1054-1059.
- TOLMAN, E. (1948). *Cognitive maps in rats and men*. In: «Psychological Review», vol. LV, n. 4, pp. 189-208.
- TURING, A.M. (1950). *Computing machinery and intelligence*. In: «Mind», vol. LIX, n. 236, pp. 433-460.
- WITTGENSTEIN, L. (1953). *Philosophical investigations*, Routledge, London.

RICERCHE

Psychology is – and should be – central to cognitive science

Max Coltheart^(a)

Ricevuto: 15 marzo 2022; accettato: 30 novembre 2022

Abstract Cognitive science is typically defined as the multidisciplinary study of mind, with the disciplines involved usually listed as philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology. Furthermore, these six “core disciplines” are generally regarded as having equal status vis-à-vis cognitive science. In contrast to the latter position, I argue that psychology has a special status here: it is central to cognitive science in a way that none of the other five disciplines is. I support this argument via both theoretical and empirical considerations.

KEYWORDS: Psychology; Cognitive Science; Interdisciplinarity/Multidisciplinarity

Riassunto *La psicologia è – e dovrebbe essere – al centro della scienza cognitiva* - La scienza cognitiva viene definita comunemente come indagine multidisciplinare sulla mente e tra le discipline che vi sono coinvolte vengono solitamente indicate la filosofia, la psicologia, l'intelligenza artificiale, la neuroscienza, la linguistica e l'antropologia. Queste sei “discipline fondamentali” sono generalmente considerate come aventi pari dignità nell'ambito della scienza cognitiva. A dispetto di quest'ultima posizione sosterrò, invece, che alla psicologia vada riconosciuto uno status speciale: la sua importanza per la scienza cognitiva è tale da non poter essere eguagliata da nessuna delle altre cinque discipline. A supporto di questa posizione porterò alcune riflessioni di natura teoretica ed empirica.

PAROLE CHIAVE: Psicologia; Scienza cognitiva; Interdisciplinarietà/Multidisciplinarietà

^(a) School of Psychological Sciences, Macquarie University, North Ryde - 2109 Sydney NSW (Australia)

E-mail: max.coltheart@mq.edu.au (✉)



1 What is cognitive science?

ACCORDING TO THE *STANFORD ENCYCLOPEDIA OF PHILOSOPHY*, cognitive science is «the interdisciplinary study of mind and intelligence, embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology».¹ Exactly this definition is widely accepted, including by Wikipedia² and many other sources.³

However, one must consider here the difference between the terms “interdisciplinary” and “multidisciplinary”. Choi and Pak draw this distinction. For them, a field is *interdisciplinary* if it «analyzes, synthesizes and harmonizes links between disciplines into a coordinated and coherent whole». A field is *multidisciplinary* if it «draws on knowledge from different disciplines but stays within the boundaries of those fields».⁴

Is cognitive science meant to be interdisciplinary, or is it meant to be multidisciplinary? Nuñez, Allen, Gao and colleagues ask the same question in a different way.⁵ They approach the question via a distinction drawn by Gardner between *strong* and *weak* conceptualizations of the nature of cognitive science. According to the strong conceptualization, in cognitive science the constituent disciplines «blend together into one seamless cognitive science»: ⁶ this is *interdisciplinarity*. According to the weak conceptualization, in cognitive science the constituent disciplines «call for cooperation among the [...] member disciplines»: ⁷ this is *multidisciplinarity*. Their critique of the field, entitled *What happened to cognitive science?*, argued that cognitive science was intended to be interdisciplinary but has failed to achieve this aim:

Cognitive science has failed to move from a multidisciplinary collection of collaborations (Gardner’s weak version) to a cohesive interdisciplinary field (Gardner’s strong version) – a new science of the mind.⁸

However, the claim that cognitive science originally set out to be interdisciplinary has been disputed. Recently Gentner⁹ asked two of the founding fathers of the field, Don Norman and Allan Collins, whether they had in mind at the beginning that cognitive science was meant to have a cohesive subject matter. Both said “No”, Collins commenting «we just thought that AI and Psychology and Computational Linguistics were addressing related issues, and it would be good to have a journal that focused on the theoretical interactions between them».¹⁰ That is multidisciplinary; and indeed the title of the journal *Cognitive Science*, which was founded in 1977, has always included the subtitle *A multidisciplinary journal*.

If one respects the distinction between interdisciplinarity and multidisciplinary, then in my view the correct way of expressing the standard

definition of the field referred to above is: Cognitive science is the multidisciplinary study of mental processes. Others agree. For example, the definition by French stresses multidisciplinary. According to him, cognitive science is

an attempt to bring together researchers who are interested in a single overarching issue – how the mind works – in order to have them exchange ideas with one another. The fundamental principle of cognitive science today, one that does not involve a predefined list of interacting disciplines, [is] researchers from different fields talking to one another about the mechanisms that give rise to cognition.¹¹

Similarly, Gentner concurs:

cognitive science cannot and should not aim for a crystalized theory [...] I propose instead the metaphor of a multilingual set of people gathered to solve a common problem. It’s unlikely that the six languages will evolve into a new combined language. Rather, each person does their best to become bi- or tri-lingual so that they can learn from others. The most productive interactions are likely to be dyadic or at most triadic, and which ones will take off cannot be predicted in advance. Every now and then, some group will hit on an arena in which enormous progress can be made, possibly leading to a new subfield. And apart from the big breakthroughs, little gems of insight will come floating along at more regular intervals. This includes disagreements – discovering that a neighboring field has made assumptions that contradict one’s own can be quite enlightening.¹²

Whenever the constituent disciplines of cognitive science are listed, one typically finds the particular six listed in the *Stanford Encyclopedia* and *Wikipedia* definitions of cognitive referred to above. These are AI (sometimes referred to as computer science or computing), anthropology, linguistics, neuroscience, philosophy and psychology, and so I will refer to these six as the “core disciplines” of cognitive science. There are also other disciplines (e.g. archaeology, economics or psychiatry)¹³ which are sometimes mentioned as occasional contributors to cognitive science.

But that is not to say that *all* work in any of these six core disciplines counts as cognitive science. Work in AI that focusses on developing automatic face recognition systems is often not intended to elucidate how people recognize faces, but is meant instead to produce practical solutions to engineering problems (e.g. as a part of a passport control system). This point was made in the editorial of the very first issue of the journal *Cognitive Sciences*: «I will list examples [...] of articles which the editors will be biased against no matter how ex-

cellent they be. 1. Descriptions of intelligent systems that do not provide any insight into human¹⁴ processing». ¹⁵ The reason Collins gave for making this point is that such work «belongs exclusively to one of the established disciplines rather than to cognitive science [so] will likely be rejected by the three editors as inappropriate». ¹⁶

The same point applies to the other five core disciplines. Work in *linguistics* sometimes focusses on language as an abstract object without being specifically interested in how language-processing tasks are actually executed. Much *neuroscience* is not about the brain at all, let alone about the mind. Only one subfield of *philosophy* – philosophy of mind – is concerned with mental processes. And not all *psychology* is cognitive psychology.

Cognitive psychology is the study of mental processes, and so is cognitive science: so one might ask what the difference is between the two. The answer is: multidisciplinary. A great deal of work in cognitive psychology does not draw upon any other disciplines i.e. is monodisciplinary: when this is so, such work, in the words of Collins, quoted above, «belongs exclusively to one of the established core disciplines¹⁷ rather than to cognitive science». ¹⁸

2 What is a cognitive scientist?

A first thought: a cognitive scientist is someone with expertise in anthropology, artificial intelligence, linguistics, neuroscience, philosophy and psychology. But that is absurd, of course, because there are no such people.

A weaker version of this thought is: a cognitive scientist is someone with expertise in at least two of the six core disciplines – but even this is a mistake.

Consider the paper by Fodor, Garrett and Bever on the syntactic determinants of sentence complexity.¹⁹ One might think of this as a classic early example of work in cognitive science on the ground that when this paper was written these authors had expertise in different core disciplines of cognitive science: philosophy (Fodor), psychology (Garrett), and linguistics (Bever). They did not have notable expertise outside their own individual disciplines. Nevertheless what they were doing in this research was cognitive science.

This was so even though it does not follow, simply because its authors came from different disciplines, that this piece of research had to be multidisciplinary (i.e. meets this criterion for being cognitive science). Whether a piece of research meets this criterion can only be determined by examining the research itself. So we need to ask: Did the work reported in this paper draw upon more than one discipline?

It did. That is why it counts as a classic early example of work in cognitive science.

Scrutiny of this paper reveals that the theoretic-

cal basis of the work came from linguistics; in particular, the work used the notions of *base structure tree*, *deep structure* and *surface structure*, and these theoretical notions were drawn from (Chomskyan) linguistics.²⁰ The notions were used in developing a theory of sentence processing. This theory, derived from linguistics, was expressed in cognitive-psychological terms. It was then experimentally tested, and the techniques used in the experiments were typical cognitive-psychological ones: paraphrasing sentences and reconstructing sentences from their scrambled words. So though this work did not draw upon the discipline of philosophy (despite the fact that one of the authors of the paper was a philosopher), it was genuinely multidisciplinary, because it did draw upon two distinct disciplines: linguistics and psychology. That is why it qualifies as an example of a piece of work in cognitive science.

This argument provides an answer to the question we are considering (*What is a cognitive scientist?*), the answer being that a cognitive scientist is a person who does cognitive science i.e. who participates in and contributes to multidisciplinary research on mental processes. Whether such people themselves are expert in more than one discipline is irrelevant. It is the work such people contribute to that must be multidisciplinary, not necessarily the people themselves.

3 Cognitive psychology and cognitive science

Could there be research in which there was no involvement of cognitive psychology that nevertheless counts as cognitive science?

Imagine that a group of AI people wanted to develop a new predictive text processing system for mobile phones. It would be bizarre if they did not draw upon what linguistics has told us about semantic and syntactic constraints on sentence structure to guide them in this development, and if they did so they might well produce an effective system. Such a project would be multidisciplinary (AI plus linguistics), and it would result in a system that performs a task that human cognizers can do (namely, enhance their sentence processing by using prediction). But this would not count as a piece of work in cognitive science, because this project need not have any implications for understanding how human readers do predictive processing. If that is what the AI group had wanted to do, they would have had to draw upon what cognitive psychology has told us about this. It is because of the absence of cognitive psychology here that this hypothetical piece of work does not qualify as an example of cognitive science, even though it is multidisciplinary and even though both disciplines involved in the work are core disciplines of cognitive science.²¹

The same point can be made re any pair of cognitive science's core disciplines that excludes cognitive psychology. When linguists and anthropologists get together to study a variety of languages in exotic locations in order to investigate whether there are language universals – properties shared by all languages no matter how dissimilar they seem – the object of study here is language itself, not the mind. That is why this multidisciplinary project again does not count as an example of cognitive science.

This is a theoretical argument in support of the conclusion that (cognitive) psychology is central to cognitive science while the other five core disciplines are peripheral. But empirical support for this conclusion can also be sought – by, for example, analysing the papers published in the journal *Cognitive Science* (which according to Nuñez and colleagues is «the flagship journal of the Cognitive Science Society»)²² to obtain a picture of the current state of the art in cognitive science.

■ 3.1 *The bibliometric work of Nuñez and colleagues (2019)*

Something of this kind was attempted by Nuñez and colleagues. They adopted a bibliometric approach (rather than scrutinising the contents of the papers published in the journal, as I have done and report later in this paper). They used two bibliometric measures: the departmental affiliation of the authors of papers in the journal, and journal – journal citation patterns and environments of papers in the journal.

■ 3.1.1 *Authors' affiliations in Cognitive Science*

Nuñez and colleagues made two points from their author-affiliation data.

The first was that «Overall, data show that after 2000, less than 10% of affiliations came from departments or programs in cognitive science».²³ But such data do not show that the field of cognitive science has been changing over time (e.g. shrinking), because most work in cognitive science has not been carried out in cognitive science department, simply because there are not many such departments; indeed, if there were no departments of cognitive science at all in the world, it would not follow that no work in cognitive science is being done.

And if the number of cognitive science departments in the world's universities has indeed been shrinking, that does not mean that less and less work in cognitive science is being done, since it could equally be the case that more and more work in cognitive science is being done in departments with other affiliations – psychology departments, for example.

The second point these authors made from their analyses of the affiliation data was:

authors with psychology affiliations have continued to be consistently overrepresented among the core disciplines [...] making up more than half of the affiliations in Cognitive Science [...] In stark contrast, anthropologists and philosophers are functionally absent in the journal, accounting for just 1% and 3% of the total affiliations, respectively. Interestingly, affiliations in neuroscience are also underrepresented in the journal (7%). We also observe a small contribution from authors with affiliations in computer science (8.5%) [...] Overall, the disciplinary distribution of affiliations of Cognitive Science authors has been consistently characterized by strong lop-sidedness over the years.²⁴

What bearing do these particular bibliometric findings have on the status of cognitive science? If all of a paper's authors come from departments of psychology, does it follow that the work reported in this paper cannot count as cognitive science? Obviously not. What matters is the nature of the work reported in the paper, not what departments the authors happen to be employed by. That is why the lop-sidedness referred to in the quote above is not informative concerning the status of cognitive science.

■ 3.1.2 *Journal-journal citation patterns and environments: Scientometrics*

Nuñez and colleagues analysed the citation environments²⁵ of papers in the journal for the years 2000, 2007 and 2014. They reached the following conclusions:

For each of the 3 years examined (2000, 2017, 2014), the environment shows a dominance of psychology journals, a nearly complete lack of anthropology and philosophy journals, and a weak presence of neuroscience journals. Over the years the number of journals in the citation environment decreases and becomes less diverse.²⁶

and

In all 3 years, the journal's citation environment showed a complete lack of anthropology and philosophy journals, a weak presence of neuroscience and an overrepresentation of psychology journals. Moreover, across the analyzed period, the number of journals in the citation environment shrank by 23%. This reduction was not uniform; the environment became less diverse, as computer science and other miscellaneous journals were no longer present [... there has been ...] a decline in interdisciplinarity in both the knowledge base and audience of cognitive science's literature.²⁷

Discipline(s)	2020	2021
	N=417 papers submitted N=599 fields selected	N=393 papers submitted N=537 fields selected
Anthropology	2.3	1.5
Biology + neuroscience	8.2	9.9
AI + Computer science	9.3	10.4
Education	4.0	5.6
Linguistics	19.5	18.8
Philosophy	9.0	8.2
Psychology	47.6	45.6

Table 1. For papers submitted to *Cognitive Science* in 2020 or 2021, percentage for each field selected as a function of total number of field selections.

These bibliometric findings are, again, not informative about the status of cognitive science. All they tell us is that authors working in research in cognitive science are publishing their work more and more in psychology journals. It does not at all follow from this that research in cognitive science itself is shrinking, or changing in any way.

3.2 The bibliometric work of Cooper (2019)

At the point of manuscript submission, the journal *Cognitive Science* asks authors to select, as relevant to their paper, any number of fields (including none) from a list of nine fields (these nine fields are listed in *Table 1*). As Cooper notes, such data «provide a more accurate characterization of manuscript content and orientation than departmental affiliation». ²⁸ So Cooper carried out and reported analyses of these field choices for all submissions to the journal between August 1, 2002 and August 31, 2019. Here I report further analyses of this kind for the years 2020-2021, to enable comparisons with the data from the 2021 volume of the journal which I report in Section 3.3.

Table 1 indicates author choices of the fields of their papers for the two most recent complete years of the journal. ²⁹ This table combines data from biology and neuroscience, and from artificial intelligence and computer science, to yield the traditional six core disciplines of cognitive science, and adds a seventh discipline (education).

As reported by Nuñez and colleagues, ³⁰ psychology is much more strongly represented here

than any other discipline; and for this method of analysis, Linguistics is the second-most represented discipline. After Linguistics come AI, Neuroscience and Philosophy, with fairly similar rates of representation (in the range 8.2% - 10.4%), then Education, and lastly Anthropology.

More informative perhaps are data on the number of different fields selected for each paper upon submission. These data are shown in *Table 2*.

One thing stands out here. Although the journal refers to itself as a multidisciplinary journal, around 44% of authors chose just a single discipline as characterizing their paper. If it is correct to infer from this that around 44% of the journal's papers do not report multidisciplinary work, then that is a major departure from the strictures offered in the first issue of the journal by one of its editors, according to which monodisciplinary work «will likely be rejected by the three editors as inappropriate». ³¹

I return to this point in section 3.3. Here I will just point out that if an author chooses during the process of submission only one of the core disciplines, the inference that the paper does not draw upon any of the other disciplines is not a legitimate one. Bibliometric analyses cannot tell us what we want to know here. The only way to establish this securely is to actually read the paper.

3.3 An alternative to the bibliometric approach

I have argued that bibliometric analysis of the contents of the journal *Cognitive Science* has not

Number of fields selected	2020	2021
	N submissions = 417	N submissions = 393
0	14.1	14.8
1	44.6	44.0
2	28.8	33.1
3	9.4	6.9
4	2.2	1.0
5	1.0	0
6	0	0
7	0	0.3

Table 2. For papers submitted to *Cognitive Science* in 2020 or 2021, percentage of submissions as a function of number of fields selected.

yielded results that permit us to draw any conclusions about the state of cognitive science. What must be done is to analyse the actual content of the papers published in this journal – that is, to read them. I have done this for the most recent full volume of the journal.

A recent volume of this journal, *Cognitive Science* (2021), consists of 12 issues containing a total of 131 papers. These are listed in my *Supplementary Table 1* (ST1), indexed in the first column of that table by issue number and name of first author. In this section, I refer to these papers by issue number plus author's first name e.g. 2/Smith.

I read all of these papers so as to be able find the answers to three questions about each paper. These three questions, and the answers to them, were as follows.

Question 1: Did the paper offer any conclusions about mental processes, and if so what were these conclusions? The current *Author Guidelines* for the journal require this: they state that «*Cognitive Science* is an online journal of the Cognitive Science Society focusing on the multidisciplinary study of minds and other intelligent systems».

Column 2 of ST1 provides answers to *Question 1*. There were three papers for which I could not easily identify what conclusions about mental processes the paper was offering. These were:

5/Beekhuizen: This project sought to determine whether the ambiguity of a word and the relatedness of its senses influence its semantic representation in a detectable way by analysing the dictionary definitions of 5395 words. It was found that similarities for monosemes are higher than those of polysemes, which are in turn higher than those of homonyms. But no attempt was made to investigate whether such findings are related to human mental processing. The authors acknowledged this «While we have shown that distributional semantic representations created from natural corpora exhibit this ambiguity structure, it remains as future work to see whether the representations would show the behavioral correlates found in [computational cognitive] models». (p. 20). So this was purely a paper about linguistics.

10/Richter. According to the Abstract: «We provide a neural dynamic account for both perceptual grounding, in which relational concepts enable the attentional selection of objects in the visual array, and for the generation of descriptions of the visual array using relational concepts [...] The model points toward a neural dynamic theory of higher cognition». But nothing is said about what the model might tell us about how humans perform the relevant cognitive-perceptual tasks. So this purely a paper about AI.

11/Jing: This project analysed the Universal Dependencies corpus database to look at 6 forms of dependency in 71 languages. The Abstract says: «These findings challenge received theories of across-the-board effects of complexity on word order and word order variation and call for theoretical models that relativize effects to specific kinds of syntactic structures and dependencies». But no evidence is considered that would allow one to decide whether or not any of these specific kinds of syntactic structures and dependencies have any influences upon human language processing. So this was another paper purely about linguistics.

Thus virtually all of the 131 papers from this volume of the journal did indeed offer conclusions about mental processes.

Question 2: Did the paper draw upon the literature of more than one of the core Cognitive Science disciplines for its rationale, method and/or discussion (i.e. was it multidisciplinary)?

Column 4 of ST1 provides answers to this question. A “No” in column 3 indicates that in my judgement the paper did not draw upon more than one discipline.

There are 54 papers with a No in this column (these included all 3 papers referred to above in relation to *Question 1*): that amounts to 41% of all the papers published in this volume of the journal. If my judgement about each of these papers is correct – the judgement that the paper drew only upon one discipline – then these papers do not report work that is multidisciplinary (which confirms what is reported above in *Table 2*). Hence one might ask how they could be deemed to meet the journal's *Author Guidelines* quoted above. And one might also wonder what to make of the journal *Cognitive Science* referring to itself as *A multidisciplinary journal* if almost half of the papers it publishes are not multidisciplinary.

The remaining 77 papers (59%) all drew upon at least two of the core disciplines i.e. were multidisciplinary. The disciplines that these papers drew upon are listed in Column 4 of ST1. The representations of the core disciplines were: Linguistics 38 papers, AI 26 papers, Philosophy 18 papers, Cognitive Neuroscience 8 papers and Anthropology 6 papers – and Psychology 77 papers i.e. 100% of them. Which brings us to the third question.

Question 3: Did the paper draw upon findings from cognitive psychology for its rationale, method and/or discussion, and if so what parts of that literature were drawn upon?

Column 3 of ST1 provides my answer to this question. My judgement was that every one of the

131 published papers (which included the 3 papers referred to above in relation to *Question 1*) did draw upon findings in cognitive psychology in one way or another (no matter whether the paper was multidisciplinary or not). Column 3 provides the evidence for this claim: it gives examples for each paper of work in cognitive psychology that the paper drew upon.

In sum: Making the judgements upon which ST1 is based involved some element of subjectivity, so others making these judgements might come up with numbers differing slightly from those above. But I believe that anyone who carries out this exercise would reach the same three conclusions that I have reached, namely:

1. Virtually all of the papers in this volume of the journal offered some claims about mental processes and how they work.
2. All of the papers drew upon cognitive psychology.
3. A substantial proportion of the papers – between a third and a half – did not draw upon any discipline other than cognitive psychology, and hence were monodisciplinary rather than multidisciplinary. So something over half of the papers I surveyed reported work that was both multidisciplinary and offered conclusions about mental processes i.e. satisfied the standard definition of cognitive science with which I introduced this paper.

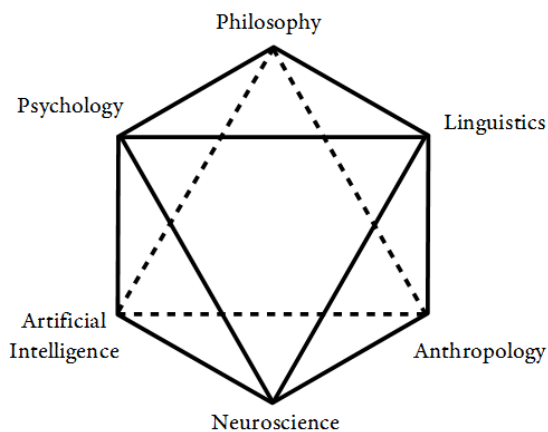


Figure 1(A). The cognitive science hexagon model

4. Conclusions

I therefore propose, for both theoretical and empirical reasons, that psychology has a different status vis-a-vis cognitive science than any other discipline, including the other five core disciplines. To count as cognitive science, a piece of work

must involve cognitive psychology. No other discipline is *necessarily* involved in this way, though there must always be *some* other discipline(s) involved (if there were not, this would be cognitive psychology, not cognitive science).

This is not at all to claim that any piece of work counts as cognitive science only if cognitive psychology *predominates* in that work: I claim only that cognitive psychology *must be present* in that work.

So what I am proposing here is that work counts as cognitive science if it is cognitive-psychological research that has been augmented (supplemented, enhanced, enriched) by taking into account and using findings from one or more of the other five core disciplines (or even other disciplines that are not one of these five).

Figure 1(A) shows the *Cognitive Science Hexagon*.³² This model is often included as an accompaniment to the kinds of definition of Cognitive Science referred to above, in the first paragraph of this paper. I am proposing a different diagrammatic representation of cognitive science, a Cognitive Science Hub and Spokes Model, as shown in *Figure 1(B)*.

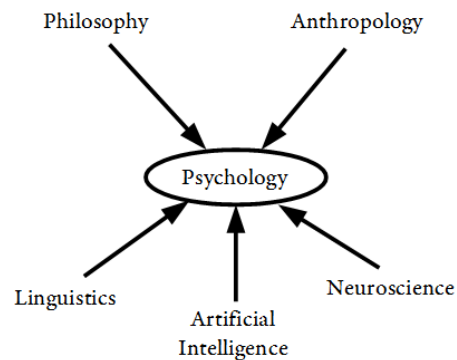


Figure 1(B). The cognitive science hub-and-spokes model

To the best of my knowledge, there are no other treatments of cognitive science which propose, as I have done here, that cognitive psychology is central to cognitive science. However, there certainly is widespread agreement that cognitive psychology happens to be represented much more strongly in cognitive science than any of the other core disciplines, and analyses reported above confirm this.

Numerous authors have expressed dismay about this state of affairs. Cooper refers to views recently expressed in the literature such as «cognitive science is being consumed by cognitive psychology»³³ and «cognitive psychology is taking control of the discipline by stealth».³⁴

Rosenbloom and Forbus consider that «cognitive science has become dominated by cognitive psychology, to the detriment of the original goals of the field».³⁵ Gentner believes that «we in the Cognitive Science movement need to take steps to prevent the “victory” of Psychology over Cognitive Science»,³⁶ and wonders «will Psychology complete its conquest of Cognitive Science?».³⁷ Re the journal *Cognitive Science* and the annual conference of the Society for Cognitive Science, Gentner considers that «Psychology has become far too dominant in the journal and the conference».³⁸ And according to Goel «although cognitive science started in the 1970s as a multidisciplinary field with the goal of becoming an interdisciplinary one over time, it is now dominated by cognitive psychology».³⁹

And in the April 2021 issue of *Cognitive Science*, the incoming chief editor, Rick Dale, had this to say: «I resonate with the concern that our multidisciplinary definition may be at risk due to entrenchment of particular research traditions. Psychology has assumed this role [...] To address these issues in 2021, the Editorial Board will be joined by two philosophers, two anthropologists, two psychologists who are also cognitive neuroscientists, and an expert on human factors and applied cognition».⁴⁰

To the authors quoted in the above three paragraphs, and others who would wish to challenge the picture I have painted of psychology's having a prioritised role in cognitive science, I pose the following two questions, one at a theoretical level and the other at an empirical level:

1. Can you offer some examples – even hypothetical examples – of work that you would count as cognitive science that does not draw at all on findings from cognitive psychology? If not, is there any other discipline that is as ubiquitous in cognitive science as cognitive psychology is?
2. If cognitive psychology does not have a special status for cognitive science, why is it that every one of the multidisciplinary papers in the most recent complete volume of *Cognitive Science* made use of findings from cognitive psychology, whereas the next most represented discipline (Linguistics) is represented in fewer than 50% of these papers?

Some examples of the practice of cognitive science as I see it:

- A cognitive psychologist completes a typical project in cognitive psychology, and then wonders: “Are my conclusions about cognition true only for people from WEIRD⁴¹ societies, or might they also hold in cultures very different from the one I studied?” *calls anthropologist*

to propose a collaboration.

- A cognitive psychologist completes the construction of a model of how people perform some cognitive task, a model in which process X precedes process Y, and then wonders whether there might be brain signatures of when X and Y are happening, and whether there are brain imaging techniques that can tell you whether X really does happen before Y in real time as the task is being performed. *calls neuroscientist* to propose a collaboration.

- A cognitive psychologist completes a very detailed single case study of an aphasic person with a semantic impairment and develops from these data a theory about the nature of semantic representations, then wonders what linguists have said on this topic: might there be linguists who have developed theoretical arguments that support or conflict with this theory about semantic representation? Or might such arguments even be used to refine or extend the theory? *calls linguist* to propose a collaboration.

- A cognitive psychologist studying word recognition wonders how people avoid confusion between sets of words composed of exactly the same letters (now/won/own; pirates/parties) and hears of a technique used in machine learning that has been successful in discriminating items that differ only in the order of their elements. This is the “string kernels” technique.⁴² *calls AI person* to propose a collaboration.

- A cognitive psychologist is interested in learning more about the nature of normal belief processes by studying pathologies of belief such as delusional conditions. The starting point is the proposal by William James that the adoption of a delusional belief is an attempt to explain an unpredicted and unusual experience. The inference from an experience to its explanation is not deductive or inductive: it is abductive. The concept of abductive inference was developed by the philosopher Charles Sanders Peirce. What did Peirce mean by this concept? *calls philosopher* to propose a collaboration.

Of course, the initiator of such multidisciplinary projects does not have to be the cognitive psychologist. The cognitive psychologist can be the person called, rather than the caller.

Acknowledgement

I thank Martin Davies and Jelle Bruineberg for critiques of a previous version of this paper.

Notes

- ¹ Cf. P. THAGARD, *Cognitive science*.
- ² Cf. URL: https://en.wikipedia.org/wiki/Cognitive_science.
- ³ Cf. R. NUÑEZ, M. ALLEN, R. GAO, C.G. MILLER RIGOLI, J. RELAFORD-DOYLE, A. SEMENUKS, *What happened to cognitive science?*, p. 783.
- ⁴ B.C.K. CHOI, A.W.P. PAK, *Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness*, p. 359.
- ⁵ R. NUÑEZ, M. ALLEN, R. GAO, C.G. MILLER RIGOLI, J. RELAFORD-DOYLE, A. SEMENUKS, *For the sciences they are a-changin': A response to commentaries on Nuñez et al.'s (2019) "What happened to cognitive science?"*, here p. 792.
- ⁶ H. GARDNER, *The mind's new science: A history of the cognitive revolution*, p. 389.
- ⁷ *Ibid.*, pp. 389-390.
- ⁸ R. NUÑEZ, M. ALLEN, R. GAO, C.G. MILLER RIGOLI, J. RELAFORD-DOYLE, A. SEMENUKS, *For the sciences they are a-changin'*, p. 800; see also R. NUÑEZ, M. ALLEN, R. GAO, C.G. MILLER RIGOLI, J. RELAFORD-DOYLE, A. SEMENUKS, *What happened to cognitive science?*.
- ⁹ Cf. D. GENTNER, *Cognitive science is and should be pluralistic*.
- ¹⁰ *Ibid.*, p. 887.
- ¹¹ R.M. FRENCH, *Missing the forest for the trees: Why cognitive science circa 2019 is alive and well*, p. 882.
- ¹² D. GENTNER, *Cognitive science is and should be pluralistic*, p. 886.
- ¹³ Consider here journals such as *Cognitive Neuropsychiatry*, which publishes multidisciplinary work combining philosophy and psychiatry (cf., e.g., M. COLTHEART, M. DAVIES, *Failure of hypothesis evaluation as a factor in delusional belief*), as does *Mind and Language* (cf., e.g., T. STONE, A.W. YOUNG, *Delusions and brain injury: The philosophy and psychology of belief*). *Philosophy, Psychiatry & Psychology* is a journal devoted specifically to work in the area of overlap between philosophy, psychiatry, and psychology.
- ¹⁴ The adjective "human" is unnecessarily restrictive here, if there is such a thing as (nonhuman) animal cognition – which is widely recognized (cf. the titles of journals such as *Journal of Experimental Psychology: Animal Learning and Cognition* or *Animal Cognition*). Hence cognitive science includes the study of mental processes in nonhuman animals, as indicated by the fact that studies of nonhuman animals' cognition can be found in the pages of the journal *Cognitive Science* (cf. S.M.W. WOOD, J.N. WOOD, *Distorting face representations in newborn brains*).
- ¹⁵ A. COLLINS, *Why cognitive science?*, p. 2.
- ¹⁶ *Ibidem*.
- ¹⁷ Psychology, in this case.
- ¹⁸ A. COLLINS, *Why cognitive science?*, p. 2 – italics added.
- ¹⁹ Cf. J.A. FODOR, M. GARRETT, T.G. BEVER, *Some syntactic determinants of sentential complexity. II: Verb structure*.
- ²⁰ Cf. N. CHOMSKY, *Aspects of the theory of syntax*.
- ²¹ That is not to say that every project which draws upon two or more cognitive-science disciplines automatically therefore contributes to cognitive science: such a project must in addition yield inferences about human cognition. I thank Jelle Bruineberg for draw-

ing my attention to this point.

- ²² R. NUÑEZ, M. ALLEN, R. GAO, C.G. MILLER RIGOLI, J. RELAFORD-DOYLE, A. SEMENUKS, *What happened to cognitive science?*, p. 788.
- ²³ R. NUÑEZ, M. ALLEN, R. GAO, C.G. MILLER RIGOLI, J. RELAFORD-DOYLE, A. SEMENUKS, *What happened to cognitive science?*, p. 785.
- ²⁴ *Ibidem*.
- ²⁵ The citation environment of a paper consists of the set of papers it cites plus the set of papers which cite it.
- ²⁶ *Ibid.*, p. 785-787.
- ²⁷ *Ibid.*, p. 787.
- ²⁸ R. COOPER, *Multidisciplinary flux and multiple research traditions within cognitive science*, p. 874.
- ²⁹ I am indebted to Rick Cooper for carrying out this analysis and the analysis reported in *Table 2*, and for providing me with the results of these analyses
- ³⁰ Cf. R. NUÑEZ, M. ALLEN, R. GAO, C.G. MILLER RIGOLI, J. RELAFORD-DOYLE, A. SEMENUKS, *What happened to cognitive science?*.
- ³¹ A. COLLINS, *Why cognitive science?*, p. 2.
- ³² Cf. D. GENTNER, *Cognitive science is and should be pluralistic*, adapted from the 1978 Sloan Foundation State of the Art report on cognitive science.
- ³³ R. COOPER, *Multidisciplinary flux and multiple research traditions within cognitive science*, here p. 872.
- ³⁴ *Ibid.*, p. 878.
- ³⁵ P.S. ROSENBLOOM, K.D. FORBUS, *Expanding and repositioning cognitive science*, p. 918.
- ³⁶ D. GENTNER, *Psychology in cognitive science: 1978-2038*, p. 333.
- ³⁷ *Ibid.*, p. 336.
- ³⁸ D. GENTNER, *Cognitive science is and should be pluralistic*, p. 885.
- ³⁹ A. GOEL, *A cognitive reformation*, p. 892.
- ⁴⁰ Cf. R. DALE, *The mindset of cognitive science*.
- ⁴¹ Cf. J. HENRICH, S.J. HEINE, A. NORENZAYAN, *The weirdest people in the world?*, here p. : «western, educated, industrialized, rich and democratic».
- ⁴² Cf. T. HOFMANN, B. SCHOLKOPF, A.J. SMOLA, *Kernel methods in machine learning*.

Literature

- CHOI, B.C.K., PAK, A.W.P. (2006). *Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness*. In: «Clinical and Investigative Medicine», vol. XXIX, n. 6, pp. 51-364.
- CHOMSKY, N. (1966). *Aspects of the theory of syntax*, MIT Press, Cambridge (MA).
- COLLINS, A. (1977). *Why cognitive science?*. In: «Cognitive Science», vol. I, n. 1, pp. 1-2.
- COLTHEART, M., DAVIES, M. (2021). *Failure of hypothesis evaluation as a factor in delusional belief*. In: «*Cognitive Neuropsychiatry*», vol. XXVI, n. 4, pp. 213-230.
- COOPER, R. (2019). *Multidisciplinary flux and multiple research traditions within cognitive science*. In: «Topics in Cognitive Science», vol. XI, n. 4, pp. 869-879.
- DALE, R. (2021). *The mindset of cognitive science*. In: «Cognitive Science», vol. XLV, n. 4, Art.Nr. e12952 - doi: 10.1111/cogs.12952.

- FODOR, J.A., GARRETT, M., BEVER, T.G. (1968). *Some syntactic determinants of sentential complexity. II: Verb structure*. In: «Perception and Psychophysics», vol. III, n. 6, pp. 453-461.
- FRENCH, R.M. (2019) *Missing the forest for the trees: Why cognitive science circa 2019 is alive and well*. In: «Topics in Cognitive Science», vol. XI, n. 4, pp. 880-883.
- GARDNER, H. (1987). *The mind's new science: A history of the cognitive revolution*, Basic Books, New York.
- GENTNER, D. (2010). *Psychology in cognitive science: 1978-2038*. In: «Topics in Cognitive Science», vol. II, n. 3, pp. 328-344.
- GENTNER, D. (2019). *Cognitive science is and should be pluralistic*. In: «Topics in Cognitive Science», vol. XI, n. 4, pp. 884-891.
- GOEL, A. (2019). *A cognitive reformation*. In: «Topics in Cognitive Science», vol. XI, n. 4, pp. 892-901.
- HENRICH, J., HEINE, S.J., NORENZAYAN, A. (2010). *The weirdest people in the world?*. In: «Behavioral and Brain Sciences», vol. XXXIII, n. 2-3, 2010, pp. 61-83.
- HOFMANN, T., SCHOLKOPF, B., SMOLA, A.J. (2008). *Kernel methods in machine learning*. In: «The Annals of Statistics», vol. XXXVI pp. 1171-1220.
- NUÑEZ, R., ALLEN, M., GAO, R., MILLER RIGOLI, C.G., RELAFORD-DOYLE, J., SEMENUKS, A. (2019). *What happened to cognitive science?*. In: «Nature Human Behaviour», vol. III, n. 8, pp. 782-791.
- NUÑEZ, R., ALLEN, M., GAO, R., MILLER RIGOLI, C.G., RELAFORD-DOYLE, J., SEMENUKS, A. (2020). *For the sciences they are a-changin': A response to commentaries on Nuñez et al.'s (2019) "What happened to cognitive science?"*. In: «Topics in Cognitive Science», vol. XII, n. 3, pp. 790-803.
- ROSENBLUM, P.S., FORBUS, K.D. (2019). *Expanding and repositioning cognitive science*. In: «Topics in Cognitive Science», vol. XI, n. 4, pp. 918-927.
- STONE, T., YOUNG, A.W. (1997). *Delusions and brain injury: The philosophy and psychology of belief*. In: «Mind and Language», vol. XII, n. 3-4, pp. 327-364.
- THAGARD, P. (2023). *Cognitive science*. In: E.N. ZALTA (ed.), *Stanford encyclopedia of philosophy*, Fall edition, 1996; substantive revision, 2023, Winter edition – URL: <https://plato.stanford.edu/entries/cognitive-science/>
- WOOD, S.M.W., WOOD, J.N. (2021). *Distorting face representations in newborn brains*. In: «Cognitive Science», vol. XLV, n. 8, Art. Nr. e13021 – doi: 10.1111/cogs.13021.

Appendix

Supplementary Table 1

First author	Claims made re mental processes	Cognitive psychology domains drawn upon	Other disciplines drawn upon
ISSUE 1 - 10 papers			
Pyers	Inhibiting gestures reduces TOT resolution for those with weak verbal STM.	Tip of Tongue. Verbal STM. Spatial STM.	None
De Deyne	Multimodal information is important for capturing both abstract and concrete words. Fully representing word meaning requires more than purely linguistic information.	Semantic networks. Affect.	AI
Chen	Subject relatives with an inanimate head are harder to process than object relative counterparts.	Sentence processing. Surprisal / expectation.	Linguistics; AI
Jraissati	Odors are represented in local conceptual spaces.	Categorization. Conceptual spaces.	Anthropology; Linguistics
Henne	People more frequently elect norm-violating factors, relative to norm-conforming ones, as the cause of some future outcome.	Causal reasoning. Norms.	Philosophy
Goupil	Improvised vs scripted joint actions are more continuous with one another than it first seems, and they differ merely in the extent to which they rely on emergent or planned coordination mechanisms.	Goal representations. Shared intentions.	None
White	Culture and cognition have independent roles in supporting different kinds of supernatural beliefs in distinct cultural contexts.	Mentalizing. Intuitive thinking. Cognitive biases.	Anthropology; Philosophy
Pooney	Within-domain flexibility in humans derives from task representations composed of propositional rules written in terms of objects and relational categories.	Transfer learning. Reinforcement learning.	AI
Mangiarulo	Participants' judgements are deeply affected by impact though they should only depend on probability distributions over features. Dissociation between evidential impact and posterior probability increased number of errors.	Visual features. Probabilistic reasoning.	AI
Wu	Seemingly small differences in languages can cause us to describe the world in surprisingly different ways.	Language production. Referential communication.	Linguistics
ISSUE 2 - 5 papers			
Reins	People have a broader conception of lying than is usually assumed.	Research on lying and on conversational implicature.	Philosophy
Fairchild	Differences in ToM are associated with pragmatic competence in neurotypical individuals	Studies on response to underinformative sentences and on scalar implicatures, including in autism.	Philosophy; Linguistics
Zhang	Instructed hand movements – even when presented as an unrelated, secondary task – can affect students' learning of a complex concept.	Studies of role of gesture in teaching and learning. Studies related to embodied cognition.	None
Stocco	Individual differences in reward processing underpin human fluid reasoning abilities.	Research on relationship between working memory and Ravens performance and on strategies used in Ravens performance.	Cognitive neuroscience; AI
Dubey	Social information plays an important role in shaping our curiosity.	Research on curiosity, and its role in learning. Research on social influences on decision and action.	None
ISSUE 3 - 6 papers			
Hawkins	In communicative situations, people are not simply "mind-blind"; they use contextually appropriate expectations to navigate the division of labor with their partner.	Research on perspective-taking in communicative situations.	Philosophy
Freudenthal	The factors that conspire to make English verb inflection particularly challenging for learners with weak sequential learning abilities are much reduced or absent in Spanish. This provides an explanation for why learning Spanish verb inflection is relatively unaffected in children with DLD.	Specific language impairment. Developmental language disorder. Sequential learning.	Linguistics; AI
Zhang	In speech perception there is dynamic mapping of input to category representations and it is flexibly tuned according to interactive processing accommodating both lexical knowledge and idiosyncrasies of the acoustic input.	Literature on speech perception extensively used.	None
Ou	Categorization gradience may be a consistent within-individual property in speech perception. Listeners who	Literature on speech perception extensively used.	None

	show greater categorization gradience tend to adopt a buffered processing strategy, especially when cues arrive asynchronously in time.		
Garofalo	Adjectives can shape the sensorimotor activation elicited by nouns of graspable objects, highlighting that language simulation goes beyond the single-word level.	Extensive use of literature concerning how affordances derive from perception-action patterns stored in memory, resulting from consolidated and constant (or relatively constant) experiences across different contexts of hand-object interaction.	Cognitive neuroscience; Linguistics.
Van der Mijl	Assessed the role of attention in a naturalistic setting that more directly generalizes to real-world settings than typical laboratory studies.	Research on interval timing, foreperiod effects, the Attentional Gate theory, and the role of attention in timing.	None
ISSUE 4 – 22 papers			
Balzarotti	Editing density of moving images by increasing the number of cuts effectively altered viewers' experience of time.	Time perception.	Cognitive neuroscience
Yoon	Perceptions about the relevant discourse history are a key determinant of how language is used in the moment. Conversational partners form asymmetric representations of the discourse history.	Psycholinguistic studies of discourse comprehension.	None
Sun	Visual complexity is extracted efficiently and automatically, and even arouses a kind of "perceptual curiosity" about objects that encourages subsequent attentional engagement.	Behavioural work on the determinants of perceived visual complexity.	None
Luthra	There is strong evidence in favor of computational models of spoken word recognition that include top-down feedback.	Context effects in spoken word recognition and phoneme identification.	Cognitive neuroscience
Hamami	A search procedure for counterexamples, which proceeds object-wise, could underlie diagram-based geometric reasoning. Transposing points, lines, and circles to our spatial environment, the present study may ultimately provide insights into how humans reason about topological relations between positions, paths, and region.	Literature on spatial deductive reasoning and psychology of geometric reasoning.	Philosophy
Vogelzang	Cognitive capacities may constrain the acquisition of linguistic forms and their meanings in various ways.	Sentence comprehension especially pronoun comprehension.	Linguistics; AI
Martin	All of the same signature properties that accompany the outcome-to-intent shift in moral judgement in childhood characterize the "intent-to-outcome" shift obtained under cognitive load in adults.	Literature on moral judgement and its development.	None
Tasimi	Identified the mechanisms that contribute to the belief that money carries traces of its moral history.	Work in cognitive psychology on money being thought to carry traces of its moral history, especially developmental research.	None
Berent	People view psychiatric disorders as more likely to be innate and immutable when diagnosis is supported by a brain test as compared to a behavioral test. So people spontaneously essentialize psychiatric conditions that are linked to the brain, even when the brain probe offers no additional diagnostic or genetic information. This bias suggests that people consider the biological essence of living things as materially embodied.	Empirical research on two principles of intuitive psychology. Dualism and Essentialism.	None
Li	The results provide novel insights into how polysemy emerges from contextual processing of word meaning from both a theoretical and computational point of view.	Psycholinguistic work on people's ratings of semantic similarity and of context effects on reading times.	Linguistics; AI
Thorn	Human subjects are sensitive to the difference between fitted and unfitted classes and are more credulous concerning inheritance inference based on fitted classes.	Cognitive-psychological studies of concept formation.	Philosophy; AI
Duan	The efficiency of an action bounds the generalization of the action across social group members through a process that is spontaneous and implicit. This constrained action generalization may be due to inefficient actions being represented as culture-specific conventional forms.	Work on effects of group membership on categorization of individuals (stereotyping).	None
Lisson	Assuming an activation-based model, intermittent deficiencies may be the best explanation for the cause of sentence processing impairments in aphasia, although slowed syntax and lexical delayed access may also play a role.	Review of behavioural research on sentence processing impairments in aphasia. Data comparing aphasics and controls in a self-paced sentence comprehension task.	Linguistics
Kirkland	Standard textbook word problems may be able to be rewritten in ways that mitigate a "senseless" mindset.	Research on neglecting world knowledge in problem solving.	None
Kurtz	Human learners often extend the alternation pattern – a finding of critical interest given that leading theories of	Research on human category learning.	None

	categorization based on similarity or dimensional rules fail to predict this behavior.		
Perry	Iconicity not only can serve to help children identify the referent of novel words but can also support their ability to retain even noniconic word-referent mappings.	Children's word learning.	None
Annerer-Walcher	What eye parameters are best suited as indicators of internal versus external attentional focus in different settings.	Eye parameters as indices of internal vs external focus of attention.	AI
Rottman	Tree-huggers have relatively amplified tendencies to attribute mental capacities to animals and relatively reduced tendencies to attribute mental capacities to out-group members – thus having elevated rates of both anthropomorphism and dehumanization. These findings necessitate a reconceptualization of both the extension of moral worth and the attribution of minds.	Moral attitudes in childhood. Moral cognition.	None
Cassani	Words with more coherent diachronic usage patterns tend to be acquired earlier. The results support theories positing a link between ontogenetic and ethnogenetic processes in language.	Research on factors influencing AoA of words.	AI; Linguistics
Gong	Cognitive reflection, rational thinking, and normative thinking dispositions converge even in a culture that emphasizes holistic nonanalytic thinking.	CRT: research on analytic vs intuitive thought.	Anthropology
Monaghan	Iconicity is revealed as a feature of ultra-conserved words and potentially also as a property of vocabulary early in the history of language origins.	Experiments investigating production of iconicity in communicative studies. Subjects' ratings of word iconicity.	Linguistics
Aryawibawa	Balinese conceptualizes directness of causation in a similar way to speakers of languages unrelated to Balinese.	Work in which adult native speakers rated causative events in terms of four properties. Research with Balinese making grammaticality judgements.	Linguistics; AI
ISSUE 5 – 13 papers			
Saide	There is a connection between children's animistic and anthropomorphic reasoning for unobservable entities, and an indirect effect of cultural input on such reasoning.	Research on young children's application of heuristics concerning animacy and anthropomorphism. Studies of how children and adults conceptualize God and germs in terms of general psychological and physiological traits.	None
Napoli	Re sign language: once iconicity is considered, handshape and movement parameters interact at the sublexical level. Thus, consideration of iconicity makes transparent similarities in grammar across both modalities, allowing one to maintain certain key findings of phonological theory as evidence of cognitive architecture.	Phonological aspects of signing.	Linguistics
Baggio	Offers a dual-stream model of language-processing.	Evidence from animal and human research shows that memories of events may be encoded in two "formats". Research on redundant representations of items in working memory tasks. Research on idiom processing.	Cognitive neuroscience; Linguistics; AI
Beekhuizen	There is a factor influencing latencies in a speech production task: linguistic complexity, which translates as increasing processing demands when a word is merely retrieved from long-term memory when in addition it undergoes morphosyntactic computation, and in case on top of the two preceding operations, it requires phonological computation.	Experimental work on various aspects of lexical processing (lexical decision, semantic priming, sentence processing).	Linguistics
Akita	Phonation types make a considerable contribution to Japanese speakers' sound-symbolic ratings of size and shape.	Research on sound symbolism.	Linguistics
Thornton	The work indicates a unity of composition in thought and language, and identifies extensional superposition as the underlying mechanism. Relationships between linguistic and conceptual composition are identified.	Research on concepts.	Linguistics
Ludusan	Speaker variability can impair certain learning algorithms (supervised algorithms) but help other ones (unsupervised algorithms). This makes the prediction that as infants develop and become more able to exploit top-down information, speaker variability should have a progressively facilitatory effect.	Evidence on the impact of variability on robust phonetic category learning.	AI
Guo	On some measures that primarily relied on metalinguistic cues, gender concepts resembled dual-character concepts.	Research on factors affecting category membership judgement.	None

	However, on other measures that depicted transgressions of traditional gender norms, neither “man” nor “woman” appeared dual-character-like.		
Dubova	The human visual system adapts its encodings to the feature structure of the environment, uses categorical expectations for robust reconstruction, allocates encoding resources with respect to categorization utility, and adapts to prevent miscategorizations.	Research on perceptual learning and perceptual categorization.	None
Thorne	Epistemic appraisals of concepts form a psychologically important yet previously overlooked aspect of the structure of concepts.	Took inspiration from a study by Haslam, Rothschild, and Ernst (2000) on the different ways a social concept could be based on an “essence”.	Philosophy
Devitt	There are both descriptive and causal-historical elements to the reference determination of some natural kind terms. We should abandon the common assumption that any one theory of reference fits all natural kind terms.	Research on subjects judging natural kind membership under conditions of ambiguity.	Philosophy
Loos	Re sign language: what matters in echo phonology is the visual percept of temporally coordinated movement that repeats a salient movement property in such a way as to give the visual impression of a copy. Echoes are not obligatory motor couplings of two or more articulators but may enhance phonological distinctions that are otherwise difficult to see.	Research on head movements that accompany signing.	Linguistics
Speed	There are associations between the sound of a word and its odor valence.	Research indicating that emotion is a critical dimension on which we encode and understand odor.	Linguistics
ISSUE 6 – 11 papers			
Meylan	There is no cross-linguistic evidence of a stronger correlation between in-context predictability and word length than word frequency and word length.	Empirical research on Zipf’s Law as applied to human language use.	Linguistics
Benitez-Burraco	Language evolution is a multifactorial process, with each player acting upon the other, engaging in an intense mutually reinforcing feedback loop. Language evolution as a gradual process, continuous with the pre-linguistic cognitive abilities, which are engaged in a positive feedback loop with linguistic innovations, and where gene-culture co-evolution and cultural niche construction were the main driving forces.	Influences of cultural niche construction on cognition. Influence of one’s language on one’s working memory. Relationship between reactive aggression and aspects of social cognition. Impact of self-domestication on cognition.	Anthropology; Linguistics
Van Schijndel	A full explanation of syntactic disambiguation difficulty may require recovery mechanisms beyond predictability	Garden-path effects on sentence comprehension.	Linguistics; AI
Messenger	Neither immediate nor long-term priming effects differ between children and adults but both children and adults show significant immediate and persistent effects of the priming when the test phase occurred immediately after the exposure phase and when a short delay separated the exposure and test phase.	Syntactic priming effects with human speakers.	Linguistics
Levin	Eye movement studies of cognition during dynamic naturalistic stimuli show that the tyranny of dynamic stimuli is real, and that links between eye movements and cognition are highly dependent on task and stimulus properties.	Research on top-down cognitive processes influencing on eye movements.	None
Martin	There is a causal link between perceptions of “internal wickedness” and moral judgment.	Effects on social cognition of priming of essentialist thinking.	None
De Jonge-Hoekstra	This work extends the phenomenon of gesture-speech mismatches to difficult tasks in general. Aksoi shows how temporal alignment, semantic similarity, and complexity matching are related, and how they predict participants’ task performance.	Work on gesture-speech synchronization in adults and children during e.g. problem-solving.	None
Barrett	While dualist thinking is a possible mode of thought enabled by evolved human psychology, such thinking does not constitute a default mode of thought. Rather, our data support what we will call <i>intuitive materialism</i> – the view that the underlying intuitive systems for reasoning about minds and death produce, as a default, the judgment that mental states cease to exist with bodily death.	Empirical research on kinds of afterlife belief in children and adults. Intuitive physics and intuitive psychology.	Anthropology
Stanley	Results have implications for competing theoretical views on the relationship between memory and morality and for functional theories of counterfactual thinking.	People’s memories of their past moral transgressions. Episodic counterfactual thinking.	None
De Cesarei	If the visual characteristics of the environment are not learned by CNNs, their visual categorization may depend only on a subset of the visual information on which humans	Research on categorization of visually presented stimuli by humans.	AI

	rely, for example, on low spatial frequency information.		
Yu	The magnitude of the OAE differs by accent, indicating that not all “other” accents are processed in the same way. The OAE can disrupt talker recognition as severely as the LFE. While confidence is related to accuracy within a given accent, listeners’ confidence ratings do not faithfully reflect accent-based differences in talker recognition performance.	Research on influences of talker accent on talker recognition.	None
ISSUE 7 – 9 papers			
Willett	Experience-based learning over long timeframes exhibits similar strengths and weaknesses as in short timeframes. However, learning over long timeframes may become more impaired with more complex tasks.	Empirical and theoretical studies of causal learning.	None
Zhang	Children’s word learning is a continuous process by means of real-time information integration.	Studies of word learning in young children.	None
Thomas	Once people’s behavior on a set of problems has been used to establish individual-level cognitive models, the cognitive models can predict how those people would have behaved for other problems.	Research on “wisdom of the crowd” using sequential decision-making tasks	None
Abramov	Children with higher cognitive abilities convey significantly more meaning via gesture and less solely via speech. These findings suggest that young children’s use of cospeech representational gesturing is positively related to their mental representation and reasoning abilities.	Research on how speech and gesture work together	None
Cranford	We developed an instance-based learning (IBL) model of an attacker using the Adaptive Control of Thought-Rational (ACT-R) cognitive architecture to investigate how humans make decisions under deception in cyber-attack scenarios.	Research on how deception succeeds through the exploitation of human processing constraints and perceptual, cognitive, and social biases.	AI
Siew	Global feature distinctiveness is associated with earlier age of acquisition ratings, and is a stronger predictor of age of acquisition than local feature distinctiveness. These results suggest that the global structure of the semantic feature network could play an important role in language acquisition, whereby globally distinctive concepts help to structure vocabulary development over the lifespan.	Experimental studies of language development.	None
Pouw	Gestures can come to embody the linguistic system at the level of interrelationships between communicative tokens, which should calibrate our theories about form and linguistic content.	Experiments in which agents are tasked to learn a novel set of signals. These signals are iteratively transmitted to later generations and/or used in communication by later generations (iterated learning + communication), where, over many cycles of learning and use, they are affected by various transmission biases	Linguistics; AI
Capuano	Linguistic negation acts very differently from a logical operator, proving itself to be in fact highly restrictive in the suggestion of alternative scenarios: naturally produced alternatives are very similar to the negated entity.	Research on processing difficulties of negation in sentence comprehension tasks	Linguistics
Martínez-Huertas	Words with early maturation and subsequent semantic definition promote emotional propagation. Second, an interaction effect between age of acquisition and abstractness explains model performance.	Research on children’s acquisition of abstract words	AI
ISSUE 8 – 11 papers			
Gumbsch	Event-predictive learning combined with active inference may be critical for eliciting goal-anticipatory gaze behavior in infants.	Eye-tracking studies of infants watching goal-oriented actions.	AI
Myers	Finding of metrical regularity inhibiting speech production is inconsistent with an abstract metrical structure that is planned independently of phonemes at the point of phonological encoding.	Research on metrical priming in speech comprehension and production tasks.	None
Caddick	Participants were tasked with repeatedly choosing policies. Evidence of motivated reasoning is found despite financial incentives for accuracy. This research advances the field of causal learning by studying the role of prior preferences, and in doing so, integrates the fields of causal learning and motivated reasoning.	Research on motivated reasoning and on causal learning.	None
Paape	Using a self-paced reading study investigating number attraction effects in Eastern Armenian. our data are better accounted for by an encoding-based model of agreement attraction, compared to a retrieval-based model.	Empirical work on the effect of agreement attraction in sentence acceptability and sentence production tasks.	AI

Dotlacil	Develops a novel psycholinguistic parser and tests it against experimental and corpus reading data.	Research on memory in parsing focusing on the role of memory during comprehension. Modelling data from self-paced reading studies.	Linguistics; AI
Wood	Slowly changing visual experiences play a critical role in the development of human-face recognition by chicks, akin to basic-level object recognition. Thus, face recognition is not a hardwired property of vision but is learned rapidly as the visual system adapts to the temporal structure of the animal's visual environment.	Studies exploring the origins of face recognition in newborn humans. Research on effects of experience on monkeys' face recognition abilities.	Cognitive neuroscience
Hannakainen	There is cross-cultural convergence in people's understanding of the nature of law. There is a striking degree of universality in beliefs about the essential properties of law, despite abundant historical and regional variation in the way actual laws manifest.	Work on subjects reporting their moral reactions – for example, of perceived shame and wrongness – in response to hypothetical violations of outdated laws. When asked to simultaneously consider actual and possible laws, participants reported that laws often abide by the procedural principles, though it would be possible for laws to violate them.	Anthropology; Philosophy
Drenea	Both foveal and extrafoveal processing are simultaneously involved during a categorical search, and the specificity of their interaction is determined by the spatial orientation of objects, type of distractors, the prohibition to use overt attention, and individual characteristics of the participants.	Treisman feature-integration theory. Work on object semantics being processed extrafoveally and then used to guide saccadic programming in a top-down manner. Research on covert attention and on categorical search.	None
Carcassi	Monotone quantifiers are easier to learn, and are also widespread in language <i>because</i> of their learnability.	Evidence that humans find monotone quantifiers easier to learn.	Linguistics; AI
St. Pierre	To help infer the meanings of novel words, children frequently capitalize on their current linguistic knowledge to constrain the hypothesis space.	Studies of children learning the meanings of function words.	None
Richie	Modelled human judgments of similarity among cohyponyms of various categories, using various similarity metrics, which are all related to classic psychological models of similarity like common features/additive clustering, distinctive features, and multidimensional scaling.	Research on effects of similarity on various cognitive tasks and on tasks where participants are given a cue word and asked to respond with whatever words come to mind.	AI
ISSUE 9 – 10 papers			
Donnelly	The lexicons of 30-month-olds exhibit lexical-level competition, with competition increasing with vocabulary size.	Neighbourhood size effects in spoken word recognition in infants and adults.	None
Western	The name-based model of the representation of categories has superior explanatory potential with regard to concept acquisition. Illustrates the complexity of the relationship between language and the world, the acknowledgment of which is paramount to enhancing our understanding of both natural language and human cognition.	Research with experiments that collect human judgments of category relatedness and category membership	Linguistics; AI
Abudarham	Our findings validate human perceptual models of face recognition, enable us to use DCNNs to test predictions about human face and object recognition as well as contribute to the interpretability of DCNNs. The approach we used here for the study of face recognition can be similarity applied to other cognitive tasks to improve our understanding of human cognition and the interpretability of artificial neural networks.	Human psychophysical experiments revealing a subset of view-invariant facial features that are critical for human face recognition	AI
Fay	Our evolved cognitive preferences are moderated by our social goals.	Experimental studies showing that compared to positive information, negative information has a greater impact on impression formation and is more believable. Transmission chain studies <i>a la</i> Bartlett	None
Valtonen	People distort claims about the brain from the wider culture to fit their dualist belief that minds and brains are distinct, interacting entities: Exposure to cultural discourse about the brain as the physical basis for the mind prompts people to posit that mind-brain interactions are asymmetric, such that the brain is able to affect the mind more than vice versa.	Research on lay intuitive theories about the mind and the brain, and how these develop in children.	Philosophy
Karjus	People tend to colexify similar meanings with artificial languages. Speakers' communicative needs affect their lexification choices.	Research with pairs of participants faced with the task of communicating single-word messages using a small set of artificial words.	Linguistics
Goldwater	Spontaneous use of causal principles in analogy generati-	Research on the influence of expertise on	Philosophy

	on is particularly characteristic of experts.	encoding processes during problem-solving, and on analogical retrieval and generation.	
Shin	For Korean-speaking children's comprehension of a transitive construction, the Agent-First strategy is activated properly only in conjunction with other types of interpretive cues.	Studies of early acquisition of word order.	Linguistics
Kryven	People attribute intelligence based on plans given sufficient context and cognitive resources and rely on the outcome when computational resources or context are limited.	Research on factors governing humans' attributions of intelligence.	None
Lowder	Results provide important information regarding the locus of relative-clause processing effects and highlight the importance of carefully considering how intervening material might unintentionally alter the structure or the meaning of a sentence.	Psycholinguistic research on comprehension of sentences containing relative clauses.	Linguistics

ISSUE 10 – 17 papers

Holtgraves	Documented how the interpersonal context influences the communication process and how language users are sometimes unaware of their communication failures.	Work on miscommunication e.g. sender-receiver asymmetry in communicative confidence.	Philosophy; Linguistics
Kneer	Ordinary people are willing to ascribe deceptive intentions to artificial agents and to blame them for this.	Empirical work on the Folk's concepts of lying.	Philosophy; AI
Haward	The distinction between properties that bear principled connections to their kinds, and merely generic-supporting properties structures novel kind representations, provides strong evidence that this distinction is part of the formal structure of kind representations.	Work showing that distinction between generic-supporting and idiosyncratic/accidentally related properties emerges early in ontogenesis, reflected in the child's language and thought by at least two and a half years of age.	Philosophy
Entzmann	There is no automatic prioritization of emotional faces, at least for saccades with short latencies, but salient local face features can automatically attract attention.	Work on guidance of attention by face stimuli.	Cognitive neuroscience
Flanagan	Children and adults hold a nuanced understanding of free choice that is sensitive to both the agent type and constraints within a given scenario.	Research showing that from early in life, children understand choice as the ability to consider alternative means for achieving a goal, and to flexibly switch between these alternatives when faced with new constraints.	None
Canessa	Blind individuals show many of the same processing differences between abstract and concrete concepts found in sighted individuals, but our model shows that those differences are noticeably less pronounced than in sighted individuals.	Empirical studies of concepts using the Property List paradigm.	None
Nancekivell	Findings reveal the complexity of learning style beliefs, and how they interact with evidence in previously undocumented ways.	Research on efficacy of multimodal learning.	None
Braithwaite	Results have implications procedural knowledge, conceptual knowledge, and metacognitive processes in math problem solving.	Research on use of conceptual knowledge during problem solving.	None
Foppolo	A verb's perfective aspect triggers the culmination inference incrementally during sentence processing. This offers novel evidence for the continuous integration of linguistic processing with real-world visual information.	Research on effects of verb type on sentence comprehension.	Linguistics
Kumar	Both associative and distributional models can capture relatively unconstrained search processes in a cooperative game setting, and Connector I is particularly suited to examine communication and semantic search processes.	Research on retrieval from semantic memory.	None
Scheer	These findings may thus redound to the advantage of psycholinguistic studies by identifying two new factors, as well as produce results that speak to the linguistic quarrels.	Cognitive models of language production and evidence related to them.	Linguistics
Payir	Children appraise the likelihood of story events actually happening in light of their beliefs about causal regularities. A religious upbringing does not impact the frequency with which children invoke causal regularities in judging what can happen, even if it does impact the type of causal factor that children endorse.	Developmental research on children's grasp of the distinction between reality and pretence.	None
Richter	The model points toward a neural dynamic theory of higher cognition. This demonstrates how elements of higher cognition, including the capacity to sequentially operate on objects, to form and reject hypotheses, and to apply local networks to global representations, may be	Discusses research on how siblings establish joint attention via conversation.	AI

	grounded in sensory-motor processing.		
Banks	Results support the linguistic short-cut hypothesis in semantic processing and provide strong evidence that both linguistic and grounded representations are inherent to the functioning of the conceptual system.	Research with the category exemplar production task.	Linguistics; AI
Verosky	Generalized scale degree associations are informed by both adjacent and nonadjacent relationships between melodic notes influence listeners' melodic predictions above and beyond n-gram context, highlighting the need to consider a broader range of statistical learning processes that may underlie listeners' expectations for upcoming musical events.	Data from trained human singers performing a musical cloze task, to assess musical knowledge.	AI
Leontyev	The response time (RT) and motion-based decision tasks converge well at a fundamental level, and mouse-tracking features such as area under the curve and maximum velocity indicate the degree of decision conflict and impulsivity.	Research indicating that mouse trajectory features such as AUC (area under the curve) reveal the subject's perceptual, cognitive, and social conflicts in the decision-making process.	None
Kim	A recent study proposed a computational model that combined a model of trial-to-trial changes with a model for the internal scaling of discrete numbers. Here, we tested a surprising prediction of the model – a situation in which children's estimates of numerosity would be better than those of adults.	Developmental studies of numerical estimation.	AI
ISSUE 11 – 9 papers			
Hoicka	High-proportion quantifiers, including generics, lead adults, and to some extent children, to generalize, but not essentialize, about novel social kinds.	Work showing that generics help adults and children learn novel information about known and novel kinds.	None
Jing	These findings challenge received theories of across-the-board effects of complexity on word order and word order variation and call for theoretical models that relativize effects to specific kinds of syntactic structures and dependencies.	Effects of sentence complexity on self-paced reading.	Linguistics
Skovgaard-Olsen	Both stances – neutrality about the antecedent in indicatives and the falsity of the antecedent in subjunctives – are conveyed by conversational implicatures.	Experimental data indicating that participants prefer indirect to direct meanings when committing problematic acts when the hearer is likely to be antagonistic and when the potential costs are high. Data indicating that participants prefer to trust speakers who implied (more technically, “implicated”), rather than explicitly said or presupposed, false information.	Linguistics
Kempe	Detection of systematic form-meaning linkages requires stable combinatorial units that can guide learners toward the structural properties of signals.	Research using Bartlett iterative communication technique.	Linguistics
Favela	Person-plus-tool systems can be classified as extended cognitive systems. A framework for quantifying system-level properties of these systems is proposed.	Results of experiments analyzing the movement dynamics of participants engaged with tools during a task centered on affordance judgments.	Philosophy
Valentini	Bilingual children exposed to English earlier have larger vocabularies and better morpho-syntactic skills.	Evidence that older children between the ages of 5 and 7 and adults too can use syntactic information to deduce the meanings of unfamiliar verbs. This provides further support for the role of syntactic bootstrapping past the earlier stages of lexical acquisition.	None
Lev-Ari	Widely spoken languages rely on different sound symbolic patterns than less common languages. Community size can thus shape linguistic forms and influence the tools that languages use to facilitate communication.	Empirical research of the effects of sound symbolism on word acquisition in children.	Linguistics
Sehl	Children do not anticipate the sunk cost bias in first person scenarios, or in interpersonal sunk cost scenarios, where costs are sunk by others. Young children may struggle to understand and predict irrational behavior.	Developmental research demonstrating that young children come to expect others to behave rationally in light of their beliefs and desires, and environmental constraint.	None
Boduch-Grabka	A cognitive factor, processing fluency, can lead individuals to distrust information when it is delivered in a foreign accent. Ensuring exposure to foreign accent can reduce discrimination against non-native speakers.	Empirical research on effects of processing fluency on information evaluation.	None

ISSUE 12 – 8 papers

Fitch	There are similarities and differences in the information that learners may have access to in acquiring signed versus spoken languages.	Language acquisition research on extensively on how do infants take arbitrary word forms from the input and map each onto the correct meaning and on a variety of cues that children employ, including joint attention and iconicity.	None
White	Happening is perceived at a moment and is constituted by an information structure of connected recent historical information.	Evidence that perceptual asynchronies can be corrected or recalibrated, though Errors of synchronization do occur.	None
Slooman	“Exploration” may not be a single cognitive property, but rather the emergent result of three distinct behavioral and cognitive mechanisms, namely, (a) breadth of generalization, (b) quality of prior expectation, and (c) relative valuation of self-obtained information.	Cognitive phenomena involved in social learning.	None
Berent	Mind–body interactions elicit a latent dualist dissonance in intuitive psychology.	Research on infants’ perceptual expectations. Theory of mind research.	None
Tourtour	A bounded-rational account of overspecification is proposed according to which even redundant words can be beneficial to comprehension to the extent that they facilitate the reduction of listeners’ uncertainty regarding the target referent.	Research showing that (a) In situated contexts the visual input rapidly informs listeners’ incremental language processing; (b) Contra the Gricean view – which deems only minimal specifications to be optimal – speakers do frequently overspecify.	Philosophy; Cognitive neuroscience
Biesaga	People high in the Need For Cognition can maintain cognitive motivation for a longer period and consequently on average generate more random series.	Research indicating that people cannot produce perfectly random series.	None
Dewi	These results and conclusions have important implications for learning theories because they demonstrate that a shift from counting to retrieval over training cannot be deduced from verification time differences between outside and within-count equations in an alphabet-arithmetic task.	Research showing that Some knowledge can be learnt directly and deliberately by the memorization of associations, while other knowledge can be created after repeated practice of counting procedures. In the latter case, a shift from procedural to retrieval strategies necessarily occurs during learning.	None
Loy	Speakers align with their partner’s linguistic behavior to produce overspecific or minimally specific descriptions. Little evidence that the alignment was enhanced by lexical or semantic repetition across prime and target trials.	Research on overspecification posing a problem for theories of language production that emphasize communicative efficiency.	None

RICERCHE

Two open questions in the reformist agenda of the philosophy of cognitive science

Aurora Alegiani,^(α) Massimo Marraffa^(α) & Tiziana Vistarini^(β)

Ricevuto: 1 maggio 2022; accettato: 25 aprile 2023

Abstract In this paper we carve out a *reformist* agenda within the debate on the foundations of cognitive science, incorporating some important ideas from the 4E cognition literature into the computational-representational framework. We are deeply sympathetic to this reformist program since we think that, despite strong criticism of the concept of computation and the related notion of representation, computational models should still be at the core of the study of mind. At the same time, we recognize the need for a liberalization of the computational and representational framework that can address deep dissatisfaction with the anti-biologism and radical internalism of classical cognitive science. However, reform is a difficult task, so in this article we focus on two open questions within the reformist agenda. The first concerns the possibility of combining mechanistic-computational and dynamical explanations. The second concerns related changes in the notion of representation and its use (with special attention to Andy Clark's radical predictive processing).

KEYWORDS: Continuum of Representational Genera; Enactivism; Predictive Processing; Radical Embodied Cognition Thesis; Representationalism

Riassunto *Due problemi aperti nell'agenda riformista della filosofia della scienza cognitiva* – In questo lavoro identifichiamo un'agenda *riformista* nel dibattito sui fondamenti della scienza cognitiva che incorpora alcune idee centrali provenienti dalla letteratura sulla cognizione 4E all'interno di una cornice computazionalista e rappresentazionalista. Tale agenda considera il quadro computazionalista e rappresentazionalista ancora imprescindibile ai fini dello studio integrato della mente e del cervello, ma ne persegue una liberalizzazione nell'intento di renderlo idoneo ad accogliere alcuni importanti spunti emersi dalla letteratura sulla cognizione delle 4E. Tuttavia, riformare è un compito difficile. In questo articolo ci concentriamo su due problemi aperti nell'agenda riformista. Il primo riguarda la possibilità di mettere assieme le spiegazioni meccaniciste e computazionaliste con quelle dinamiche. Il secondo riguarda i cambiamenti relativi alla nozione di rappresentazione e al suo impiego (con particolare attenzione all'elaborazione predittiva radicale di Andy Clark).

PAROLE CHIAVE: Continuum dei generi rappresentazionali; Elaborazione predittiva; Enattivismo; Tesi della cognizione incarnata radicale; Rappresentazionalismo

^(α)Dipartimento di Filosofia, Comunicazione e Spettacolo, Università di Roma Tre, via Ostiense, 234/236 - 00146 Roma (IT)

^(β)Dipartimento di Scienze Filosofiche, Pedagogiche ed Economico-Quantitative, Università degli Studi di Chieti-Pescara "G. D'Annunzio", via dei Vestini, 31 - 66100 Chieti (IT)

E-Mail: aurora.alegiani@uniroma3.it; massimo.marraffa@uniroma3.it (✉); tvista3@gmail.com



1 The radical embodied cognition thesis

BETWEEN THE 1950S AND 70S, cognitive science took on the intellectual and institutional structure that we today define as “classical”. This occurred, however, at the cost of narrowing the disciplinary scope that characterized its gestation period. In fact, while this early phase was characterized by the dense interweaving of contributions from various fields of research (studies on artificial neural networks, symbolic artificial intelligence, Chomskyan linguistics, neuroscience, and psychology), in classical cognitive science, symbolic artificial intelligence established hegemony. Nevertheless, starting from the mid-eighties, cognitive science began to regain – and indeed expand – its original scope, extending in two directions.¹ On the one hand, it expanded *vertically*, toward the brain, leading cognitive neuroscience to assume the centrality that artificial intelligence had previously held. As part of this downward expansion, the emergence of subsymbolic connectionism was hailed by many as a paradigm shift.²

On the other hand, cognitive science has also expanded *horizontally*, toward the environment. This expansion resulted in a critical review of both the methodological individualism of classical cognitive science (it is not possible to study the mind by putting the physical and social environment in which it operates in brackets), and the metaphysical thesis according to which the mind depends only on the brain (mental processes are in the first instance control systems of a *body* that moves, acts and, by moving, retroacts on the brain and mind). Accordingly, a number of highly diverse research programs have adopted an *externalist* conception of explanation that is organically linked to a conception of cognition as embodied, embedded, enacted and extended – so-called “4E cognition”.³ This conception can be more or less radical.

A first form of externalism consists in the methodological thesis that an adequate understanding of cognitive processes requires study of the environment in which these processes take place.⁴ This thesis is fully compatible with evolutionary computational psychology, which assumes that cognitive processes have been “designed” (by evolution or learning) for specific physical or social environments.⁵

A second form of externalism is the so-called “sensorimotor paradigm”, a family of theories that share a critical stance towards computational theories of perception and classical computationalism in general. This paradigm also informs sensorimotor *enactivism*, a research program that can be characterized as today’s updated version of the ecological perspective introduced by James J. Gibson. It takes the form of *radical* enactivism when it follows Gibson’s ecologism not only in considering perception to be *constituted by* (not only de-

pendent on) the possession of sensorimotor competence,⁶ but also in subscribing to his critique of the concept of mental representation.⁷

This radicalization is certainly not mandatory. Sensorimotor enactivism can accord, within an ecological theoretical framework, some role to computational models, more or less consistently revised to make them congruent with the ecological point of view. But there have also been attempts to synthesize computationalism and ecologism in a different way, i.e., by trying to account, within the computational paradigm, for some theses from the Gibsonian school.

The idea of an integration of the computationalist and ecological theoretical frameworks is only of course possible if Gibson’s ban on the notions of representation and computation is overturned. The first attempt in this direction was made by Ulrich Neisser, whose notion of *anticipatory schema* is much closer to the concept of representation in cognitive science than to the Gibsonian idea of direct extraction of invariants.⁸

On this track, Pierre Jacob and Marc Jeannerod’s dual theory of vision tried to account for some aspects of ecologism within the computational paradigm.⁹ The authors believe that a clear distinction should be made between vision-for-action and vision-for-perception, both characterized in terms of computations and representations. Vision-for-perception is realized in the ventral pathway and is conceived, in broad terms, according to Marr’s framework. Vision-for-action is realized in the dorsal pathway; and it is here that some of Gibson’s theses are computationally reinterpreted. For example, the viewing of affordances is reconceptualized as a modular computational process in which a visuomotor representation is constructed to function as a kind of anticipatory schema.

In this way, the computationalist can capitalize on an important point emphasized by Gibson’s ecological psychology. The perceptual process can be understood as both that which allows us to control motor behavior and that which allows us to recognize objects; to neglect either aspect is to forego offering a complete account of perceptual activity. In particular, while it is true that we have historically favored the use of “perception” to denote the process that culminates in the formation of an empirical belief, it should be clear by now that this traditional linguistic usage inappropriately neglects the fact that perception is also a system of action control.

However, as already mentioned, sensorimotor enactivism claims to go far beyond the conclusion that perception is *also* an action control system, advancing the thesis that visual perception is a form of action that does not require construction of mental representations. This is the *radical* strand of enactivism; and when joined with that form of externalism that denies or tends to deny

the existence of a *clear boundary between agent and environment*, it corresponds to what Clark has called “the radical embodied cognition thesis”.¹⁰

The radical embodied cognition thesis is the outcome of a Kuhnian interpretation of the dynamical approach to cognition (sometimes referred to simply as “dynamicism”). The application of tools of dynamical systems theory to psychological phenomena has been presented as the advent of “a third contender” in the debate on the foundations of cognitive science.¹¹ In this connection, a standard reference is van Gelder and Port’s *Mind as Motion*, a collective volume that was the first major presentation of the dynamical approach to cognition. According to the editors, «to see that there is a dynamical approach is to see a new way of conceptually reorganizing cognitive science as it is currently practiced». ¹² Such reorganization takes a stand against not only classical but also connectionist computationalism – and this despite the fact that the connectionists were the first to apply dynamical systems theory to the study of cognition.¹³ However, van Gelder and Port argue, the limit of connectionism lies in the use of dynamical systems tools within a paradigm that is still computationalist and representationalist, though in a brain-like variant. This is not enough for the dynamicists.

First, the dynamicist dissolves the boundary between the cognitive system and the system’s environment. Coupling between the equations describing a cognizing system and those describing the environment gives rise to complex “total system” behaviors.¹⁴ In this perspective, «the cognitive system is not just the encapsulated brain; rather, since the nervous system, body, and environment are all constantly changing and simultaneously influencing each other, the true cognitive system is a single unified system embracing all three». ¹⁵

Second, the dynamicist cuts ties with mechanistic and computationalist explanations. The dynamicist expansion into the environment implies a very different explanatory model from the mechanistic one underlying vertical expansion. In the 1950s the early cognitivists’ appeal to mechanistic explanatory strategy was the logical conclusion of the battle waged against behaviorism and mathematical psychology, which conceived of psychological explanation as the discovery of laws or mathematical regularities in behavior.¹⁶ The dynamical approach, however, relaunches the covering law conception of explanation.¹⁷ The dynamical analysis identifies the critical variables characterizing the state of a system and attempts to construct laws (a set of differential equations) to account for the system’s trajectory through state space. The system can no longer be decomposed into subsystems (modules) that involve computations on representations. Consequently, the dynamical explanation is seen as incompatible with the explanatory style of the computationalist mechanism.¹⁸

So we have arrived at the radical embodied cognition thesis: to understand the complex interplay of brain, body, and environment we do not need either the concepts of internal representation and computation or the mechanistic decomposition of a cognitive system into a multiplicity of inner neuronal or functional subsystems; all we need are the analytic tools and methods of dynamical systems theory.¹⁹ We think, however, that in this form the dynamicist project is not a third contender in the controversy on the foundations of cognitive science but, rather, the denial of the possibility of such a science. In other words, it is not “radical embodied cognitive science” but, rather, the confirmation of the current relevance of some *behavioristic* metaphysical and methodological challenges.²⁰

Fortunately, this obituary for cognitive science has been opposed by a “reformist” perspective, according to which the computational and representational paradigm can be reconstructed making due allowances for «the environmental embedded, corporeally embodied, and neurally “embrained” character of natural cognition»,²¹ without collapsing into the radical embodied cognition thesis.

2 A continuum of representational genera: An early continuist take

In pursuit of this reformist program, Andy Clark grafted the computational and representational framework onto a three-tiered explanatory strategy: (a) a *dynamical* explanation of the gross behavior of the agent-environment system; (b) a *mechanistic* account, describing how the components of the agent-environment system interact to produce the collective properties described in (a); and (c) a *representational* and *computational* explanation of the components identified in (b).²²

This tripartite strategy (*minimal representationalism*) is then situated within the larger framework of *active externalism*. Unlike *semantic externalism*, where the mental contents of a subject depend on aspects of the environment which are clearly external to the subject’s cognitive processes, active externalism asserts that the environment can play an active role in constituting and driving cognitive processes.²³

In the wake of Gibson’s ecologism, this environment is conceived by Clark as a complex of affordances, which are however the source of a particular variety of internal states, namely, *action-oriented representations*. Unlike the symbols of classical computationalism (prototypically, language-of-thought symbols), action-oriented representations are *personal* (in that they are related to the agent’s needs and the skills that it has), *local* (in that they relate to the circumstances currently surrounding the agent) and *computationally cheap* (compared with David Marr’s rich inner models of

the visual scene).²⁴ Thus, Clark affirms, when the agent represents the environment as a complex of affordances, the type of internal states that are generated are representations that describe partial aspects of the world and prescribe possible actions and interventions.²⁵

However, action-oriented representations are only one *genus* of representations. The concept of inner representations was introduced in cognitive science to account for cases in which a cognitive system must coordinate its behaviors with environmental features that are not always present and manifest to the system.²⁶ In such cases the cognitive system is able to decouple from the external environment and act in an offline fashion by creating some kind of inner item that stands in for the absent phenomena. These *inner stand-ins* are what cognitive scientists have termed “inner representations”. Such cases of environmentally decoupled cognition are really a tough nut for the anti-representationalists to crack, since they are exclusively concerned with cases of “adaptive hookup”, i.e., cases in which the inner states of a system must coordinate behaviors with specific environmental contingencies.²⁷ Such cases, however, cannot ground a *general* anti-representationalist argument since they are not sufficiently “representation hungry”.²⁸

In view of this, Clark replaces the classical notion of mental representation with *a continuum of representational genera*. At one end of the spectrum are inner states that border on simple causal correlation and environmental control. At the other end of the spectrum, we find the type of inner stand-in that allows us to deal with the representation-hungry problems. Between these two poles are the action-oriented representations.

According to Clark, therefore, depending on the coupling or decoupling between agent and environment, one must respectively appeal to the dynamical non-representational or the representational explanation. It can be objected, however, that this implies a division of labor between the two styles of explanation, and not their complementarity; as a result, they cannot be tiers (i) and (iii) of a single explanatory strategy, as Clark would like.

This ambiguity echoes in Clark’s most recent proposal, which appeals to predictive processing to refine the reformist agenda. We return to this point in Section 4.

3 Integrating mechanistic and dynamical explanations

Clark’s active externalism leaves another thorny question open: what role is there for the “highest” dynamical level, if we have already a mechanistic explanation of the interactions between the components of the global system? We

understand that one can say that a certain type of system does not lend itself to mechanistic description, and so a dynamical model must be used; but we do not understand why, as Clark seems to suggest, we should (1) describe the global agent-environment system as a dynamical system and (2) describe the interactions between the components of the system with a mechanistic model: if the interactions between the components are describable in mechanistic terms, the reason for thinking of the global system in dynamical terms falls away.

The philosopher who has worked most on the prospects for the integration of dynamical modeling with mechanistic analysis has been William Bechtel.

In the ground-breaking *Discovering Complexity*, Bechtel and Richardson note that in the early stage of the process of developing mechanistic models scientists often assume that the processes they are considering are performed serially.²⁹ But when it is not possible for scientists to develop a linear model that is adequate for a phenomenon, they start to introduce feedback loops and other non-linearities in their attempts to develop adequate models. The outcome is what the authors define as “functionally integrated systems”.

Again, as in the case of Clark’s representationalism, a continuum emerges. At one end of the spectrum, we have *fully decomposable* (or *highly modular*) systems, which are composed of subsystems that are completely independent except for the mutual exchange of outputs.³⁰ If the interactions among the subsystems are weak but not negligible, the system is *nearly decomposable*. As the complexities of the interactions among parts increase, the explanatory burden shifts from the parts (or, more precisely, the interactions *within* subsystems) to their organization (i.e., the interactions *between* subsystems). Thus, we reach the other end of the spectrum, where we find *holistic* systems, whose components are functionally equivalent and hence interchangeable. In between the nearly decomposable systems and the holistic ones, there are the integrated systems. In these systems, unlike the holistic systems, it is possible to isolate different parts that make distinctive contributions but also give rise to a complex set of interactions that are nonlinear, and hence much stronger than those of a nearly decomposable system.

Now, Bechtel believes that psychobiological cognition occupies the intermediate space between near decomposability and holism, namely that of integrated systems.³¹ This allows him to denounce as spurious the opposition between an ultra-modularist conception of the parts of biological mechanisms as totally isolated and a radically holistic view that rejects the very possibility of decomposing the mind-brain.³²

Indeed, Bechtel considers the collection of studies by David van Essen and his collaborators – which provide an almost complete map of the are-

as of the Macaque monkey visual system – to be an exemplar of a *mechanistic* analysis of how the brain performs a cognitive function.³³ The researchers have identified over 30 different areas in the macaque visual cortex and more than 300 connections between these areas; and the tool-kit of dynamical analysis can be very useful for modeling this vast number of feedforward, feedback, and collateral connections. However, although these regions are highly interconnected, we can still determine what each area contributes to visual information processing. That is, it is not a holistic system, but an integrated one; and in an integrated system, mechanistic analysis «provides the foundation for dynamical analysis»³⁴ since the latter has explanatory force only insofar as it describes «the operations of the underlying mechanism»,³⁵ only to the extent that it reveals «aspects of the causal structure of a mechanism».³⁶ In this vein, Bechtel and Abrahamsen coined the phrase “mechanistic-dynamic explanations” for those explanations that integrate the mechanistic decomposition of systems into parts and operations with the quantitative tools offered by dynamical systems theory.³⁷

Mechanistic-dynamic explanations, however, require a reconceptualization of the notions of part and operation:

Dynamic mechanistic explanations are still mechanistic, and so make reference to operations localized within parts, but respect the dynamic processes that require characterizing both parts and operations relationally in terms of how they are situated in endogenously active dynamic networks.³⁸

In this perspective, then, individual brain regions are “endogenously” active and as a result of this activity organize themselves into specialized processing components. And yet, even as brain regions specialize, they remain integrated with other regions according to a form of organization known as “the small-world network”. In such an organization, local clustering gives rise to specialized regions, but long-range connections link processing in these regions with activity taking place in other parts of the system, allowing activity in these other parts to modulate the behavior of local clusters.

Do Bechtel’s dynamic mechanistic explanations succeed in harmonizing mechanistic-computational with dynamical explanations as Clark’s three-tiered explanatory strategy requires? It is unclear. The appeal to dynamical models is invoked for integrated systems which are very *weakly* modular, since each of their parts is influenced by the activity in some other parts of the system. Is this degree of modularity sufficient for the standard required by a mechanistic model including computational explanations? Certainly not for those cognitive scientists who argue that a mechanistic-computational explanation

requires constraints on the concept of part far more demanding than what is required for the notion of integrated system, namely, some form of informational encapsulation and massive modularity.

An example is provided by Peter Carruthers who, in dealing with Fodor’s problem of the computational intractability of holistic central cognition, distinguishes a “narrow-scope” form of encapsulation from a “wide-scope” variety.³⁹ Influenced by the “simple heuristics” research program, he argues that such computational tractability does not require Fodor’s “narrow-scope” form of encapsulation in which the encapsulated system cannot draw on any information held outside of it during the course of processing but only that, on any given occasion, it draws only on a subset of the “exogenous” information – a property that Carruthers calls “frugality” or “wide-scope encapsulation”.

Now, within this modularist framework, Carruthers suggests an account of central cognition based on a functional architecture known as the “Global Workspace” (GW). Initially proposed by Bernard Baars based on hypotheses made by Tim Shallice and Michael Posner, this architecture was later developed by Stanislas Dehaene (2014) as the “Global Neural Workspace” (GNW) hypothesis.⁴⁰

Analyses of functional connectivity patterns in the human brain, Carruthers notes, have demonstrated just the sort of neural architecture necessary to realize the main elements of a GW model.⁴¹ More specifically, these studies show the existence of two main neurocomputational spaces within the brain, each characterized by a distinct pattern of connectivity. The first space is a processing network, composed of a set of parallel, distributed, and functionally specialized processors or *modular* subsystems subsumed by topologically distinct cortical domains with highly specific local or medium-range connections that encapsulate information relevant to its function. The subsystems compete with each other to access the GNW, which is implemented by long-range cortico-cortical connections, mostly originating from the pyramidal cells of layers 2 and 3 that are particularly dense in prefrontal, parieto-temporal and cingulate associative cortices, together with their thalamo-cortical loops. This GNW breaks the modularity of the nervous system. When one of these subsystems accesses the GNW, its outputs are broadcast to an array of specialized executive, conceptual, and affective consumer systems – e.g., systems that “consume” the perceptual input to form judgments or make decisions. Moreover, GNW makes possible the development and subsequent benefits of a working memory system which exploits the mechanisms of global broadcast to subservise a wide variety of central-cognitive purposes.⁴²

Thus, a GNW architecture arranges parallel specialized conceptual systems around the global broadcast of attended perceptual information and

thus enforces competitive entry into a general-purpose working memory system. Such design features seem to enable us to circumvent many aspects of Fodor's scepticism about holistic central cognition being amenable to computational modeling.⁴³

Hohwy proposed an account of conscious access that integrates the GNW into the framework of the Predictive Processing framework, a view that Whyte terms "the predictive global neuronal workspace".⁴⁴ This reference makes it particularly interesting for the purposes of our reflection on reformism in philosophy of cognitive science Andy Clark's attempt to develop the explanatory framework of active externalism by adding to the resources of connectionist and dynamical psychology those of the Predictive Processing (henceforth, PP).

4 Radical predictive processing and 4E cognition

PP is a renowned and highly articulated framework that has emerged as one of the prominent alternatives for a reformist program. PP, it appears, is capable of pursuing many notable objectives: above all, that of conjugating the "pragmatic turn" taking place in the cognitive sciences with the more "conservative" desire to preserve a representational outlook on cognition.

Some ambiguities in the project still remain, though. These involve the employment of the concept of representation. Indeed, it seems, and understandably so, that when it comes to representations, the middle ground can easily lead to fuzziness. A further look at this unwelcome consequence may not only point to open questions in the reformist agenda but also to strategies we could employ in its future development.

In this respect, Clark's recent integrative proposal is a leading example. If it is true that the notion of representation has undergone several changes, his approach pushes these modifications to their limit. The development of Clark's methodology reflects the rationale of the three-tiered scheme presented in Section 3. The "continuist" approach adopted there is now cast within the popular framework of PP. Before looking at Clark's stance, we highlight a few key points of PP.

4.1 A nod at predictive processing (PP)

In PP, organisms proceed by trial and error in an uncertain world. They have access to the sensory consequences of environmental causes exclusively and, based on the former, must reconstruct the latter. To do so, in light of the models they acquire through experience (generative models), systems proceed inferentially to predict the causes of the effects they receive. Upon failure, error signals ensue. The minimization of these signals is the

fundamental goal of cognition, for all its processes («perception and action and everything mental in between») can be traced back to the overarching principle of prediction error minimization.

Error minimization calls for a complex balancing game between "model intervention" and "world intervention". When an error ensues, a system must decide whether to correct the model upon which the predictions were made or to change the world to conform to the model. This decision is not straightforward and requires further harmonizing between the reliability of the models on the one hand and the errors on the other. The tradeoff between such reliabilities is modulated by precision.

A crucial aspect of this theory lies in its hierarchical nature. While there are different approaches to how they should be algorithmically sorted and implemented, the general idea is that a predictive architecture is made of different levels, each comprising different units: predictions and prediction errors. At any level predictions are sent downstream in the hierarchy while prediction errors are sent upstream; each level is in turn constrained by predictions coming from higher and errors coming from lower levels. The hierarchical matrix of PP is mirrored in the generative models it proposes. The highly structured data they contain (the interactions between hidden causes and their effects) require, in fact, a multilevel kind of organization that tracks different interactions at different spatial and temporal scales. In this way, the causal-probabilistic relationships standing between causes and expected outcomes trace the environmental variables as closely as possible.

4.2 Clark's radical synthesis

Based on these premises, Clark has developed what he himself dubbed "*radical* predictive processing" (rPP), opposing it to the contrasting "conservative" view (cPP).⁴⁵

His reasoning can be summarized schematically: PP is compatible with 4E approaches; PP is (in some shape or form) representational;⁴⁶ hence, PP affords a strategy to integrate the two perspectives on cognition. Importantly, PP is not just another version of the "traditional" computational-representational view of cognition (as "conservative" proponents, seemingly, claim), but is a paradigm that enables a virtuous synthesis between apparently opposing tendencies. In this sense, Clark contends, PP allows us to "predict peace" in the long-standing "representational wars" raging in the cognitive sciences. But before such peace can be stipulated, it seems that a few dangling threads must be tied together.

Clark provides a captivating option for the reformist agenda. However, noting a few matters can provide an indication as to how a reformist proposal must evolve to be more robust. Clark's alternative is convincing from a general perspec-

tive. By this, we mean that his reformist proposal compellingly points to the compatibility of PP and 4E, leveraging key points that are consistent with the “pragmatic turn” of PP. This aspect of the integrative proposal, though, is also supported by proponents of cPP:⁴⁷ the fact that the brain’s objective is not *plainly* “reconstructive”⁴⁸ is part and parcel of PP’s program. If both parties are in general agreement as to the compatibility of PP and 4E, then the crux of the dispute lies in singling out exactly what *features* of such representations significantly differ from those of the “conservative” camp. This should further enable rPP to distinguish the types of processes and levels at play in the synthetic framework.⁴⁹ Given that cPP already undertakes a shift in the understanding of representational activities, a further refinement of this shift puts the “radical” reformist in dire straits, as shall become clear below.

Clark’s two-faced account reveals a risk that looms large on reformist approaches when they try their hand at changing the features of the representations they postulate while still resorting to a computational-representational gloss. This problem acquires the form of a question: *how far can we take these modifications?*

It seems, in fact, that while the stricter claims coming from the conservative camp provide more rigorous criteria for representations,⁵⁰ Clark’s deflationary⁵¹ proposal has some unfinished business. We emphasize that this article does not intend to unravel these complex knots. Rather, it intends to point out, in light of Clark’s proposal for rPP, the risks lurking in the reformist agenda.

4.3 The background and development of the radical synthesis

As anticipated, various aspects of the proposed compatibility between PP and 4E are endorsed by both rPP and cPP.

In Clark’s work, this compatibility has immediate repercussions on the descriptions of the representations at play in cognition. As illustrated in the previous Sections, it is important to note that his view harkens back to both embodied⁵² and enactive projects,⁵³ as well as to the suggestions coming from behavior-based robotics⁵⁴ and Beer’s dynamical approach.⁵⁵ Upon this reading, in outlining cognitive processes, we should not consider the slavish reconstruction of the world to be their prerogative. On the contrary, cognitive science must be in the business of sketching «the common principles or lawful linkages between sensory and motor systems that explain how action can be perceptually-guided in a perceiver-dependent world».⁵⁶ In this sense, cognition is modulated by “scaffolding” and “productively lazy” processes. According to the latter, cognitive systems operate strategically to reduce the load of their enterprises;

in line with the former, systems exploit the environment and their bodies to minimize cognitive costs. One interesting instance is that of infants in category learning and concept formation processes: here, «self-generated motor activity, such work suggests, acts as a “complement to neural information-processing”».⁵⁷ In sum, by poking, sucking, and grasping, infants balance out the cognitive load of “information structuring” and “information processing” across their brains, bodies, and the environment.

Clark makes the case that PP is compatible with these standards. Two of the major points he leverages fall under the balancing game mentioned above (sect. 4.1). For one, in model selection, we opt *for less complex* alternatives. As famously claimed by Clark: «the goodness of a predictive model is determined by accuracy minus complexity» (this is pointed out by Hohwy himself).⁵⁸ To this end, our model choices will tend to be less costly in computational terms. Second, precision calls for frugality. Its *context-sensitive* precision assignments – based on which we select our cognitive strategies – both consist of the alteration of «patterns of “effective neuronal connectivity”» and rest on an impending situation on which we must get a quick grip.

In sum, it is not coincidental that the two terms “fast” and “frugal” stick together. Keeping up with the requirements of the environment requires us to be “fast”. But being fast is only possible through frugality.

Another example related to “scaffolding” illustrates the link between “fast” and the “frugal”. The case is Optical Acceleration Cancellation (OAC). When running to catch a fly ball, we do not resort to detached representations to determine where it will land to *thereafter* run to reach the expected spot. Rather, as long as we move so as to cancel any «apparent changes in the ball’s optical acceleration»,⁵⁹ we will reach the point where the ball strikes the ground.

Instead of relying exclusively on independent structures to successfully interact with it, we harness the world itself to achieve this very same goal. In the case above, “throwing away the world” would amount to doing all the cognitive work on representations before acting. To live up to this function, such representations would be complex and articulated. In short, the lack of representational frugality comes with slowness.

On cPP, Clark notes, «the model-rich cognizer is able to “throw away the world” and select her actions and responses by manipulating the inner model instead».⁶⁰ Just like Clark, Hohwy promotes his stance based on the very motives of PP. It is the uncertainty we navigate in our experiences that calls for rich and (p)reconstructive models. In this view, PP subscribes to a more canonical understanding of the computationalist-representationalist perspective.

Two points carry weight, though. The appeal cPP proponents make to the reconstructive aspects of representations should not have us thinking that they argue that the pragmatic traits of cognition are a drop in the bucket. On the contrary, they oftentimes emphasize that the rejection of the idea of «representation for representation's sake»⁶¹ is not at odds with the preservation of «internal representation *as such*». In short, claiming that representations “copy” does not equate to saying that the reconstruction they impart does not bolster “coping”.⁶² Hence, cPP proponents do not negate the relevance of action or coping. Rather, they claim that pursuing such cognitive goals does not require dispense with the reconstructive and detachable traits of representations altogether.

Another conviction we should let go of as far as cPP is concerned is that the “more canonical” representations we have been mentioning thus far are the traditional, “*linguaformal*” structures of Fodoran inspiration.⁶³ PP’s representationalism, in fact, comes with a whole different wealth of representations, one that exhibits substantial differences from the “traditional” understanding of the concept. This is even more evident when we pause on the fact that *generative models* are the standard of mental representation on PP. Much more should be specified here, for the implications of this transition are immense. But we will limit ourselves to note that the “recapitulative” function performed by these kinds of representations is starkly different from the linguistic alternative. Here, the neocortex does not explicitly represent the parameters indicating the causal-probabilistic relations between the represented environmental variables; rather, just like a graph or a map, it *acquires* the form, the structure, of the reciprocal relationships between these variables.⁶⁴ As Daniel Williams has noted, the implication is that «[i]n doing so, [these relationships] effectively realize a dynamical model (albeit a causal-probabilistic one) of the body and environment» and, further, «brains deploy the very kind of representation that advocates of dynamical systems theory argue *we* should use to model the brain». ⁶⁵ Importantly, this seems to imply compatibility between the deployment of representational structures and dynamical approaches to cognition.

This conclusion gives hope to the reformist program. If cPP already provides substance to this integration, what does rPP add to the picture?

4.4 rPP representations

Clark’s talk of representations gains shape while he is in the affair of distancing them from the rich, reconstructive, and detached models of cPP. This is a thorny issue, as it seems that the key features of Clark’s representational paradigm must be taken up by those very traits that are intended to distinguish

them from what we generally consider representational. This is further complicated by the fact that both these “paradigmatic” structures and the processes that tap into them combine with other structures and operations, thus making the term difficult to outline and understand straightforwardly.⁶⁶ Some points can be drawn, though.

First, to be genuinely “pragmatic” and pursue the control-theoretic goal cognition performs, representations must be fast, frugal (that is, “quick-and-dirty”), and “attached” (in the sense of “non-detached”), as well as observer-dependent.⁶⁷ To be a structure of this kind is to be an action-oriented representation.

Second, such structures are intermingled with other “styles” of processing, that go from higher-order, rich modeling, to reflex-like processes. The relations between these kinds of processing, as well as those between the structures handled (or not handled) in each case, lie on a *continuum* and combine with one another based on the system’s needs (these observations recall section 2). In this, Clark explicitly draws on computational neuroscience, such as the correlation between “model-free” and “model-based” styles of processing.⁶⁸ In general, these two types of processing are considered two extremes of the cognitive spectrum. Model-free processing resembles the non-representational end, while the model-based end traces higher-level cognition. Both these extremes and what is in between combine in varied and dynamic ways.

Third, action-oriented representations, to exert their effects, lean on rich models. It is such models that make the default patterns of action-oriented representation possible, Clark claims when he observes that «slower processes of learning and adaptive plasticity have already sculpted patterns of neural connectivity in ways that make the low-cost strategy available».⁶⁹

To recapitulate then: The “pragmatic turn” pervades the representational agenda. Cognition is *not* in the business of reconstructing rich, slow, and abstract models. While the concept of representation is preserved, unlike in cPP, the notion is (mainly) action-oriented. Representations are frugal, fast, and observer-dependent. However, they are not all there is to cognition, which is made up of different styles of processing that variously combine, spanning from higher, abstract levels down to heuristic, reflex-like processing. The possibility of engaging in action-oriented representations itself “rests” on high-level knowledge.⁷⁰

4.5 A glimpse into the issues

The picture indeed exhibits ambiguities. Clark’s insistence that representations be fast and frugal leads to surprise when he alludes to the presence of rich models. This sensation is as pronounced as ever when he claims that action-

oriented representations (which are central in rPP, if only because of Clark's insistence) are *made available* by rich models (thus granting the latter a key role after having largely downsized their benefits). Further, it is not very clear why action-oriented representations should still be deemed as such: what are the genuinely representational traits they impart? Last, it is not uncontroversial that these different styles of processing can all be maintained in PP.⁷¹

We shall group these uncertainties into three main issues: first, the representational status of action-oriented representations; second, the possibility of intermingling between model-based and model-free approaches; third, the role of high-level models.

All of this aims at highlighting possible flaws of the reformist program. We will not by any means resolve these issues but leverage them to provide food for thought for the future development of the reformist agenda.

Clark's description poses issues as to the representational status of the structures he proposes. As mentioned above, his favored strategy in describing action-oriented representations is to distinguish them from PP's commonplace understanding of the notion (that of cPP). The general impression is that he does not systematically explain in what capacity they still are, for all intents and purposes, representations. If, on the one hand, Clark tells us why we should shy away from the canonical understanding of the notion in PP, he does not rigorously indicate why we should still consider action-oriented structures to have representational status.

This sounds even more dissonant if we think that cPP does indeed propose structures that are both representational and, ultimately, action-guiding (see above). It would be misleading to suggest that the representations at play in PP are action neutral. In cPP, the process of construction of a structural representation of the environment «is not an end in itself, but a tool for self-maintenance».⁷² Further, «what is "reconstructed" in internal models of prediction-error-minimizing-agents are those aspects of the environment which constitute the organism's *Umwelt*, i.e., the ones which the organism depends on in its practical engagements with the environment».⁷³ Suffice it to say that it is not sufficient for a state to be "structurally similar" to its domain for it to qualify as representational. To acquire this status, the relation of structural similarity must be *exploitable*.⁷⁴

In light of these general observations, it might seem that the hybrid structures put forward by Clark add complication. cPP poses entities that, in reconstructing the environment, do not aim at mirroring it, but rather at acting in it, in this way fully embracing the pragmatic turn and the 4E approach. Concurrently, cPP allows us to endorse this option while

at the same time providing a structured way of motivating the representational status of the structures at play.⁷⁵ It seems that, as long as there is an option that allows us to get "the best of both worlds" by aptly justifying both endorsements (that is, the computational-representational stance on the one hand and the pragmatic role of cognition on the other), it makes less sense to endorse a two-faced notion⁷⁶ such as action-oriented representations. This consideration is bolstered by the concurrent need, on Clark's end, for the postulation of a non-representational end on the one hand and of a "canonically" representational one on the other, as well as for the "fully" representational structures to be the condition for availability of the action-oriented representations.

Related to this haziness about the representational weight that action-oriented representations may bear, is the fact that they often appear to be compatible with the notion of "affordance". Even more so, the representations postulated by Clark aptly correlate with Shepard's⁷⁷ development of Gibson's theory. This reference of Barrett is on point: Shepard, in fact, proposes a model where selection and learning provide the basis for apt 'resonance' to the environment. Thus, we are «tuned to resonate» to patterns that are meaningful to us. Such a "resonating system" imparts higher levels that resonate to a «wider class of objects and events». Very interestingly, systems are calibrated to reduce the complexity of these resonances. Higher levels are invoked when things get complicated. If not, systems deploy simpler resonance, which, at the lowest level, is "direct perception" in a Gibsonian sense⁷⁸. The similarity between this program and Clark is evident. Thus, it seems Clark could set up a very similar project even without the notion of representation.

This issue, as anticipated, cannot be solved here. But it points to a major risk looming on the reformist agenda. *If, in adopting a reformist approach, we change the notion of representation to fit 4E, then we must be careful to provide a more robust account of the terms by which it should still genuinely be "representational"*. The absence of this clarification, in fact, might make the postulated structures compatible with approaches that do not ratify representationalism, thus making its endorsement appear trivial.

This leads us to the second point. Clark admits to the possibility of coexistence and cooperation between representational and nonrepresentational processes. This point is further argued for by casting the findings of computational neuroscience⁷⁹ into the PP framework. Model-based and model-free cognitive styles coopt one another and dynamically combine based on our current needs. This point puts pressure on the endorsement of the representational jargon, though. PP, in fact, promotes an overarching view in which cognition is constructed in a fractal-like manner. If lower levels of cognition operate in a

model-free, non-representational way, then how can higher levels operate so differently? This claim necessitates for instance the adoption of “double standards”, or of a “double gloss”: as Anderson and Chemero⁸⁰ have argued, for example, the notion of prediction at low levels is used differently from that at high levels. They frame this operation as an arbitrary one: «it is a theoretical choice not necessitated by evidence».⁸¹ The conclusion they draw is, in short, that since PP can be framed in a non-representational gloss,⁸² we could settle things once and for all by excluding the notion of representation altogether. Thus, the presence itself of model-free processing puts pressure on the persistence of the notion of representation⁸³ from *two perspectives* (from both the representational and the non-representational readings of PP). On the one hand, assuming that PP is representational, PP’s hierarchical, overarching view is at odds with this stark distinction, as generative models are supposed to be at work everywhere throughout cognition; on the other, the fact that we encounter non-representational processes, conjoined with PP’s overarching framework, might lead us to think that the notion of representation is ultimately uninformative.

This second point emphasizes the risks of a hybrid architecture in promoting the reformist agenda. Not only are the required double standards at odds with the general rationale of PP (leading to objections from proponents of both camps), but the difference between model-based and model-free processing seems to be too radical to grant, as Clark would wish, the possibility of seeing these processing styles as the two ends of a continuum.

A similar argument has been made by Hohwy with regard to the relationship that is supposed to stand between action-oriented representations and higher levels of the generative model. Recall, the latter make low-cost strategies *available*. This is a strong statement, for it argues that the action-oriented representations’ operations are made possible by rich representations. Not only is this a strong statement, but a confusing one too, since Clark devotes much ink to rescaling the role of the latter. This point is well argued by Hohwy: «there is [...] a potential tension [...] between allowing and withholding a role for rich models».⁸⁴

This tension is especially highlighted by the following considerations. Clark “withholds” a role for rich models in two main ways. Firstly, he extends the non-reconstructive prerogative to include higher levels of elaboration. The latter do not aim at *depicting* the world, but at *ensuring an adequate grip on it*. The pragmatic rationale goes all the way up to high-level cognition. Secondly, the use we make of the latter is avoided as much as possible.

However, not only does he concurrently preserve the jargon of richness (for example, he claims that higher levels are “more *abstract*”, thus

implying some form of detachment),⁸⁵ but he also ascribes an important role to them. For instance, he argues:

[...] high-level states (of the generative model) target large-scale, increasingly invariant patterns in space and time. Such states help us to keep track of specific individuals, properties, and events despite large moment-by-moment variations in the stream of sensory stimulation.⁸⁶

This seems to point to slow and definitely richer, as well as detached levels of elaboration. This is where Clark “allows” a role for rich models.

This ambiguity provides food for thought for reformists: we who are sympathetic to this camp should be careful not to superimpose the *features* of the representations on the nature of their *use*. This oscillation is present in rPP: when Clark argues that higher levels of the generative model exert an “active” role what he means, most probably, is that in their *use* they are fundamentally pragmatic. In short, when we put them into use, we do not deploy them *as rich reproductive models as such*, but we use them to highlight what is meaningful to us in that context. In principle, however, Clark does agree that we have an «immense storage of causal knowledge».⁸⁷ Thus, while the two stances seem to converge on the general need for high-level information in our models, on the one hand (cPP) this information is used to «repeatedly infer when we are in situations where low-cost strategies are viable»; in this sense invoking all the apparatus of knowledge we enjoy, while on the other hand (rPP), once we have acquired such information (once we have learned), we do not repeatedly put to use the entire wealth of knowledge we have gained over time. Rather, this acquired knowledge allows us to operate in a default manner whenever possible. More precisely, we exert «default precision-assignments that install the transient organizational structure that best confronts that kind of puzzle in that kind of context».⁸⁸

The take-home message goes as follows: since a reformist approach embraces disparate concepts and structures in their differences, to conjugate these, it must *be rigorous about the features and the use of the structures it postulates*.

Further reflection can be conducted as to the possibility of preserving all this richness only to shut it away in tangible experience. This point highlights that the amphibious position of action-oriented representations still meets with inconsistencies. It would seem that we stockpile a great deal of information without actually invoking it in experience.⁸⁹

Recapitulating. Clark’s “radical reformism” still holds the notion of representation, but it emphasizes its pragmatic nature. The paradigmatic case of a representation is action-oriented representations. Concurrently, Clark retains rich models,

claiming both that they make quick-and-dirty operations available and, concurrently, that their objective is pragmatic. Further, these different structures interact with one another, as well as with model-free operations. All these processes lie along a *continuum*.

The open questions are thus condensed: in what measure are action-oriented representations still such? Why should we retain the notion of representations if there are non-representational processes (given the “unifying” rationale of PP)? Or, complementarily, why shouldn’t we retain the notion of representations if there are representational processes? More generally: why resort to a hybrid notion if we do have a well-grounded notion of representation (that of cPP) that does not neglect the role of action, but rather emphasizes its prominence in our cognitive processes?

5 Conclusions

In this paper we have carved out a reformist agenda within the debate on the foundations of cognitive science, a reformist agenda that aims to incorporate some important ideas from the literature on 4E cognition in the computational-representational framework.

We are deeply sympathetic to this reformist program since we think that, despite some strong criticisms of the concept of computation and the related notion of representation, computational models are still at the core of the disciplines of the mind. Computational models and, more generally, mechanistic explanations are still the dominant methods in cognitive science. Indeed, on the one hand, the complexity of animal and specifically human behavior requires an appropriately complex model, like computational models provide; on the other hand, more traditional nomological explanations do not appear apt to hit the explanatory target: psychological explanation is closer to the biological than the physical kind.⁹⁰

At the same time, we recognize the need for a liberalization of the computational and representational framework that offers a satisfactory response to the deep dissatisfaction with the antibiologism and radical internalism of classical cognitive science. In this perspective, the evaluation of the tenability of the reformist program is the main task of the philosophy of cognitive science. Our article focused on two paradigmatic cases of reformism. First, we conducted some reflections on Bechtel’s attempt to combine mechanistic-computational explanations and dynamicist explanations. Then, we critically examined Clark’s strategy for integrating representationalism and 4E cognition through radical predictive processing.

As we have seen, many open questions remain -- reforming is undoubtedly an arduous task. It is therefore not surprising that, as early as 1998,

Bechtel, Graham and Abrahamsen contemplated the possibility that «the simultaneous pulls *downwards* into the brain and *outwards* into the world may prove to be too much pulling, and lead to the disintegration of cognitive science».⁹¹ However, the authors hastened to point out that the volume they introduced (a companion to cognitive science) was an attempt to combine the two thrusts. Reformism is the belief that a coherent synthesis can be found between vertical and horizontal expansion.

Notes

¹ W. BECHTEL, G. GRAHAM, A. ABRAHAMSEN, *The life of cognitive science*, p. 77.

² W. SCHNEIDER, *Connectionism: Is it a paradigm shift for psychology?*

³ Cf., e.g., A. NEWMAN, L. DE BRUIN, S. GALLAGHER (eds.), *The Oxford handbook of 4E cognition*.

⁴ Cf. E. HUTCHINS, *Cognition in the wild*.

⁵ Cf. E. MACHERY, *Philosophy of psychology*, pp. 273-274.

⁶ That is, the tacit knowledge of the patterns of covariation between our bodily movements and the sensory stimulus.

⁷ Cf. L. SHAPIRO, S. SPAULDING, *Embodied cognition*, §2.4.

⁸ U. NEISSER, *Cognition and reality*.

⁹ P. JACOB, M. JEANNEROD, *Ways of seeing*. The authors build on Ungerleider and Mishkin’s famous what vs. where distinction, which was then modified by Goodale and Milner and became the action vs. perception distinction. Another attempt to make a synthesis between environmentalism and computationalism is in J. NORMAN, *Two visual systems and two theories of perception*.

¹⁰ A. CLARK, *Being there*, p. 108.

¹¹ C. ELIASMITH, *The third contender*.

¹² R. PORT, T.J. VAN GELDER (eds.), *Mind as motion*, p. 4.

¹³ Neural networks are a class of dynamic systems: the evolution of a network can be described by a system of differential equations.

¹⁴ Note that the existence of a direct systematic dependence between movements and stimulation and the related intrinsically dynamic nature of perception are the main reasons that radical enactivists identify complex dynamical systems as the most appropriate models for the sensorimotor system. Such dependence is in fact an exemplification of what dynamicists call “coupling relations”.

¹⁵ T.J. VAN GELDER, *What might cognition be, if not computation?*, p. 373.

¹⁶ W. BECHTEL, GRAHAM, ABRAHAMSEN, *The life of cognitive science*, p. 96.

¹⁷ Cf. W. BECHTEL, *Representations and cognitive explanations*; J. WALMSLEY, *Explanation in dynamical cognitive science*; A. CHEMERO, *Radical Embodied Cognitive Science*.

¹⁸ Cf., e.g., A. CHEMERO, M. SILBERSTEIN, *After the philosophy of mind*. The authors offer the Haken-Kelso-Bunz model of bimanual coordination as evidence that some explanations in cognitive science and neuroscience are nonmechanistic and that for many complex behavioral and neural systems the primary explanatory tools are the mathematical methods of nonlinear dynamical systems theory.

¹⁹ Cf., e.g., A. CHEMERO, *Radical embodied cognitive science*; G. SANCHES DE OLIVEIRA, V. RAJA, A. CHEMERO,

Radical embodied cognitive science and “real cognition”.

²⁰ Cf., e.g., N. ALKSNIS, J. REYNOLDS, *Revaluing the behaviorist ghost in enactivism and embodied cognition.*

²¹ T.J. VAN GELDER, *Dynamic approaches to cognition*, p. 244.

²² A. CLARK, *Being there*, p. 126.

²³ M. ROWLANDS, J. LAU, M. DEUTSCH, *Externalism about the mind.*

²⁴ Cf. D. BALLARD, *Animate vision.*

²⁵ The reference here is to R. MILLIKAN, *Pushmi-pullyu representations.*

²⁶ Cf. J. HAUGELAND, *Having thought*, p. 172.

²⁷ A. CLARK, *Being there*, p. 147.

²⁸ A. CLARK, J. TORIBIO, *Doing without representing?*, p. 403. Shaun Gallagher has recently called this the “scaling-up problem,” i.e., the problem of making current approaches to «lower-order or basic types of processes involving perception and action» also work for explaining «higher-order cognitive abilities, such as memory, imagination, reflective judgment, and so on» (S. GALLAGHER, *Enactivist interventions*, p. 187).

²⁹ W. BECHTEL, R. RICHARDSON, *Discovering complexity*, Ch. 7.

³⁰ This would be the case with Jerry Fodor’s encapsulated modules.

³¹ W. BECHTEL, *The compatibility of complex systems and reduction.* Cf. also A. CLARK, *Being there*, p. 114.

³² This radically holistic conception is endorsed by G.C. VAN ORDEN, B.F. PENNINGTON, G.O. STONE, *What do double dissociations prove?*

³³ W. BECHTEL, *Mental mechanisms*, Ch. 3.

³⁴ W. BECHTEL, *The compatibility of complex systems and reduction*, p. 483.

³⁵ D.M. KAPLAN, W. BECHTEL, *Dynamical models*, p. 443.

³⁶ D.M. KAPLAN, C.F. CRAVER, *The explanatory force of dynamical and mathematical models in neuroscience*, p. 602. Milkowski and colleagues (*From wide cognition to mechanisms*, p. 2393) subscribe to Kaplan and Craver’s thesis, situating it within a broader theoretical framework in which “broad” perspectives on cognition (i.e., those of the 4E approach) are reinterpreted as research heuristics to construct mechanistic explanations.

³⁷ W. BECHTEL, A. ABRAHAMSEN, *Dynamic mechanistic explanation*; W. BECHTEL, *Addressing the vitalist’s challenge to mechanistic science.*

³⁸ W. BECHTEL, *Referring to localized cognitive operations in parts of dynamically active brains*, p. 263.

³⁹ P. CARRUTHERS, *The architecture of the mind*, p. 58.

⁴⁰ S. DEHAENE, *Consciousness and the brain.* For a recent review of GW-Inspired Architectures, cf. A. SIGNA, A. CHELLA, M. GENTILE, *Cognitive robots and the conscious mind.*

⁴¹ Cf., e.g., E. BULLMORE, O. SPORNS, *Complex brain networks*; G. GONG, Y. HE, L. CONCHA, C. LEBEL, D.W. GROSS, A.C. EVANS, C. BEALIEU, *Mapping anatomical connectivity patterns of human cerebral cortex*; M. SHANAHAN, *Embodiment and the inner life.*

⁴² Cf. P. CARRUTHERS, *The centered mind.*

⁴³ Cf. E.C. DEISE, *Frame problems, Fodor’s challenge, and practical reason*; P. CARRUTHERS, *On central cognition.*

⁴⁴ J. HOHWY, *The predictive mind*; C. WHYTE, *Integrating the global neuronal workspace into the framework of predictive processing: towards a working hypothesis.*

⁴⁵ Cf. J. HOHWY, *The predictive mind*; P. GŁADZIEJEWSKI, *Predictive coding and representationalism*; W. WIESE, *What are the contents of representations in pre-*

dictive processing?; D. WILLIAMS, *Predictive processing and the representation wars.*

⁴⁶ We note that, despite their importance to the debate, we here set aside anti-representationalist takes on PP since they would countenance no reformism whatsoever – at least not in the terms we hope for here. Cf., e.g., J. BRUINEBERG, J. KIVERSTEIN, E. RIETVELD, *The anticipating brain is not a scientist.*

⁴⁷ Cf., e.g., P. GŁADZIEJEWSKI, *Just how conservative is conservative predictive processing?*; J. HOHWY, *Quick’n’lean or slow and rich?*

⁴⁸ Cf. M. ANDERSON, *After phrenology.*

⁴⁹ This is noted in K. DOŁĘGA, *Moderate predictive processing.*

⁵⁰ Cf. references in footnote 55.

⁵¹ We call his approach “deflationary” as it seems to downplay the degree of (at least the canonical) “representationality” of representations as will appear below.

⁵² Cf. A. CLARK, *Radical predictive processing*; M.K. MCBEATH, D.M. SHAFFER, M.K. KAISER, *How baseball outfielders determine where to run to catch fly balls.*

⁵³ F.J. VARELA, E. THOMPSON, E. ROSCH, *The embodied mind.*

⁵⁴ Cf. R.A. BROOKS, A.M. FLYNN, *Fast, cheap and out of control.*

⁵⁵ Cf., e.g., R. BEER, *Dynamical approaches to cognitive science*; R. BEER, *The dynamics of active categorical perception in an evolved model agent.*

⁵⁶ F.J. VARELA, E. THOMPSON, E. ROSCH, *The embodied mind*, p. 173.

⁵⁷ M. LUNGARELLA, O. SPORNS, *Information self-structuring: Key principles for learning and development*, p. 25 as in A. CLARK, *Radical predictive processing*, p. 12.

⁵⁸ However, conservative views, while leaning towards rich and thorough representations, do not correlate these necessarily with complexity: Hohwy diffusely refers to models as both “simple” and “low-cost, energy-efficient”.

⁵⁹ A. CLARK, *Radical predictive processing*, p. 11.

⁶⁰ *Ibid.*, p. 15.

⁶¹ M. WILSON, *Six views of embodied cognition*, as in D. WILLIAMS, *Predictive processing and the representation wars*, p. 148.

⁶² These points are well said by D. WILLIAMS, *Predictive processing and the representation wars*, pp. 148-149.

⁶³ Cf. J.A. FODOR, *The language of thought.*

⁶⁴ Cf. D. WILLIAMS, *Predictive processing and the representation wars*, p. 154; GŁADZIEJEWSKI, *Predictive coding and representationalism.*

⁶⁵ D. WILLIAMS, *Predictive processing and the representation wars.*

⁶⁶ Note that we do not mean to say that there are no issues in cPP or that there is nothing controversial with its concept of representation. Aspects of these criticalities have been reviewed, (e.g., DOŁĘGA, *Moderate Predictive Processing*), not to mention the issues uncovered by the non-representational camp (e.g., BRUINEBERG ET AL., *The anticipating brain is not a scientist*).

⁶⁷ See references to Clark’s work in footnote 45.

⁶⁸ N.D. DAW, S.J. GERSHMAN, B. SEYMOUR, P. DAYAN, R.J. DOLAN, *Model-based influences on humans’ choices and striatal prediction errors.*

⁶⁹ A. CLARK, *Radical predictive processing*, pp. 16-17.

⁷⁰ Clark does not refrain from using this expression: A. CLARK, *Whatever next?*, p. 200.

⁷¹ The cogency of these ambiguities is highlighted by

several contributions from M. COLOMBO, E. IRVINE, M. STAPLETON (eds.), *Andy Clark and his critics*. There, informatively, Clark is in fact pressured by both ends of the battlefield (i.e., by Hohwy's cPP approach on the one hand and by Barrett's and Anderson and Chemero's Gibsonian and ecological approaches on the other). Cf. L. BARRETT, *Supercharged apes versus super-sized minds*; M. ANDERSON, A. CHEMERO, *The world well gained*. We see Clark's ecumenical answers, especially to members of the "affordance/ecological" camp, to indicate the existence of these open questions in his proposal.

⁷² Cf. P. GŁADZIEJEWSKI, *Just how conservative is conservative predictive processing?*

⁷³ *Ibid.*, p. 105

⁷⁴ e.g., N. SHEA, *VI-exploitable isomorphism and structural representation*.

⁷⁵ We note again that this claim is not plainly uncontroversial.

⁷⁶ Note that the issue is not with the ambiguity or duplicity of the notion, but with the rigor with which it is able to explain cognitive features. Thus far, it seems that cPP is more able to convey such rigor.

⁷⁷ R.N. SHEPARD, *Ecological constraints on internal representation* as in L. BARRETT, *Supercharged apes versus super-sized minds*.

⁷⁸ J.J. GIBSON, *The senses considered as perceptual systems*.

⁷⁹ Cf., e.g., N.D. DAW, S.J. GERSHMAN, B. SEYMOUR, P. DAYAN, R.J. DOLAN, *Model-based influences on humans' choices and striatal prediction errors*.

⁸⁰ M. ANDERSON, T. CHEMERO, *The problem with brain GUTs*.

⁸¹ M. ANDERSON T. CHEMERO, *The problem with brain GUTs*, p. 205. K. DOŁĘGA, (*Moderate predictive processing*, p. 11), emphasized that PP requires parity from the formal point of view with regard to prediction estimations: «from a formal perspective, each PE level [Prediction Estimation level] in the hierarchy is performing the same kind of basic function – carrying out probabilistic inferences aimed at producing hypotheses, which are best at accommodating the currently available data (incorporating the information fed from the level downstream into the next estimation about the activity on that level)». This supports the contention that double standards are hard to import into a framework like PP.

⁸² This point is based, as D. Williams notes, on the fact that the predictive processing architecture can be seen as a "process theory" (D. WILLIAMS, *Predictive processing and the representation wars*, p. 159).

⁸³ Cf. N. ORLANDI, *Bayesian perception is ecological perception*. It must be further noted that this is a critique that has been moved towards the representational interpretation of as a whole. I couldn't decode this sentence.

⁸⁴ J. HOHWY, *Quick'n'lean or slow and rich?*, p. 195.

⁸⁵ Given that fastness and frugality are related (see above), one point to be better explained is how the pragmatic prerogative of high-level cognition can correlate with richness.

⁸⁶ A. CLARK, *Radical predictive processing*, p. 22.

⁸⁷ J. HOHWY, *Quick'n'lean or slow and rich?*, p. 195.

⁸⁸ A. CLARK, *Replies to critics*, pp. 289-90.

⁸⁹ J. HOHWY in *Quick'n'lean or slow and rich?* highlights how it is precisely this information that allows for sparse sampling to occur.

⁹⁰ Cf. M. MARRAFFA, A. PATERNOSTER, *Models and mechanisms in cognitive sciences*.

⁹¹ W. BECHTEL, G. GRAHAM, A. ABRAHAMSEN, *The life of cognitive science*, p. 96.

Literature

ALKSNIS, N., REYNOLDS, J. (2021). *Revaluating the behaviorist ghost in enactivism and embodied cognition*. In: «Synthese», vol. CXCVIII, n. 6, pp. 5785-5807.

ANDERSON, M. (2014). *After phrenology*, MIT Press, Cambridge (MA).

ANDERSON, M., CHEMERO, A. (2013). *The problem with brain GUTs: Conflation of different senses of "prediction" threatens metaphysical disaster*. In: «Behavioral and Brain Sciences», vol. XXXVI, n. 3, pp. 204-205.

ANDERSON, M., CHEMERO, A. (2019). *The world well gained: On the epistemic implication of ecological information*. In: M. COLOMBO, E. IRVINE, M. STAPLETON (eds.), *Andy Clark and his critics*, Oxford University Press, Oxford, pp. 161-174.

BALLARD, D. (1991). *Animate vision*. In: «Artificial Intelligence», vol. XLVIII, n. 1, pp. 57-86.

BARRETT, L. (2019). *Supercharged apes versus super-sized minds*. In: M. COLOMBO, E. IRVINE, M. STAPLETON (eds.), *Andy Clark and his critics*, Oxford University Press, Oxford, pp. 99-112.

BECHTEL, W. (1998). *Representations and cognitive explanations: Assessing the dynamicist challenge in cognitive science*. In: «Cognitive Science», vol. XXII, n. 3, pp. 295-318.

BECHTEL, W. (2001). *The compatibility of complex systems and reduction: A case analysis of memory research*. In: «Minds and Machines», vol. XI, n. 4, pp. 483-502.

BECHTEL, W. (2007). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*, Routledge, London.

BECHTEL, W. (2012). *Referring to localized cognitive operations in parts of dynamically active brains*. In: A. RAFTOPOULOS, P. MACHAMER (eds.), *Perception, realism and the problem of reference*, Cambridge University Press, Cambridge, pp. 262-284.

BECHTEL, W. (2013). *Addressing the vitalist's challenge to mechanistic science: Dynamic mechanistic explanation*. In: S. NORMANDIN, C.T. WOLFE (eds.), *Vitalism and the scientific image in post-enlightenment life science, 1800-2010*, Springer, Dordrecht, pp. 345-370.

BECHTEL, W., ABRAHAMSEN, A. (2010). *Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science*. In: «Studies in History and Philosophy of Science - Part A», vol. XLI, n. 3, pp. 321-333.

BECHTEL, W., GRAHAM, G., ABRAHAMSEN, A. (1998). *The life of cognitive science*. In: W. BECHTEL, G. GRAHAM (eds.), *A companion to cognitive science*, Blackwell, Oxford, pp. 1-104.

BECHTEL, W., RICHARDSON, R. (2010). *Discovering complexity*, Princeton University Press, Princeton (NJ).

BEER, R. (2000). *Dynamical approaches to cognitive science*. In: «Trends in Cognitive Sciences», vol. IV, n. 3, pp. 91-99.

BEER, R. (2003). *The dynamics of active categorical perception in an evolved model agent*. In: «Adaptive Behavior», vol. XI, n. 4, pp. 209-243.

BROOKS, R.A., FLYNN, A.M. (1989). *Fast, cheap and out*

- of control: *A robot invasion of the solar system*. In: «Journal of the British Interplanetary Society», vol. XLII, pp. 478-485.
- BRUINEBERG, J., KIVERSTEIN, J., RIETVELD, E. (2016). *The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective*. In: «Synthese», vol. CXCIV, n. 6, pp. 2417-44.
- BULLMORE, E., SPORNS, O. (2009). *Complex brain networks: Graph theoretical analysis of structural and functional systems*. In: «Nature Reviews Neuroscience», vol. X, pp. 186-198.
- CARRUTHERS, P. (2006). *The architecture of the mind*, Oxford University Press, Oxford.
- CARRUTHERS, P. (2014). *On central cognition*. In: «Philosophical Studies», vol. CLXX, n. 1, pp. 143-162.
- CARRUTHERS, P. (2015). *The centered mind*, Oxford University Press, Oxford.
- CHEMERO, A. (2009). *Radical embodied cognitive science*, MIT Press, Cambridge (MA).
- CHEMERO, A., SILBERSTEIN, M. (2008). *After the philosophy of mind: Replacing scholasticism with science*. In: «Philosophy of Science», vol. LXXV, n. 1, pp. 1-27.
- CLARK, A. (1997). *Being there*, MIT Press, Cambridge (MA).
- CLARK, A. (2013). *Whatever next? Predictive brains, situated agents, and the future of cognitive science*. In: «Behavioral and Brain Science», vol. XXXVI, n. 3, pp. 181-204.
- CLARK, A. (2015). *Radical predictive processing*. In: «The Southern Journal of Philosophy», vol. LIII, Splindel Supplement, pp. 3-27.
- CLARK, A. (2015). *Predicting peace: The end of the representation wars. A reply to Michael Madary*. In: T. METZINGER, J.M. WINDT (eds.), *Open MIND*, MIND Group, Frankfurt a.M. – doi: 10.15502/9783958570979.
- CLARK, A. (2016). *Surfing uncertainty*, Oxford University Press, New York.
- CLARK, A. (2019). *Replies to critics: In search of the embodied, extended, enactive, predictive (EEE-P) mind*. In: M. COLOMBO, E. IRVINE, M. STAPLETON (eds.), *Andy Clark and his critics*, Oxford University Press, Oxford, pp. 266-302.
- CLARK, A., TORIBIO, J. (1994). *Doing without representing?* In: «Synthese», vol. CI, n. 3, pp. 401-431.
- DAW, N.D., GERSHMAN, S.J., SEYMOUR, B., DAYAN, P., DOLAN, R.J. (2011). *Model-based influences on humans' choices and striatal prediction errors*. In: «Neuron», vol. LXIX, n. 6, pp. 1204-1215.
- DEHAENE, S. (2014). *Consciousness and the brain*, Viking, New York.
- DEISE, E.C. (2008). *Frame problems, Fodor's challenge, and practical reason*, Ph.D. dissertation, University of Maryland, College Park.
- DOŁĘGA, K. (2017). *Moderate predictive processing*. In: T. METZINGER, W. WIESE (eds.), *Philosophy and predictive processing*, MIND Group, Frankfurt a.M. – doi: 10.15502/9783958573116.
- ELIASMITH, C. (1996). *The third contender: A critical examination of the dynamicist theory of cognition*. In: «Philosophical Psychology», vol. IX, n. 4, pp. 441-463.
- ELIASMITH, C. (2007). *How to build a brain: From function to implementation*. In: «Synthese», vol. CLIII, n. 3, pp. 373-388.
- FELDMAN, H., FRISTON, K.J. (2010). *Attention, uncertainty, and free-energy*. In: «Frontiers in Human Neuroscience», vol. IV, Art.Nr. 215 – doi: 10.3389/fnhum.2010.00215.
- GALLAGHER, S. (2017). *Enactivist interventions*, Oxford University Press, Oxford.
- GIBSON, J.J. (1966). *The senses considered as perceptual systems*, Allen and Unwin, London.
- ĞŁADZIEJEWSKI, P. (2016). *Predictive coding and representationalism*. In: «Synthese», vol. CXCIII, n. 2, pp. 559-582.
- ĞŁADZIEJEWSKI, P. (2017). *Just how conservative is conservative predictive processing?* In: «Hybris», vol. XXXVIII, pp. 98-122.
- GONG, G., HE, Y., CONCHA, L., LEBEL, C., GROSS, D.W., EVANS, A.C., BEALIEU, C. (2009). *Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography*. In: «Cerebral Cortex», vol. XIX, n. 3, pp. 524-536.
- HAUGELAND, J. (1998). *Having thought*, Harvard University Press, Cambridge (MA).
- HOHWY, J. (2013). *The predictive mind*, Oxford University Press, Oxford.
- HOHWY, J. (2012). *Attention and conscious perception in the hypothesis testing brain*. In: «Frontiers in Psychology», vol. III, Art.Nr. 96 - doi: 10.3389/fpsyg.2012.00096.
- HOHWY, J. (2019). *Quick'n'lean or slow and rich? Andy Clark on predictive processing and embodied cognition*. In: M. COLOMBO, E. IRVINE, M. STAPLETON (eds.), *Andy Clark and his critics*, Oxford University Press, Oxford, pp. 191-205.
- HUTCHINS, E. (1995). *Cognition in the wild*, MIT Press, Cambridge (MA).
- JACOB, P., JEANNEROD, M. (2003). *Ways of seeing: The scope and limits of visual cognition*, Oxford University Press, Oxford.
- KAPLAN, D.M., BECHTEL, W. (2011). *Dynamical models: An alternative or complement to mechanistic explanations*. In: «Topics in Cognitive Science», vol. III, n. 2, pp. 438-444.
- KAPLAN, D.M., CRAVER, C.F. (2011). *The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective*. In: «Philosophy of Science», vol. LXXVIII, n. 4, pp. 601-627.
- LUNGARELLA, M., SPORNS, O. (2005). *Information self-structuring: Key principles for learning and development*. In: «Proceedings 2005 IEEE International Conference on Development and Learning», pp. 25-30.
- MACHERY, E. (2010). *Philosophy of psychology*. In: F. ALLHOFF (ed.), *Philosophies of the sciences: A guide*, Wiley-Blackwell, Malden (MA), pp. 262-292.
- MARRAFFA, M., PATERNOSTER, A. (2017). *Models and mechanisms in cognitive sciences*. In: L. MAGNANI, T. BERTOLOTI (eds.), *Springer handbook of model-based science*, Springer, Dordrecht/Heidelberg/London/New York, pp. 929-952.
- MCBEATH, M.K., SHAFFER, D.M., KAISER, M.K. (1995). *How baseball outfielders determine where to run to catch fly balls*. In: «Science», vol. CCLXVIII, n. 5210, pp. 569-573.
- MILKOWSKI ET AL. (2018). *From wide cognition to mechanisms: A silent revolution*. In: «Frontiers in Psychology», vol. IX, Art.Nr. 2393 – doi:10.3389/fpsyg.2018.02393.
- MILLIKAN, R. (2001). *Pushmi-pullyu representations*. In: L. MAY, M. FRIEDMAN, A. CLARK (eds.), *Mind and morals*, MIT Press, Cambridge (MA), pp. 145-161.

- NEISSER, U. (1976). *Cognition and reality*, W.H. Freeman, San Francisco.
- NEWEN, A., DE BRUIN, L., GALLAGHER, S. (eds.) (2018). *The Oxford handbook of 4E cognition*. Oxford University Press, Oxford.
- ORLANDI, N. (2016). *Bayesian perception is ecological perception*. In: «Philosophical Topics», vol. XLIV, n. 2, pp. 327-352.
- PORT, R., VAN GELDER, T.J. (eds.) (1995). *Mind as motion*, MIT Press, Cambridge (MA).
- RAO, R., BALLARD, D. (1999). *Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive field-effects*. In: «Nature Neuroscience», vol. II, n. 1, pp. 79-87.
- ROWLANDS, M., LAU, J., DEUTSCH, M. (2020). *Externalism about the mind*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, URL: <https://plato.stanford.edu/archives/win2020/entries/content-externalism/>.
- SANCHES DE OLIVEIRA, G., RAJA, V., CHEMERO, A. (2021). *Radical embodied cognitive science and "real cognition"*. In: «Synthese», vol. CXCVIII, n. 1, pp. 115-136.
- SCHNEIDER, W. (1987). *Connectionism: Is it a paradigm shift for psychology?* In: «Behavior Research Methods, Instruments, and Computers», vol. XIX, n. 2, pp. 73-83.
- SHANAHAN, M. (2010). *Embodiment and the inner life*, Oxford University Press, New York.
- SHAPIRO, L., SPAULDING, S. (2021). *Embodied cognition*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, URL: <https://plato.stanford.edu/archives/win2021/entries/embodied-cognition/>.
- SHEA, N. (2014). *VI-exploitable isomorphism and structural representation*. In: «Proceedings of the Aristotelian Society», vol. CXIV, pp. 123-144.
- SHEPARD, R.N. (1984). *Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming*. In: «Psychological Review», vol. XCI, n. 4, pp. 417-447.
- SIGNA, A., CHELLA, A., GENTILE, M. (2021). *Cognitive robots and the conscious mind: A review of the global workspace theory*. In: «Current Robotics Reports», vol. II, pp. 125-131.
- SPRATLING, M. (2017). *A review of predictive coding algorithms*. In: «Brain and Cognition», vol. CXII, pp. 92-97.
- VAN GELDER, T.J. (1995). *What might cognition be, if not computation?* In: «The Journal of Philosophy», vol. XCII, n. 7, pp. 345-381.
- VAN GELDER, T.J. (1999). *Dynamic approaches to cognition*. In: R. WILSON, F. KEIL (eds.), *The MIT encyclopedia of the cognitive sciences*, MIT Press, Cambridge (MA), pp. 244-246.
- VAN ORDEN, G.C., PENNINGTON, B.F., STONE, G.O. (2001). *What do double dissociations prove?* In: «Cognitive Science», vol. XXV, pp. 111-172.
- VARELA, F.J., THOMPSON, E., ROSCH, E. (1991). *The embodied mind*, MIT Press, Cambridge (MA).
- WALMSLEY, J. (2008). *Explanation in dynamical cognitive science*. In: «Minds and Machines», vol. XVIII, n. 3, pp. 331-482.
- WHYTE, C. (2019). *Integrating the global neuronal workspace into the framework of predictive processing: towards a working hypothesis*. In: «Consciousness and Cognition», vol. LXXIII, Art.Nr. 102763, doi: 10.1016/j.concog.2019.102763.
- WIESE, W. (2016). *What are the contents of representations in predictive processing?* In: «Phenomenology and Cognitive Sciences», vol. XVI, n. 4, pp. 1-22.
- WILLIAMS, D. (2018). *Predictive processing and the representation wars*. In: «Minds and Machines», vol. XXVIII, n. 1, pp. 141-172.
- WILSON, M. (2002). *Six views of embodied cognition*. In: «Psychonomic Bulletin and Review», vol. IX, n. 4, pp. 625-636.

RICERCHE

From unified to specific theories of cognition

Frank van der Velde^(a)

Ricevuto: 28 marzo 2022; accettato: 24 marzo 2023

Abstract This article discusses the unity of cognitive science that seemed to emerge in the 1950s, based on the computational view of cognition. This unity would entail that there is a single set of mechanisms (i.e. algorithms) for all cognitive behavior, in particular at the level of productive human cognition as exemplified in language and reasoning. In turn, this would imply that theories in psychology, and cognitive science in general, would consist of algorithms based on symbol manipulation as found in digital computing. However, a number of developments in recent decades cast doubt on this unity of cognitive science. Also, there are fundamental problems with the claim that cognitive theories are just algorithms. This article discusses some of these problems and suggests that, instead of unified theories of cognition, specific mechanisms for cognitive behavior in specific cognitive domains could be needed, with architectures that are tailor-made for specific forms of implementation. A sketch of such an architecture for language is presented, based on modifiable connection paths in small-world like network structures.

KEYWORDS: Connection Paths; Control of Activation; Small-world Networks; Symbol Manipulation; Unity of Cognition

Riassunto *Dalle teorie unificate della cognizione a quelle specifiche* – Questo articolo discute l'unità della scienza cognitiva che sembrava emergere negli Anni '50 e che era basata su una concezione computazionale della cognizione. Questa unità prevedeva l'esistenza di un singolo insieme di meccanismi (algoritmi) per tutti i comportamenti cognitivi, in particolare al livello della cognizione umana produttiva come, per esempio, linguaggio e ragionamento. A sua volta ciò implicava che le teorie psicologiche e, più in generale della scienza cognitiva, prevedessero algoritmi basati sulla manipolazione di simboli come nella computazione digitale. E, tuttavia, diversi sviluppi degli ultimi decenni hanno messo in dubbio questa unità della scienza cognitiva. Affermare che le teorie cognitive sarebbero solo algoritmi presenta problemi di fondo. Questo articolo discute alcuni di questi problemi, suggerendo che, invece di teorie della cognizione unificate, si potrebbe aver bisogno di meccanismi specifici per il comportamento cognitivo in specifici domini cognitivi, con architetture ritagliate per specifiche forme di implementazione. Questo articolo presenta uno schizzo di una simile architettura per il linguaggio, basata su vie di connessione modificabili in piccoli mondi come le strutture di reti.

PAROLE CHIAVE: Vie di connessione; Controllo dell'attivazione; Reti di piccoli mondi; Manipolazione di simboli; Unità della cognizione

^(a)Cognition, Data, and Education, BMS, University of Twente, Drienerlolaan, 5 – 7522NB Enschede (NL)

E-mail: f.vandervelde@utwente.nl; veldefvander@outlook.com (✉)



1 Introduction

THE AIM OF THIS SPECIAL issue is to address the future of cognitive science(s). The formulation of this aim seems to imply the question of whether we could speak of future of cognitive science, or should speak of the future of cognitive sciences. I will address this issue in particular by focusing on the question of whether a theoretical foundation could be formulated that would account for the unity of cognitive science (both human-level cognition and artificial cognition).

Such a foundation seemed to have emerged in the 1950s, in a period that has been regarded as the beginning of cognitive science.¹ The reason for this is that, at that time, at least three developments came together: the shift from behaviorism to cognitive psychology, the start of artificial intelligence (AI) and the emergence of modern linguistics.

Behaviorism itself, as it arose in the first half of the 20th century, could perhaps be seen as a unified account of the basis of behavior (and in the work of Hull even of human and artificial cognition).² Yet, as formulated by Watson,³ behaviorism originally aimed to explain human behavior based on learning only, specifically conditioning. So, if all human (and animal) behavior resulted from learning, this could imply that humans and animals are adapted to their environments, as these would be the sources of all their learned experiences. This leaves open the possibility that the mechanisms of cognition developed in this way are tailor-made for the specific link between the organism and its environment. In other words, there could be a close link between the specific cognitive architectures involved, the specific way they are implemented, and the cognitive domains they operate in. I will return to this possibility later on.

The emergence of the computational account of cognitive psychology and AI changed this view, at least for some time (e.g., up to the re-emergence of connectionism in the 1980s). A clear example is given by Newell's aim of «unified theories of cognition», which are all based on «a single set of mechanisms for all cognitive behavior».⁴ In Newell's view, this single set of mechanisms is based on symbol manipulation as found in digital computing (e.g., in the von Neumann architecture). In turn, this would be needed to solve the problem of “controlled distal access” (or “logistics of access”),⁵ which in Newell's view is required for any productive cognitive system, as I will discuss in section 5.1.

Another example of the computational view as a unified account of cognition is given by the critique on the re-emergence of connectionism in the 1980s by Fodor and Pylyshyn.⁶ In their critique, they formulated three main features of human-level cognition, given by *productivity*, *compositionally* and *systematicity*, which in their view were not found in connectionist systems. Instead, they

would require symbol manipulation, implemented in digital computational architectures.

The views of Newell⁷ and Fodor and Pylyshyn⁸ are closely related. In fact, the features *productivity*, *compositionally* and *systematicity* each require a computational architecture that possesses logistics of access as analyzed by Newell. Furthermore, these features also play a key role in the current debate on whether deep learning, as given in models like GPT-3, can provide human-level cognition.⁹

Therefore, I will rely in particular on Newell¹⁰ and Fodor and Pylyshyn¹¹ as the basis for the view that computational architectures provide a unified theory of cognition. Here, however, I will only address a few issues related to this view. A discussion of other aspects related to this view can be found in van der Velde.¹²

Recent developments seem to cast doubt on whether unified theories of cognition as intended by Newell¹³ would be possible, which also raises the question of what this would mean for the development of theories of human cognition. I will argue that the aim for unified theories of cognition may be out of reach. Instead, it would perhaps be better to look for specific cognitive architectures, implemented in specific ways and acting within specific cognitive domains. However, although this implies taking the cognitive domains and forms of implementation into account, it does not imply a simple return to behaviorism as we know it. To see why, I will start by briefly describing the transition from behaviorism to cognitive science.

2 From behaviorism to cognitive science

Around 1950, or even earlier, it became clear that classical behaviorism failed to explain human behavior in terms of conditioning only. The first mechanism explored was classical conditioning, based on an already existing (inborn) coupling between a stimulus and a response (reflex). This inborn response is then associated with a new stimulus, as illustrated with the famous study of Pavlov¹⁴ on salivation.

Although classical conditioning undoubtedly works, it is problematic as the basis of all human behavior. Either because we would have to assume that all human behavior derives from just a few existing (inborn) responses, which makes the variability of human behavior difficult to explain. Or, we would have to assume a wide range of already existing, hence inborn, responses as the basis of human behavior, even for language and reasoning. This would be a very problematic assumption for a theory that boasted on explaining all of human behavior on the basis of learning alone, instead of inheritance. In the words of Watson: «we draw the conclusion that there is no such thing as an inheritance of capacity, talent, temperament, mental constitution, and characteristics».¹⁵ This claim is

essentially meaningless if all our behavior derives from a large set of inborn responses.

However, the problem with classical conditioning seemed to be solved with the development of operant conditioning by Skinner. Instead of starting with an existing unconditioned response (behavior), operant conditioning could modify any form of behavior using reinforcement (e.g., reward). So if, say, a monkey accidentally pulled a lever upon which a reward emerged, the monkey would pull that lever more often.

Operant conditioning undoubtedly works as well. Because it can be used for any kind of behavior, it would seem it has solved the problem of classical conditioning. That is, on the assumption that the original behavior it starts with is just "random", inherited forms of behavior need not be assumed. The random behavior could then be modified into more purposeful behavior using reinforcement.

However, the notion of "random" behavior as the basis for learning is very unclear. In the case of a monkey initially pulling a lever, it is clearly not the case that the animal moves its arms and legs randomly until one of them accidentally pulls the lever. Instead, it would look selectively at the lever first and then pull it. Hence, the initial behavior should be seen more as explorative than as random, which again begs the question of where it comes from.

Other examples of explorative behavior were found in behaviorist experiments as well.¹⁶ For example, a rat left on its own in a maze would be able to find the shortest path even before it was trained (rewarded) to do so. To deal with this problem, behaviorists introduced the concept of the "drive", like a "curiosity" drive that would produce the exploratory behavior of an animal. Drives such as curiosity are then rewarded by the behavior of the animal. This would eliminate the need for an external reward to account for learning, but would maintain the idea that all behavior results from learning based on some reward.

But again, drives have to be assumed to exist beforehand, which diminishes the importance of learning in explaining behavior. Also, the sheer amount of different drives needed resulted in an incoherent view about their nature and their relations. In the 1950s, all of this resulted in the shift from behaviorism to cognitive psychology, which emphasized that behavior results from the processing of information, instead of just learning associations.

However, it is remarkable to see that a similar criticism of the behavioristic approach was already formulated much earlier.¹⁷ In the 1910s, Köhler studied how chimpanzees solved so-called "detour" problems. For example, a banana would be visible but just out of reach. However, by making use of material available (e.g., wooden boxes to be used for climbing) the chimpanzees were able to solve the problem,

in that they could get access to the fruit. Although these studies were conducted before the development of operant conditioning, Köhler already noticed the problem of relying on initial random behavior to get learning started. In his words:

In the description of these experiments it should have been apparent that the chief essential is lacking for an explanation by chance actions, that is to say, the means by which the solution is composed out of chance parts is not apparent. Certainly it is not a characteristic of the chimpanzee, when he is brought into an experimental situation, that he should make chance movements out of which, among other things, a non-genuine solution could arise. Very seldom is a chimpanzee seen to attempt any action that would have to be considered accidental in relation to the situation [...] all distinguishable stages of his behavior [...] tend to appear as complete attempts at solutions, of which none appears as the result of accidentally arranged parts. [...] Never, in real and convincing cases, does the solution merge from the disorder of blind impulses. The action is smooth and continuous and can be resolved into parts only *by the abstract thinking* of the observer. In *reality* the parts do *not* appear independently. Thus [...] our theory cannot permit the supposition that [...] the solutions that came as wholes could possibly have arisen from mere chance.¹⁸

This quote and its timing raise the question of why it took until the 1950s before the conclusion could be reached that human (and animal) behavior cannot be (fully) described by means of learning based on conditioning. One reason could be that behaviorism simply had to run its course, before its limitations became more apparent and convincing.

Another reason might be found in philosophy, in particular the philosophy of science. Positivism and later logical positivism were dominant in the first half of the 19th century, and it is clear that Watson was strongly influenced by it.¹⁹ Although Popper had already criticized aspects of logical positivism in the 1930s, it was not until the 1950s that it lost its dominant position in the philosophy of science, based on the work of, e.g., Hanson²⁰ and Kuhn.²¹ It is an interesting question (but beyond the scope of this article) to see if there is indeed a relation between the rise of post-positivism and cognitive psychology (insights from Gestalt psychology were certainly used in post-positivism).

Yet another and perhaps decisive reason could have been the development of the computer, not only as an abstract model of information processing, as with the Turing machine, but also as a practical tool. This gave the idea of what information processing could be and how it could be

developed and tested, as exemplified with the start of AI. Initially, cognitive psychology and AI developed along similar lines, based on the idea that cognition derives from computational processes in the form of symbol manipulation.²²

The emergence of modern linguistics strengthened the notion that cognitive processing is based on symbol manipulation. In 1957 Skinner published a book on verbal behavior, in which he argued that we produce and understand a sentence based on learned associations between words.²³ In the same year Chomsky published a book in which he argued that a sentence has a syntactic structure, which cannot be understood as just an association between words.²⁴ The notion of syntactic structures and the program (grammar) needed to produce and analyse them fitted very well with the computational approach in cognitive psychology and AI that developed around the same time.

3 Competing approaches on the nature of cognition

In the 1950s it seemed that there is just a single cognitive science, dealing with cognition as a form of symbol manipulation (both for human cognition and artificial intelligence). However, since then competing approaches on the nature of cognition, such as connectionism and dynamical approaches, emerged. On its own that would not indicate the end of a single science of cognition. In Kuhn's terms, it could indicate instead that this science has not found its foundational paradigm yet.

However, more recent developments in AI do seem to cast doubt on the unity of cognitive science, as illustrated with Alpha Go and Alpha Go Zero.²⁵ The first is an AI program that learned to play the game Go and succeeded in beating the world champion. Later, a similar program was developed for Chess, with a similar result. However, Alpha Go itself was defeated by Alpha Go Zero (100 to 0, both in Go and Chess). A remarkable difference between these two programs resides in the way they were trained. The Alpha Go program was first trained by using knowledge that had been acquired (by humans) on how to play Go or Chess. Then, the program was developed further by playing against itself, using forms of machine (reinforcement) learning. In contrast, the only forms of knowledge used to train Alpha Go Zero were the rules of the game Go or Chess. Then, the learning procedure based on playing against itself was used to develop the program further.

The significant defeat of the Alpha Go program by Alpha Go Zero raises an important question about the human knowledge on how to play Go or Chess, used with the first program. One could assume that, during the ages, humans would have acquired a lot of knowledge on how to play, e.g., Chess, such as the best ways to start the game or how to re-

spond to the opponent in certain situations. So, it would seem that using that knowledge would be a benefit for an AI program. It would, so to say, have a kick-start with this knowledge and then could learn further. It would certainly be a benefit over a program that was not given this kick-start, but had to find out everything for itself. Or so it would seem.

But the results show otherwise. The knowledge on Chess (or Go) used as kick-start apparently hindered the program its development. It seems as if it was burdened by it and had to unlearn it before it could learn to play Chess in the proper way. But a program not burdened by human knowledge on Chess would develop further and would have the upper hand. The minimal conclusion from this is that, apparently, we do not understand the game of Chess; at least not in the way it could be understood. This raises the question of what that knowledge would be. Clearly, it is engraved in all of the relations etc. learned by the Alpha Go Zero program. But how are we to understand it, even if we could analyse all of these relations?

The example of Alpha Go Zero suggests that our knowledge (cognition) is different from the cognition that could be acquired with certain forms of machine learning. This raises the question of whether the reverse could also be true, and if so, what that would entail about the unity of cognitive science. An example is found in the language behavior program GPT-3.²⁶ This program is based on a neural network and trained on a huge amount of sentences (more than humans see in their lifetime). It can respond to questions or situations by producing answers in fluent English, which suggests that it has mastered the language.

However, Marcus and Davis²⁷ analyzed the behavior of GPT-3 with a number of scenarios. In each case, GPT-3 was given a brief description of a situation, to which it responded. Although this research was not yet intended as a systematic investigation of GPT-3's abilities on language completion or reasoning, a few observations do stand out from the replies given.

Firstly, GPT-3 will produce a response to the scenario given whether or not that response "makes sense" (e.g., is actually or even remotely related to the given scenario). Secondly, even when the response makes no sense in any meaningful way, it is not entirely random. For example, one scenario concerned the use of a cigarette to stir a drink (when a spoon is not available). GPT-3 replied by telling a story about crematoria. This story had nothing to do with the issue at hand, which suggests that it did not understand what the issue was about.

It is, of course, possible not to understand an issue. But, in general, a cognitive agent would (should) be able to reflect on that and acknowledge that it does not understand. From the responses given by GPT-3 on this and other scenarios it seems that it

does not have that form of reflective knowledge. That is, it cannot make a distinction between what it knows and what it does not. Instead, it will just give an answer, apparently based on direct or indirect associations it has learned in training. For example, indirect associative links between cigarettes and crematoria would certainly be present in the learning material for GPT-3, such as the use of fire, the production of ashes, or the strong association with death.

The inability to reflect on what you know or not and giving an answer in all cases regardless of whether it makes sense constitutes a real problem for language understanding, and reasoning (cognition) in general. This could indeed be a difference between learning language and learning to play Chess or any other game. In the latter case, there are real restrictions on what you can do, as given by the rules of the game, the space that the game is confined to, or other constraints on the moves you can make. So even it, say, a Chess program would produce a move like Knight to H9, it would be restrained from performing it because it is not possible.

In language, and cognition in general, such restrictions are far less clear. The examples with GPT-3 indicate that, apparently, these restrictions cannot come from learning sentences only (as noted, GPT-3 is already trained on more sentences than a human will encounter in a lifetime). They would also have to come from learning about the world in a more direct manner. Moreover, humans learn in a different way. Children develop their ability for language in an incremental manner, learning brief sentences and simple scenarios first. Incremental learning is very difficult for neural networks as used in machine learning, because it intervenes with the statistical analysis of the data these networks develop in learning.²⁸

So, there is a possibility that human cognition and cognition acquired with certain forms of machine learning are distinctly different. One could argue here that this conclusion is premature, because further developments with programs like GPT-3 might eventually produce programs that have learned to understand the world in the way that humans do.

But this argument misses the point. Even if it were possible to develop AI systems at the level of human cognition, it is apparently also possible to develop AI systems that are significantly different. Of course, one could still subsume all of this under the same heading of “cognitive science”, but a key issue for that science would then be to understand why, apparently, different forms of cognition are possible, and what that would entail for the general notion of “cognition”.

A first attempt to do this is to have a closer look at the unity of cognitive science that seemed to emerge in the 1950s. As outlined above, the development of the computer played an important role in this, as is also clear from the view of Fodor

and Pylyshyn²⁹ in their well-known analysis of connectionism. They argued that this constituted a return to behaviorism. On that note, Watson’s³⁰ version of behaviorism was based on a positivist’s view of psychology as a science. So, he was not interested in making models. However, the behaviorist Hull was very much interested in developing models on how behavior could be produced.³¹ His models were hand designed and consisted of long chains of stimulus-response associations (reflexes). But in their appearance they are not so different from, say, feedforward neural networks as used in connectionism.

According to Fodor and Pylyshyn, the architecture underlying cognition must be a computational architecture as found in digital computing, such as the Turing machine or the Von Neumann architecture. The reason is that these architectures provide the means to process symbolic structures in a rule-based manner. In turn, this is needed to provide the main features of human-level cognition, given by the related features of productivity, compositionality and systematicity.³²

An example is given by our ability to understand arbitrary “who does what to whom” relations in arbitrary sentences.³³ Systematicity, for example, implies that if you understand that *Sue* is the agent in *Sue eats pizza*, you cannot but understand that *pizza* is the agent in *pizza eats Sue*, even though that is semantically odd. Indeed, we know that *pizza eats Sue* is odd precisely because we identify *pizza* as the agent and *Sue* as the object (theme) of *eat*.³⁴

So, it is no surprise that these features concur with Chomsky’s³⁵ view on the unlimited productivity of language, which in the view of Fodor and Pylyshyn would be achievable only with computational architectures. Here, I want to focus on one aspect of computation as referred to by Fodor and Pylyshyn. It concerns the role of implementation in theories of cognition, discussed in the next section.³⁶

4 The role of implementation in cognition

The topic can be introduced by a quote from Fodor and Pylyshyn on whether it would be useful to know how the architectures they refer to are actually implemented in the brain. Their response is:

The answer [...] has always been that the *implementation*, and all properties associated with the particular realization of the algorithm that the theorist happens to us in a particular case, is irrelevant to the psychological theory; only the algorithm and the representations on which it operates are intended as a psychological hypothesis.³⁷

An algorithm is indeed independent of the way it is implemented. This follows from the theory of computable or recursive functions.³⁸ These are

functions for which an “effective procedure” can be found to compute the function. Recursive function theory shows that a formal definition of what this means cannot be given. But it also shows that all procedures developed thus far are equivalent with (or lesser than) the Turing machine. Turing³⁹ developed the Turing machine to give an answer on the question of what an effective procedure could be. In short, it is a program that can be executed on the Turing machine.

Fodor and Pylyshyn do indeed see the Turing machine, and related architectures such as the Von Neumann architecture, as the basis for cognitive architectures. In this way, one can describe what the aim of a unified cognitive science would be. In line with Newell,⁴⁰ it would consist of developing and studying the “algorithms and the representations on which they operate”, as asserted in the quote above. Hence, this approach equates cognitive theories and models with algorithms, and thus equates any cognitive task with a computable function.

The fundamental problem with this approach is that mathematical functions, of which computable functions are a subset, are inherently static. This is clear from the general definition of a mathematical function, which describes a function as a relation between two sets, the domain (input) and the range (output). The function is characterized by the way it assigns an element from the range to an element of the domain. For a subset of functions, a description of this relation can be given. For example, the numerical function $f(x) = 2x$ assigns the output $2x$ to the input x .

Another subset of mathematical functions is the set of computable functions. For these functions, so-called “effective procedures” or algorithms can be given that produce an output given an input.⁴¹ The function $f(x) = 2x$ belongs to this subset as well, because there are algorithms on the Turing machine and other computers that produce the value of $2x$ for the input x (assuming sufficient memory is available). It is clear now why implementation plays no role here. For example, the output of $f(x) = 2x$ for $x = 3$ is given by 6, because that follows from the function description. The role of the algorithm is only to produce the output 6 for the input 3, otherwise it would in fact compute a different function. The implementation of an algorithm could affect, for example, the time it takes to compute the output, but not the output itself. It makes no sense to say that “today the output of $f(3)$ is 5, because there is no time to compute further”.

But what about cognition? Picture a hominid living on the plains in Africa, who is confronted with an animal. Let’s say that the choice here is between a lion or a deer. At face value, this looks like a functional problem, and indeed the problem has a clear functional aspect. But it is also clear that the behavior of the hominid, and indeed its sur-

vival, critically depends on the time in which an answer is “computed”. This aspect is not included in the definition of a mathematical function. And yet, it is crucial for cognition, because the most fundamental aspect of cognition is to generate the behavior that enhances survival.⁴²

Hence, time is an important factor in cognition but it is (by definition) not included in computation theory. This shows that Fodor and Pylyshyn⁴³ are wrong: “psychological hypotheses” cannot be only “algorithms and the representations on which they operate”. However, as the choice between a lion or deer shows, there are “functional” aspects to the production of behavior. These functional aspects can be integrated with a dynamical (time) constrained description in terms of dynamical systems as functional “flows”.⁴⁴

It is important to understand that this problem cannot be solved by including time as a factor in an algorithm. For example, a computer program that generates a weather prediction will include time as a factor, because e.g. you want to know when the storm arrives and how long it will last. However, this program can be executed on two computers with the same computational precision but one faster than the other. The result will be that both computers generate the same weather prediction, because they run the same algorithm. But the execution times of the program will be different. This shows that the execution time of an algorithm is not a part of the algorithm itself. Yet, it is a part of many cognitive tasks, as illustrated above. So, the processing underlying these tasks is not only depended on an algorithm (i.e. an effective procedure for a computational function).

The main conclusion here is that the computational characterization of cognition that emerged in the 1950 is at best incomplete. This casts doubt on the unity of cognitive science, as it seemed to emerge in that period. Cognition is not just computational, or better, functional (in the mathematical sense of the word). It also depends on the satisfaction of constraints such a speed of processing needed for survival and potentially other constraints as well. These constraints and their relations could be different for different cognitive domains. In particular, certain forms of implementation could be selective for certain forms of information processing.

A glimpse of that could be found in the game of Jeopardy that IBM’s Watson played against the two best human players at that time.⁴⁵ IBM’s Watson won the competition. But there were a number of problems that the humans solved better. Not because IBM’s Watson did not know the answer. It did, but it was too slow in these cases. The (likely) reason of why IBM’s Watson was too slow is that it treated each problem in the same (analytical) manner. For the humans, however, some problems were easier to solve, presumably because

of the stronger activation of associations in their memory in these cases. In turn, this indicates or suggests that human memory is selective. This, again, would be an effect of implementation, as given by the way the brain learns, stores and retrieves information, specified for the environments or domains in which it operates.

Viewed in this way, the competing approaches on the nature of cognition could be more than just an indication that cognitive science has not found its foundational paradigm yet. It could also mean that they target different domains, based on different architectures that are tailored to the way they are implemented. So, those that succeed in winning games like Go and Chess in a way that is perhaps beyond our understanding would not necessarily be the best suited for language. At least not concerning the interactions of language with the environment, as given by the ability to deal with multiple constraints in a limited amount of time, as well as the need for incremental learning.

At this moment, these are just mere suggestions, for which future developments will show to what extent they are true. But, at least, they put a focus on closer interactions between the domains of cognitive processes, the underlying architectures, and the way the architectures are implemented (and are influenced or determined by their implementation). In case of human cognition, this would require a more profound understanding of how cognitive processing relates to the structure and dynamics of the brain. Much more indeed than anticipated or advocated by Fodor and Pylyshyn.⁴⁶ The next section discusses this relation in somewhat more detail.

5 Cognition implemented in the brain

In his *Principles of psychology*, William James stated the following assertion about the relation between cognition and brain (or psychology and neuroscience, if you will):

For the entire nervous system *is* nothing but a system of paths between a sensory *terminus a quo* and a muscular, glandular, or other *terminus ad quem*.⁴⁷

This quote relates to the central thesis of modern neuroscience, as initiated by Cajal, that neurons form connection paths in the brain.⁴⁸ But, in my view, it also reflects a deep insight into what cognition is about and how we should aim to understand it. When push comes to shove, the aim of cognition is to provide the organism with better changes of survival, which is indeed reflected in the ability to act in response to the stimulation received from the environment.

This entails a behavioristic component in cognitive science. However, to be clear, it does not indi-

cate a return to classical behaviorism, just based on conditioned reflexes and drives. As noted in section 2, this was motivated by a positivist's view on the way science (psychology) should operate, which is not implied in James' quote. This quote does not imply either that models of how the brain produces behavior should be reflexive in their nature. As analyzed by Amsel and Rashotte,⁴⁹ this was in fact the main problem of Hull's behavioristic models. Instead, cognition «intervenes in the sensorimotor loop by means of which the creature interacts with its physical and social environments».⁵⁰

In the case of human behavior, all aspects one would attribute to higher-level cognition play a role in this intervention. Fodor and Pylyshyn⁵¹ do in fact make a strong case for productivity, compositionality and systematicity as important features of human-level cognition, as exemplified in language. Incremental learning should be added here as well, also because it is closely related to these features.

Incremental learning entails that a child will learn language starting with small sentences, and is capable of gradually integrating already learned material with newly acquired knowledge. This is very difficult for models like GPT-3, which have to be retrained extensively when they aim to acquire new information. In machine learning with neural networks, re-learning already learned material is needed to prevent undoing the learned knowledge by newly learned material (sometimes referred to as “catastrophic interference”). This behavior is not just an accident but derives directly from the way these networks learn.⁵² An example is found with GPT-3. At some point a mistake was detected in the set-up of the training data. However, to address this afterwards would have required substantial relearning. Due to the costs involved it was decided not to do this.⁵³

However, the architecture that provides these features of cognition would have to be implemented in terms of the structure and dynamics of neural processing in the brain. This requirement, in my view, rules out symbolic architectures that are implemented in a neural manner.⁵⁴ But, as briefly illustrated in the next section, it could be achieved by an architecture that would provide the ability to control connection paths between perception and action, that is, control of connection paths that “intervene in the sensorimotor loop”. Features such as productivity and compositionality could be implemented if the architecture has a connection structure that resembles a small-world like network.⁵⁵ In particular, because the logistics of access needed for these features could be implemented in this way.⁵⁶

5.1 Combinatorial productivity in a small-world like network structure

A key feature of human language is “combinato-

rial productivity”, which concerns the virtually unlimited ability to combine words in arbitrary sentence structures. The ability to combine sound patterns is found in animal communication as well,⁵⁷ which would make the difference in combinatorial productivity between humans and animals merely quantitative. But this difference in quantity is in fact so huge that it becomes a different quality in its own right. The fact that human cognition is singled out by its combinatorial productivity is illustrated in, for example, movies and cartoons. Young children have no difficulty in relating to a character like an artificial sponge living in a pineapple at the bottom the sea, even though that is not derived from their direct experience.

Here, incremental learning and combinatorial productivity come together. Children will have learned what an “agent” is and what it means to live in a house. Effortlessly, they can then recombine these roles and relations even with fantasy characters and environments. For hominids, combinatorial productivity would have been a key ability for their survival, and it is a key ability in the environments we live in, much more so than, say, the ability to play Go or Chess.

Behaviorism, even in the way of Hull, would fail on this. Combinatorial productivity cannot be achieved on the basis of associations alone. For one, it cannot provide associations for relations between agents, or between agents and actions, never seen before. These include, for example, potential relations between agents that are the reverse of relations between those agents that have been learned. So, if a child has learned that a cat could chase a bird, it could also understand what happens if a bird would chase a cat, even though it has not yet seen that, or even if it would never actually occur. Again, this is an ability that makers of, say, movies or cartoons are relying on. It is also an ability that relates to the notions of systematicity and compositionality, as analysed by Fodor and Pylyshyn.⁵⁸

As noted in the introduction, combinatorial productivity in computational terms requires architectures that possess the ability to achieve controlled distal access to information and integrate it in processing,⁵⁹ which I refer to here as logistics of access.⁶⁰ As analyzed by Newell,⁶¹ the need for this ability derives from the fact that, in physical terms, the amount of information that can be stored at a local site is limited. So, with more information required, the architecture needs to have distal access to that information.

An example is given by the internet. The information we can store on a local computer is limited, but we can obtain more information by means of distal access to other computers, and then download that information to influence processing on the local computer.

Another example is found in language. With a lexicon of 60.000 words or more,⁶² word infor-

mation will be stored at different sites in the architecture (as it is in the brain).⁶³ So, distal access to these sites is needed to integrate these words in a sentence structure. This includes the ability to integrate newly learned words directly in sentence structures. As noted above, integrating newly learned information is hard for models such as GPT-3, which is a strong indication that these models do not possess logistics of access.⁶⁴

Computational architectures such as the Turing machine and the Von Neumann architecture do possess logistics of access, which is of course the reason why these architectures seemed to be the basis for cognitive science as it emerged in the 1950s. And, as outlined by Newell,⁶⁵ they achieve logistics of access by using symbols, e.g., to retrieve information and copy it so that it can affect local processing (as in the internet example given above). As a result, cognitive processing would have to consist of forms of symbol manipulation in this view.

But, as argued above, there are serious issues with these architectures from a cognitive perspective. There are also serious issues with them from a neural perspective. For example, Kreite and colleague⁶⁶ proposed that neural codes standing for words are stored in dedicated registers to represent a sentence. So, there would be registers for verbs and for nouns as agents or themes of a verb. This would allow the representation of a sentence like *Bob ate steak*. Indeed, this is how cognitive architectures based on symbol manipulation represent sentences.⁶⁷ But the idea that words are represented in the brain as neural codes that could be (and would have to be) transported to dedicated registers (e.g., using a data bus as in the Von Neumann architecture) to represent a sentence does not match with what is known about the way conceptual information is stored in the brain.

Hebb already suggested that conceptual information would be represented in the brain as interconnected structures he referred to as neural assemblies.⁶⁸ These could and would be distributed over the brain and would develop gradually, as more information about a concept is learned over time. More recent investigations of the structure of semantic memory corroborate this view.⁶⁹

A representation of a concept as a Hebbian assembly precludes its use as a neural code that could be stored in registers. Instead, it will always remain “*in situ*”⁷⁰ even when it is used as a word in a sentence structure. A sentence structure is then a connection path between these *in situ* word (concept) representations.

However, the logistics of access needed for combinatorial productivity imposes major demands on the ability to form connection paths in the architecture. For example, connection paths must be possible between all *in situ* assemblies of nouns and all *in situ* assemblies of verbs, in which any noun could be any of the arguments of any

verb (agent, theme or recipient, depending on the verb). This includes combinations never seen before, such as *bird chases cat*.

As an additional problem, it is unlikely that new connections between neurons or populations are or can be created on the fly. Of course, modifications of existing connections (in particular in the hippocampus and surrounding areas)⁷¹ can occur in short time windows, as in the time window for sentence processing. But growing new connections that would interconnect novel combinations of in situ concept assemblies during verbal communication seems unfeasible.

Given that new connections are not created on the fly in verbal communication, the logistics of access needed for combinatorial productivity has to be achieved with a “fixed” connection structure. This demand holds for any model of neural language representation or processing. For example, Kriete and colleagues’ model⁷² has to account for the fact that the neural code of the word *steak*, or indeed of any other noun, can be stored in each of the registers for nouns. Given that new connections are not created in that process, the model has to assume a fixed connection structure in the brain by which this can be achieved. An example would be a connection structure similar to the Von Neumann architecture of the digital computer. In this architecture, the content from any arbitrary register can be transported (copied) to any other

register by using the central data bus, controlled by the CPU.

However, the computing architecture in which sentence structures are connection paths interconnecting in situ word representations would be very different from the Von Neumann architecture with its data bus controlled by the CPU. Instead, the suggestion is that it could be achieved with a particular connection structure related to the notion of a “small world network”.⁷³

Here, only a brief outline of the motivation of this idea can be given, as illustrated in *Figure 1*.⁷⁴

Panel (a) of *Figure 1* illustrates a connection structure in which every noun is connected to every verb as its agent. A similar connection structure would then have to exist for the theme and recipient arguments of verbs (as with the verb *give*). Two problems arise here. Firstly, the extensive amount of connections needed, given the number of nouns and verbs adults are familiar with (in the order of 10.000 each).⁷⁵ Secondly, it is very hard to see how new nouns and verbs could easily be added to the structure. Yet, we learn new nouns and verbs throughout our lives.

In panel (b), a (very!) basic solution is presented by the introduction of “agent nodes” (neurons, neural populations) that exist in between the nouns and verbs stored in the brain. These agent nodes reduce the overall connection structure needed to connect nouns as agents of verbs. In-

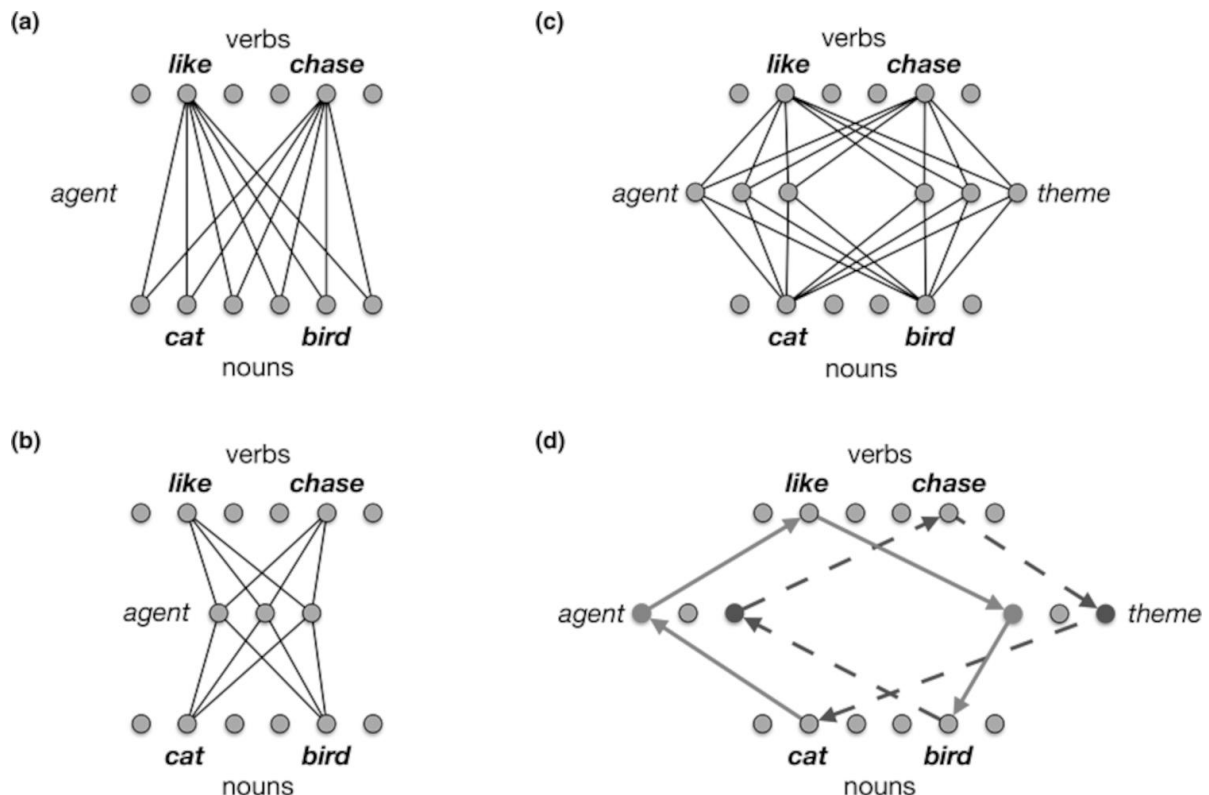


Figure 1. Illustration of connection structures between nouns and verbs. (a) Connections between all noun and verb representations. (b) Connections between nouns and verbs via agent nodes. (c) Combination of agent and theme nodes to connect nouns and verbs. (d) Connection paths for the sentences *Cat likes bird* (red solid) and *Bird chases cat* (blue dashed) in the connection structure of (c).

stead of every noun being connected to every verb, each noun is connected to a more limited set of agent nodes only, which in turn are connected to each verb.

The core of the solution illustrated in *Figure 1* is that the agent nodes operate as the “hubs” one finds in small-world networks. On the one hand, the hubs significantly reduce the number of connections needed, as illustrated by the difference between the panels (a) and (b) in this figure. And yet, on the other hand, the overall connectivity, in which every noun can be an agent of every verb as found in (a), is retained. In (c), the agent nodes are combined with theme nodes that operate on the same principle.

Panel (d) illustrates how connection paths could be established in this connection structure that represent sentences, such as *Cat likes bird* (red-solid) and *Bird chases cat* (blue-dash). The arrows indicate the order of activation in the connection paths that corresponds with hearing or speaking these two sentences. The figure also shows schematically that a word can occur concurrently in two different sentences, and in two different roles.

Of course, the simple schematic connection structure in *Figure 1* raises a number of deep issues. Van der Velde and de Kamps presented in 2006 a first attempt to solve them.⁷⁶ It shows that a reduction in connections can be achieved in this way, and that the introduction of new nouns (or verbs) is easier to account for. In terms of *Figure 1*, a new noun (verb) needs to be connected only to the limited set of agent nodes, instead of to all verbs (nouns). More recent developments of the underlying architecture show that arbitrary sentences in English can be represented and processed as connection paths in this way.⁷⁷

The key aspect of the solution as outlined in *Figure 1* is that it provides the logistics of access needed for combinatorial productivity in a manner that is different from the Von Neumann architecture. In both cases, controlled distal access is needed to information outside a local site. This follows from the physical constraint of the amount of information that can be stored at a local site, and thus applies to any architecture that aims to achieve combinatorial productivity.

However, the Von Neumann architecture uses symbols to retrieve information from a distal site to affect local processing (e.g., as illustrated in the internet example given above). In contrast, in the solution outlined in *Figure 1* information is not stored with symbols. Instead, it is embedded in the network structures at any site in the architecture, as suggested by Hebb⁷⁸ and illustrated by Huth and colleagues.⁷⁹

Distal access is then achieved by temporarily interconnecting distal and local information in a connection path in the architecture, so that information at both sites can be integrated. In this pro-

cess, the information itself remains in situ at all times. The network structure that allows the creation of these connection paths it provided by the small world network structure as outlined in *Figure 1*. Processing in this architecture does not consist of forms of symbol manipulation (as there are no symbols in the architecture) but by means of controlling the creation and use of the connection paths in the architecture. In turn, these differences in processing could have an affect on behaviour, in particular under time constraints.

However, the fact that full combinatorial productivity, as in language, can be achieved in this way shows that this solution is viable alternative for both the von Neumann architecture, which achieves productivity with symbol manipulation, and forms of deep learning as in GPT-3, which as yet lack in full combinatorial productivity.⁸⁰

Furthermore, simulations of the architecture show that sentence processing proceeds differently in architectures like this compared to symbolic processing. These differences emphasize the importance of implementation, even in the case of productive architectures that can represent and process arbitrary “who does what to whom” relations in arbitrary sentences, which is a core feature of human language.⁸¹

6 Conclusions

As Newell argued, the unity of cognitive science would be based on a “single set of mechanisms for all cognitive behavior”,⁸² consisting of computational processing as found in digital computing. Fodor and Pylyshyn⁸³ corroborated this view by equating theories in psychology (cognition) with algorithms based on symbol manipulation.

However, a number of arguments suggest that such a “single set of mechanisms” might not exist. First of all, the equation of theories in psychology or cognitive science with algorithms is simply wrong. It ignores effects of implementation, such as speed of processing. These are irrelevant for algorithms, but important and sometimes vital for cognition. Secondly, different approaches for specific cognitive domains have emerged in recent years, next to the architectures based on symbol manipulation. Some of these have obtained stunning successes, as in playing games like Go and Chess in ways that seem to exceed human understanding.

The combination of different approaches with the importance of implementation for cognition suggests a shift from unified to specific theories of cognition. Each of these would deal with a specific cognitive domain, with architectures that are tailor-made for specific forms of implementation. Of course, they could all still be subsumed under the label “cognitive science”, but this would not fulfill the aim that Newell had with unified theories of cognition.

As noted, the importance of implementation and the constraints imposed by a cognitive domain (for example, the need for survival) entail a return of a behavioristic component in cognitive science. The role of reinforcement learning in machine learning, and indeed the emphasis on learning as the basis of cognition in this field underscores this return. However, it would be a mistake to discard fundamental features of cognition recognized after, and even before, the decline of behaviorism in the 1950s. Among these are the productivity, systematicity and compositionality of human level cognition, as well as the importance of incremental learning.

In the case of human cognition, insight in the role of implementation will be provided by the interaction between neuroscience and psychology. New kinds of architectures have to be developed to satisfy the constraints of both the structure and dynamics of the brain and the fundamental features of cognition and behavior. I illustrated this with a sketch of an architecture in which sentence structures consist of (temporal) connection paths interconnecting in situ word or concept representations, as found in human semantic memory.

Further developments of such architectures are needed. But their successes would show that it is possible to implement fundamental cognitive features like productivity, systematicity and compositionality without relying on forms of symbols manipulation. Hence, instead of relying on a “single set of mechanisms”, they would be achieved with different mechanisms, based on activation control of connection paths in small-world network structures as provided by the structure of the brain.

Notes

¹ Examples are the Dartmouth Project in 1956, which is generally seen as the start of the field of Artificial Intelligence (cf. *The Dartmouth summer research project on artificial intelligence*, available at URL: <https://home.dartmouth.edu/about/dartmouth-milestones>, accessed 14 March 2023), and the review of Skinner’s book on language by Chomsky, which contributed to the decline of behaviorism as the foundation of human-level cognition (e.g., cf. B.M. THORNE, T.B. HENLEY, *Connections in the history and systems of psychology*).

² Cf. A. AMSEL, M.E. RASHOTTE (eds.), *Mechanisms of adaptive behavior*.

³ Cf. J.B. WATSON, *Psychology as the behaviorist views it*; J.B. WATSON, *Behaviorism*.

⁴ A. NEWELL, *Unified theories of cognition*, p. 15.

⁵ Cf. F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.

⁶ Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

⁷ Cf. A. NEWELL, *Unified theories of cognition*.

⁸ Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

⁹ Cf., e.g., J. BROWNING, Y. LE CUN, *What AI can tell us*

about intelligence. In: «Noema», available at URL: <https://www.noemamag.com/what-ai-can-tell-us-about-intelligence/>, accessed 30 September 2022; G. MARCUS, *The next decade in AI: Four steps towards robust Artificial Intelligence*.

¹⁰ Cf. A. NEWELL, *Unified theories of cognition*.

¹¹ J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

¹² Cf. F. VAN DER VELDE, *Towards a neural architecture of language*; F. VAN DER VELDE, *The neural blackboard theory of neuro-symbolic processing*.

¹³ Cf. A. NEWELL, *Unified theories of cognition*.

¹⁴ Cf. I. PAVLOV, *On conditioned reflexes* (1904).

¹⁵ J.B. WATSON, *Behaviorism*, p. 94.

¹⁶ Cf., e.g., D. HOTHERSALL, *History of psychology*.

¹⁷ This example shows that, even when a particular view is dominant at a given moment, different views could also, and often do, exist at the time. Examples of this in case of the dominance of the computational view in the 1950s to 1970s would be Hebb (D.O. HEBB, *The organization of behavior*) or Rosenblatt (F. ROSENBLATT, *The perceptron*).

¹⁸ Cf. W. KÖHLER, *On the insight of apes* (1917), p. 571 – italics added.

¹⁹ Cf. J.B. WATSON, *Psychology as the behaviorist views it*.

²⁰ Cf. N.R. HANSON, *Patterns of discovery*.

²¹ Cf. T.S. KUHN, *The structure of scientific revolutions*.

²² Cf., e.g., Z.W. PYLYSHYN, *Computation and cognition*; A. NEWELL, *Unified theories of cognition*.

²³ Cf. B.F. SKINNER, *Verbal behavior*.

²⁴ Cf. N. CHOMSKY, *Syntactic structures*.

²⁵ Cf. D. SILVER, J. SCHRITTWIESER, K. SIMONYAN, I. ANTONOGLU, A. HUANG, A. GUEZ, T. HUBERT, L. BAKER, M. LAI, A. BOLTON, Y. CHEN, T. LILICRAP, F. HUI, L. SIFRE, G. VAN DER DRIESSCHE, T. GRAEPEL, D. HASSABIS, *Mastering the game of Go without human knowledge*.

²⁶ Cf. T.B. BROWN, B. MANN, N. RYDER, M. SUBBIAH, I. KAPLAN, P. DHARIWAL, A. NEELAKANTAN, P. SHYAM, G. SASTRY, A. ASKELL, S. AGARWAL, A. HERBERT-VOSS, G. KRUEGER, T. HENIGHAN, R. CHILD, A. RAMESH, D.M. ZIEGLER, J. WU, C. WINTER, C. HESSE, M. CHEN, E. SIGLER, M. LITWIN, S. GRAY, B. CHESS, J. CLARK, C. BERNER, S. MCCANDLISH, A. RADFORD, I. SUTSKEVER, D. AMODEI, *Language models are few-shot learners*.

²⁷ Cf. G. MARCUS, E. DAVIS, *GPT-3, Bloviation: OpenAI’s language generator has no idea what it’s talking about*.

²⁸ Cf. F. VAN DER VELDE, *Computation and dissipative dynamical systems in neural networks for classification*.

²⁹ Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

³⁰ Cf. J.B. WATSON, *Psychology as the behaviorist views it*.

³¹ Cf. A. AMSEL, M.E. RASHOTTE (eds.), *Mechanisms of adaptive behavior*.

³² Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

³³ Cf. S. PINKER, *The language instinct*.

³⁴ The entertainment industry, for example, relies heavily on this ability, e.g. in creating fantasy worlds to which we can nevertheless relate, e.g. because they express familiar “who does what to whom” relations.

³⁵ Cf. N. CHOMSKY, *Syntactic structures*.

³⁶ For a discussion on other aspects cf., e.g., F. VAN DER VELDE, *The neural blackboard theory of neuro-symbolic processing*.

³⁷ J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cog-*

nitve architecture: *A critical analysis*, p. 65 – italics by the authors.

³⁸ Cf. H. ROGERS, *Theory of recursive functions and effective computability*.

³⁹ Cf. A.M. TURING, *On computable numbers, with an application to the Entscheidungsproblem*.

⁴⁰ Cf. A. NEWELL, *Unified theories of cognition*.

⁴¹ Cf. H. ROGERS, *Theory of recursive functions and effective computability*.

⁴² Cf. W. JAMES, *The principles of psychology*.

⁴³ Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

⁴⁴ Cf. E.A. JACKSON, *Perspectives of nonlinear dynamics*. As outlined further in F. VAN DER VELDE, *Computation and dissipative dynamical systems in neural networks for classification*.

⁴⁵ Cf., e.g., IBM's Watson Supercomputer Destroys Humans in Jeopardy, available at Endgadget-URL: https://www.youtube.com/watch?v=WFR3I0m_xhE (last visit: 22 March 2022).

⁴⁶ Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

⁴⁷ W. JAMES, *The principles of psychology*, p. 108 - italics by the author.

⁴⁸ Cf. E.R. KANDEL, J.H. SCHWARTZ, T.M. JESSELL, S.A. SIEGELBAUM, A.J. HUDSPETH (eds.), *Principles of neural science*.

⁴⁹ Cf. A. AMSEL, M.E. RASHOTTE (eds.), *Mechanisms of adaptive behavior*.

⁵⁰ M. SHANAHAN, *Embodiment and the inner life*, p. 3.

⁵¹ Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

⁵² Cf. F. VAN DER VELDE, *Computation and dissipative dynamical systems in neural networks for classification*.

⁵³ Cf. T.B. BROWN, B. MANN, N. RYDER, M. SUBBIAH, I. KAPLAN, P. DHARIWAL, A. NEELAKANTAN, P. SHYAM, G. SASTRY, A. ASKELL, S. AGARWAL, A. HERBERT-VOSS, G. KRUEGER, T. HENIGHAN, R. CHILD, A. RAMESH, D.M. ZIEGLER, J. WU, C. WINTER, C. HESSE, M. CHEN, E. SIGLER, M. LITWIN, S. GRAY, B. CHESS, J. CLARK, C. BERNER, S. MCCANDLISH, A. RADFORD, I. SUTSKEVER, D. AMODEI, *Language models are few-shot learners*.

⁵⁴ Cf. T. KRIETE, D.C. NOELLE, J.D. COHEN, R.C. O'REILLY, *Indirection and symbol-like processing in the prefrontal cortex*.

⁵⁵ Cf. D.J. WATTS, S.H. STROGATZ, *Collective dynamics of "small-world" networks*; M. SHANAHAN, *Embodiment and the inner life*.

⁵⁶ Cf. F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.

⁵⁷ Cf., e.g., M.D. HAUSER, N. CHOMSKY, W.T. FITCH, *The faculty of language: What is it, who has it, and how did it evolve?*

⁵⁸ Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture*.

⁵⁹ Cf. A. NEWELL, *Unified theories of cognition*.

⁶⁰ Cf. F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.

⁶¹ Cf. A. NEWELL, *Unified theories of cognition*.

⁶² Cf. P. BLOOM, *How children learn the meaning of words*.

⁶³ Cf., e.g., A.G. HUTH, W.A. DE HEER, T.L. GRIFFITHS, F.E. THEUNISSEN, J.L. GALLANT, *Natural speech reveals the semantic maps that tile human cerebral cortex*.

⁶⁴ F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.

⁶⁵ Cf. A. NEWELL, *Unified theories of cognition*.

⁶⁶ Cf. T. KRIETE, D.C. NOELLE, J.D. COHEN, R.C. O'REILLY, *Indirection and symbol-like processing in the prefrontal cortex*.

⁶⁷ Cf., e.g., A. NEWELL, *Unified theories of cognition*.

⁶⁸ Cf. D.O. HEBB, *The organisation of behavior: A neuro-psychological theory*.

⁶⁹ Cf., e.g., A.G. HUTH, W.A. DE HEER, T.L. GRIFFITHS, F.E. THEUNISSEN, J.L. GALLANT, *Natural speech reveals the semantic maps that tile human cerebral cortex*; M.A. LAMBON-RALPH, E. JEFFERIES, K. PATTERSON, T.T. ROGERS, *The neural and computational bases of semantic cognition*.

⁷⁰ Cf. F. VAN DER VELDE, J. FORTH, D.S. NAZARETH, G.A. WIGGINS, *Linking neural and symbolic representation and processing of conceptual structures*.

⁷¹ Cf., e.g., R.C. O'REILLY, J.W. RUDY, *Conjunctive representations in learning and memory: Principles of cortical and hippocampal function*.

⁷² Cf. T. KRIETE, D.C. NOELLE, J.D. COHEN, R.C. O'REILLY, *Indirection and symbol-like processing in the prefrontal cortex*.

⁷³ Cf. D.J. WATTS, S.H. STROGATZ, *Collective dynamics of "small-world" networks*; M. SHANAHAN, *Embodiment and the inner life*.

⁷⁴ More details on the underlying mechanisms and examples can be found in, e.g., F. VAN DER VELDE, M. DE KAMPS, *Neural blackboard architectures of combinatorial structures in cognition*; F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*; F. VAN DER VELDE, *The neural blackboard theory of neuro-symbolic processing: Logistics of access, connection paths and intrinsic structures*.

⁷⁵ Cf. P. BLOOM, *How children learn the meaning of words*.

⁷⁶ Cf. F. VAN DER VELDE, M. DE KAMPS, *Neural blackboard architectures of combinatorial structures in cognition*.

⁷⁷ Cf. F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*; F. VAN DER VELDE, *The neural blackboard theory of neuro-symbolic processing: Logistics of access, connection paths and intrinsic structures*.

⁷⁸ Cf. D.O. HEBB, *The organisation of behavior*.

⁷⁹ Cf. A.G. HUTH, W.A. DE HEER, T.L. GRIFFITHS, F.E. THEUNISSEN, J.L. GALLANT, *Natural speech reveals the semantic maps that tile human cerebral cortex*.

⁸⁰ Cf. F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.

⁸¹ Cf. S. PINKER, *The language instinct*. It is beyond the scope of this article to outline this further, for an overview cf., e.g., F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.

⁸² Cf. A. NEWELL, *Unified theories of cognition*.

⁸³ Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

Literature

AMSEL, A., RASHOTTE, M.E. (eds.) (1984). *Mechanisms of adaptive behavior: Clark L. Hull's theoretical pa-*

- pers, with commentary*, Columbia University Press, New York.
- BLOOM, P. (2000). *How children learn the meaning of words*, MIT Press, Cambridge (MA).
- BROWN, T.B., MANN, B., RYDER, N., SUBBIAH, M., KAPLAN, I., DHARIWAL, P., NEELAKANTAN, A., SHYAM, P., SASTRY, G., ASKELL, A., AGARWAL, S., HERBERT-VOSS, A., KRUEGER, G., HENIGHAN, T., CHILD, R., RAMESH, A., ZIEGLER, D.M., WU, J., WINTER, C., HESSE, C., CHEN, M., SIGLER, E., LITWIN, M., GRAY, S., CHESSE, B., CLARK, J., BERNER, C., MCCANDLISH, S., RADFORD, A., SUTSKEVER, I., AMODEI, D. (2020). *Language models are few-shot learners*. In: «ArXiv», arXiv:2005.14165v4 - doi: 10.48550/arXiv.2005.14165, last revision: 22 July 2020.
- BROWNING, J., LE CUN, Y. (2022). *What AI can tell us about intelligence*. In: «Noema», available at URL: <https://www.noemamag.com/what-ai-can-tell-us-about-intelligence/>
- CHOMSKY, N. (1957). *Syntactic structures*, Mouton, The Hague.
- FODOR, J.A., PYLYSHYN, Z.W. (1988). *Connectionism and cognitive architecture: A critical analysis*. In: S. PINKER, J. MEHLER (eds.), *Connections and symbols*, MIT Press, Cambridge (MA), pp. 73-193.
- HANSON, N.R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*, Cambridge University Press, Cambridge.
- HAUSER, M.D., CHOMSKY, N., FITCH, W.T. (2002). *The faculty of language: What is it, who has it, and how did it evolve?*. In: «Science», vol. CCXCVIII, n. 5598, pp. 1569-1579.
- HEBB, D.O. (1949). *The organisation of behavior: A neuropsychological theory*, Wiley, New York.
- HOTHERSALL, D. (2004). *History of psychology*, McGraw-Hill, Boston, 4th edition.
- HUTH, A.G., DE HEER, W.A., GRIFFITHS, T.L., THEUNISSEN, F.E., GALLANT, J.L. (2016). *Natural speech reveals the semantic maps that tile human cerebral cortex*. In: «Nature», vol. DXXXII, n. 7600, pp. 453-458.
- JACKSON, E.A. (1991). *Perspectives of nonlinear dynamics*, 2 voll., Cambridge University Press, Cambridge.
- JAMES, W. (1950). *The principles of psychology* (1890), 2 voll., Dover Publications.
- KANDEL, E.R., SCHWARTZ, J.H., JESSELL, T.M., SIEGELBAUM, S.A., HUDSPETH, A.J. (eds.) (2013). *Principles of neural science*, McGraw-Hill, New York, 5th edition.
- KÖHLER, W. (1965). *On the insight of apes* (1917). In: R.J. HERRNSTEIN, E.G. BORING (eds.), *A source book in the history of psychology*, Harvard University Press, Cambridge (MA), pp. 569-578.
- KRIETE, T., NOELLE, D.C., COHEN, J.D., O'REILLY, R.C. (2013). *Indirection and symbol-like processing in the prefrontal cortex*. In: «Proceedings of the Academy of Sciences of the United States of America», vol. CX, n. 41, pp. 16390-16395.
- KUHN, T.S. (1970). *The structure of scientific revolutions* (1962), University of Chicago Press, Chicago, 2nd edition.
- RALPH, M.A.L., JEFFERIES, E., PATTERSON, K., ROGERS, T.T. (2017). *The neural and computational bases of semantic cognition*. In: «Nature Reviews Neuroscience», vol. XVIII, n. 1, pp. 42-55.
- MARCUS, G. (2020). *The next decade in AI: Four steps towards robust Artificial Intelligence*. In: «ArXiv», arXiv:2002.06177 – doi: 10.48550/arXiv.2002.06177 – last revision 2020, February 19th.
- MARCUS, G., DAVIS, E. (2020). *GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about*. In: «MIT Technological Review», published: 22 August 2020, available at URL: <https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/>.
- NEWELL, A. (1990). *Unified theories of cognition*, Harvard University Press, Cambridge (MA).
- O'REILLY, R.C., RUDY, J.W. (2001). *Conjunctive representations in learning and memory: Principles of cortical and hippocampal function*. In: «Psychological Review», vol. CVIII, n. 2, pp. 311-345.
- PAVLOV, I.P. (1965) *On conditioned reflexes* (1904). In: R.J. HERRNSTEIN, E.G. BORING (eds.), *A source book in the history of psychology*, Harvard University Press, Cambridge (MA), pp. 564-569.
- PINKER, S. (1994). *The language instinct*, Penguin, London.
- PYLYSHYN, Z.W. (1984). *Computation and cognition: Toward a foundation for cognitive science*, MIT Press, Cambridge (MA).
- ROGERS, H. (1988). *Theory of recursive functions and effective computability*, MIT Press, Cambridge (MA).
- ROSENBLATT, F. (1958). *The perceptron: A probabilistic model of information storage and organization in the brain*. In: «Psychological Review», vol. LXV, n. 6, pp. 386-408.
- SHANAHAN, M. (2010). *Embodiment and the inner life*, Oxford University Press, Oxford.
- SILVER, D., SCHRITTWIESER, J., SIMONYAN, K., ANTONOGLOU, I., HUANG, A., GUEZ, A., HUBERT, T., BAKER, L., LAI, M., BOLTON, A., CHEN, Y., LILLICRAP, T., HUI, F., SIFRE, L., VAN DER DRIESSCHE, G., GRAEPEL, T., HASSABIS, D. (2017). *Mastering the game of Go without human knowledge*. In: «Nature», vol. DL, n. 7676, pp. 354-359.
- SKINNER, B.F. (1957). *Verbal behavior*, Appleton-Century-Crofts, New York.
- THORNE, B.M., HENLEY, T.B. (2001). *Connections in the history and systems of psychology*, Houghton Mifflin, Boston.
- TURING, A.M. (1937). *On computable numbers, with an application to the Entscheidungsproblem*. In: «Proceedings of the London Mathematical Society», vol. XLII, S2, n. 1, pp. 230-265.
- VAN DER VELDE, F. (in press). *The neural blackboard theory of neuro-symbolic processing: Logistics of access, connection paths and intrinsic structures*. In: P. HITZLER, K. SARKER, A. EBERHART (eds.), *Compendium of neuro-symbolic artificial intelligence*, IOS Press, Amsterdam.
- VAN DER VELDE, F. (2022). *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*. In: «ArXiv», arxiv.org/abs/2210.10543 – doi: 10.48550/arXiv.2210.10543.
- VAN DER VELDE, F. (2015). *Computation and dissipative dynamical systems in neural networks for classification*. In: «Pattern Recognition Letters», vol. LXIV, pp. 44-52.
- VAN DER VELDE, F., DE KAMPS, M. (2006). *Neural blackboard architectures of combinatorial structures in cognition*. In: «Behavioral and Brain Sciences»,

- vol. XXIX, n. 1, pp. 37-70.
- VAN DER VELDE, F., FORTH, J., NAZARETH, D.S., WIGGINS, G.A. (2017). *Linking neural and symbolic representation and processing of conceptual structures*. In: «Frontiers in Psychology», vol. VIII, Art. Nr. 1297 - doi: 10.3389/fpsyg.2017.01297.
- WATSON, J.B. (1913). *Psychology as the behaviorist views it*. In: «Psychological Review», vol. XX, n. 2, pp. 158-177.
- WATSON, J.B. (1924). *Behaviorism*, The People's Institute Publishing Co., New York.
- WATTS, D.J., STROGATZ, S.H. (1998). *Collective dynamics of "small-world" networks*. In: «Nature», vol. CCCXCIII, n. 6684, pp. 440-442.

RICERCHE

Rethinking cognitive architecture: A heterarchical network of different types of information processors

William Bechtel^(a)

Ricevuto: 24 marzo 2022; accettato: 6 marzo 2023

Abstract Rather than seeking a common architecture for cognitive processing, this paper argues that we should recognize that the brain employs multiple information processing structures. Many of these are manifest in brain areas outside the neocortex such as the hypothalamus, brain stem pattern generators, the basal ganglia, and various nuclei releasing neuromodulators. Rather than employing one mode of information processing, the brain employs multiple modes integrated in a heterarchical network. These in turn affect processing within the neocortex and together with the neocortex regulate vertebrate behavior, including human. Cognitive science can better understand human information processing by attending to the plurality of information-processing architectures employed in the brain.

KEYWORDS: Basal Ganglia; Cognitive Architectures; Heterarchical Networks; Hypothalamus; Pattern Generators; Neuromodulators

Riassunto *Ripensare l'architettura cognitiva: una rete eterarchica di differenti tipi di elaborazione di informazione* - Anziché cercare un'architettura comune ai processi cognitivi, in questo lavoro si sosterrà che dovremmo riconoscere come il cervello impieghi molteplici strutture per processare l'informazione. Molte di loro si trovano nelle aree cerebrali esterne alla neocorteccia, come l'ipotalamo, i generatori di pattern del tronco encefalico, i gangli basali e i vari nuclei che rilasciano i neuromodulatori. Anziché impiegare un'unica modalità per l'elaborazione dell'informazione, il cervello usa molteplici modalità integrate in una rete eterarchica. Queste, a loro volta, influenzano i processi all'interno della neocorteccia e, assieme alla neocorteccia, regolano il comportamento dei vertebrati, compreso quello umano. La scienza cognitiva può meglio comprendere l'elaborazione dell'informazione da parte degli esseri umani concentrandosi sulla pluralità delle architetture impiegate nel cervello per realizzare questa elaborazione.

PAROLE CHIAVE: Gangli basali; Architetture cognitive; Reti eterarchiche; Ipotalamo; Generatori di pattern; Neuromodulatori

^(a)Department of Philosophy, University of California San Diego - 9500 Gillman Drive, La Jolla, CA 92093 (USA)

E-mail: wbechtel@ucsd.edu (✉)



1 Introduction

Cognitive architectures have been developed in cognitive science to characterize the basic modes of processing information used in cognitive activities. These have, in general, featured a homogeneous set of capacities. Newell's SOAR architecture, for example, is a production system architecture in which matching items in working memory to the antecedents of rules results in changing the items in working memory.¹ Anderson's ACT employs a broader range of operations, but like SOAR treats cognitive activities as involving the application of rules to symbolically represented information. Developers of neural network architectures² broke with symbolic architectures by drawing inspiration from the brain. With the development of deep learning allowing for networks with many layers, investigators have developed network models of visual processing that map onto regions in the neocortex.³ While the brain provided the inspiration for neural network architectures, only one part of the brain, the neocortex, was modeled. The neocortex is not representative of the rest of the brain. Moreover, it does not function on its own but is highly integrated with processing in other brain regions, especially the thalamus and the basal ganglia, which are in turn highly integrated with areas of the brainstem and midbrain. I will argue that in characterizing cognitive architectures, one should look more broadly in the brain. When one does so, one finds a heterarchical network of components implementing different information-processing architectures. The operation of the neocortex is integrated into the larger heterarchical network. My goal in this paper is to make the case that attention to these architectures and how they interact in the control of behavior can provide a more adequate grounding for cognitive science.

In arguing for a heterarchical network of different cognitive architectures, I embrace a perspective defended by Maturana and Varela,⁴ among others, that all living organisms engage in cognition to regulate the activities through which they construct, maintain, and repair themselves. These activities must be regulated so that they are employed when and in the manner needed to maintain the organism as an autonomous system.⁵ Recently a number of investigators have identified and sought to characterize cognitive processes in prokaryotes, plants, and invertebrates.⁶ This is providing important new insights into how organisms process information. In this paper, however, I will limit my focus to vertebrates, and the role of the central nervous system in regulating the activities of vertebrates. Through extensive research comparing the human brain with that of the lamprey, a phylogenetic far distant vertebrate, Grillner⁷ has argued that, except for the cerebral cor-

tex, all vertebrate brains exhibit the same set of neural components organized in much the same manner. These include the diverse subcortical areas on which I will focus.

Although lacking a cerebral cortex (the neocortex plus structures such as the hippocampus), the lamprey does have a small pallium, a structure from which the structures in the cerebral cortex evolved. The pallium is, however, a relatively minor structure in the lamprey. Without cortical areas, the brains of lamprey, and presumably the earliest vertebrates, are able to process the information needed to perform the activities required to live. These activities include behaviors in an animal's environment, such as eating and defending itself. Moreover, research on decorticate preparations has shown that even mammals such as cats can, at least in the protected environment of the laboratory, conduct their lives without a cortex.⁸ Their activities include, among others, the basic cognitive activity of making decisions about which actions to perform in different situations.⁹ My focus will be on the different resources provided by some of the subcortical areas present in all vertebrates and preserved in decorticated mammals. With an appreciation of the type of information processing provided by these brain regions, I return in the section 6 to the neocortex and advance a perspective that its architecture, like that of the tectum that performs many of the same functions in early vertebrates, underlies a distinctive type of information processing that does contribute in important ways to the cognitive life of organisms that possess it. However, the neocortex is not a monolith, and individual regions of the neocortex are highly integrated with other brain regions that in many respects direct and regulate processing within it.

I cannot, in a short paper, present a comprehensive account of the different types of information processing architectures found in the vertebrate brain. Rather, I offer a brief introduction to five types of subcortical processing that together provide a perspective on the diversity of information-processing architectures on which vertebrates rely. I begin with the hypothalamus, whose nuclei serve to assess the needs of the organism (section 2), and central pattern generators and locomotor centers, that serve to coordinate muscle activity (section 3). The motor system is capable of generating many actions and some means of selecting which action to perform at a given time is required. This selection is performed by the basal ganglia, to which I turn in section 4.

Processing in the basal ganglia is influenced by dopamine. As I discuss in section 5, dopamine as well as other monoamines, such as serotonin and acetylcholine, are referred to as neuromodulators since they alter processing in neural circuits. I will argue that the nuclei that release these neuromod-

ulators might better be seen as setting agendas for processing information. Many accounts of cognition treat it as processing sensory information, but none of the brain regions discussed so far processes sensory inputs. However, that is clearly required for effective motor activity. Although areas in the neocortex are viewed as principal areas for sensory processing, early vertebrates and decorticated animals rely on the tectum (also known as the superior colliculus). In section 6 I will briefly characterize information processing in the tectum and how the neocortex elaborates on its mode of processing. Processing sensory inputs, on the perspective I am advancing, is only one relatively specialized type of information processing underlying cognition, one that is used by other centers engaged in different modes of information processing.

The neural systems I discuss in sections 2-6 process information in diverse ways, each useful in enabling organisms to function in the world. To make sense of these diverse means of processing information, I return in the final section to how these modes of information processing serve the needs of an organism. An analysis of the brain that starts with the hypothalamus and brainstem systems of motor control focuses attention not on high-level abstract reasoning, but on directing activity needed for an organism to maintain itself. I also briefly consider the question of how the processing in these diverse areas, relying on different architectures, is coordinated to serve the organism. I challenge the widespread view that control systems such as the brain must be organized hierarchically, with the neocortex, and especially prefrontal regions of the neocortex, in charge. Instead, I argue that components with different cognitive architectures form a heterarchical network and together provide a robust stem for directing the activities of vertebrates.

One might object that the information processing in areas other than the cerebral cortex, while important for life, is not *cognitive* information processing. Debating the use of terms is generally not productive. One could restrict the term to conceptual reasoning,¹⁰ solving problems,¹¹ or pattern recognition.¹² This, however, creates an unnatural division between domains of information processing, treating more traditionally cognitive activities as occurring autonomously from other information processing in the brain. It is noteworthy that much high-level human reasoning takes place in the context of recognized needs of the organism and is coordinated with action. Food is never far from people's minds, even when engaged high-level intellectual activities such as participating in academic conferences. Moreover, as I have already noted, neural activity in the cerebral cortex is integrated with that in the thalamus, and via the thalamus, with that of the brainstem and midbrain. The architecture in regions of the

cerebral cortex is distinctive, but the information processing it carries out is integrated with that of the rest of the brain. I will revisit this theme after first describing the type of information processing used in other brain regions.

2 The hypothalamus: Assessing the needs of the organism

If organisms are to maintain themselves as systems far-from-equilibrium with their environments, they need to perform actions appropriate to their current condition. This means that they must procure information about their condition: Do they have sufficient nutrients? Are they confronting specific stressors? These assessments are made by nuclei in the brainstem (e.g., the nucleus of the solitary tract) and, in many cases, further processed in the hypothalamus. The hypothalamus comprises several small nuclei located adjacent to the median eminence at the base of the diencephalon. This is one of the few places in the brain without a blood-brain barrier.

The lack of a blood-brain barrier is important as the nuclei of the hypothalamus extend the information processing already performed in the endocrine system. The endocrine system relies on sensors that release chemicals that are transported through the blood stream to locations where they bind receptors and trigger biochemical activity. For example, in contexts in which ATP levels are low, indicating low reserves of energy and the need to procure more energy, the sight or smell of food or the digestion of it elicits insulin secretion from pancreas β cells into the bloodstream. Various other cells have insulin receptors and, when activated, initiate related activities.¹³ For example, when liver cells detect insulin, they store glucose they don't need immediately as glycogen (without insulin, they extract glucose from glycogen and release it into the bloodstream). Insulin also promotes glucose uptake in adipose tissue where it is metabolized for fatty acid synthesis. Through these various responses, the organism directs the glucose it is acquiring but doesn't need immediately to replenishing energy reserves.

The fenestrated capillaries of the median eminence allow hormones in the blood to affect and be affected by neurons in the nuclei of the hypothalamus. Polypeptides comparable to those circulating in the blood provide a major currency for transmission between hypothalamic neurons. Polypeptides synthesized by neurons are referred to as *neuropeptides*. When neurons secrete a neuropeptide, it disseminates widely through the extracellular matrix. These neuropeptides can then bind to other neurons that have appropriate G-protein coupled receptors (GPCRs). Unlike ionotropic receptors, which respond by directly opening or closing ion channels, these receptors are

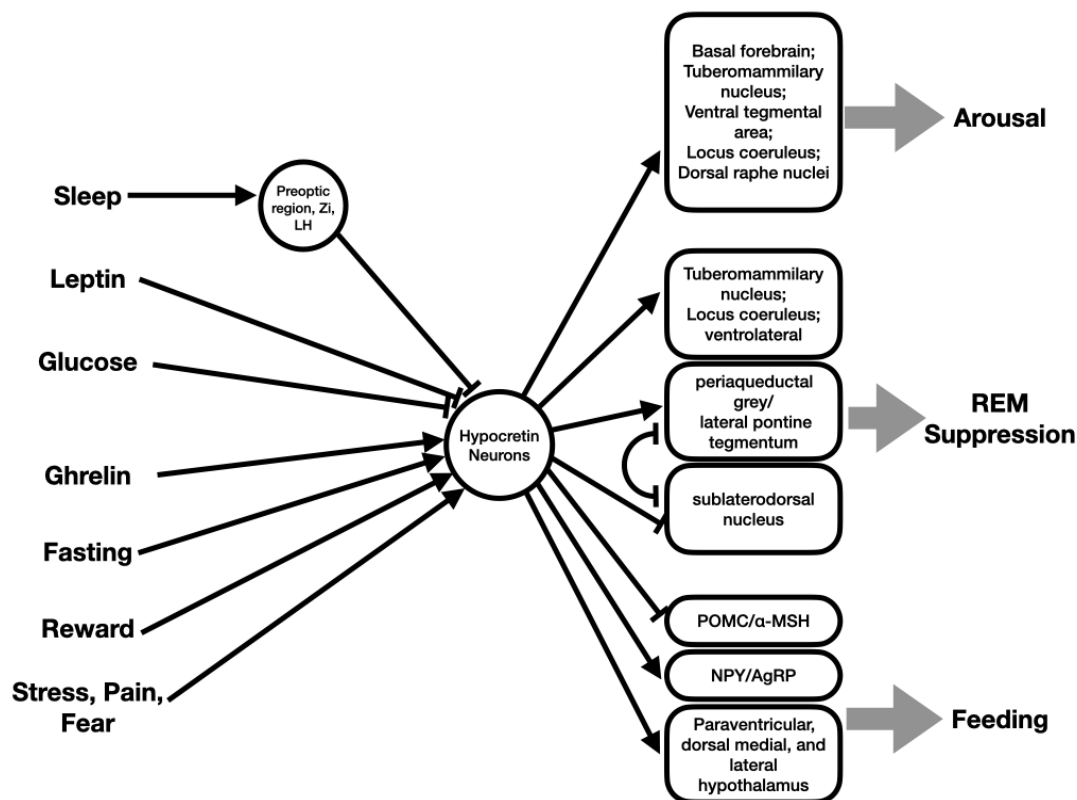


Figure 1. Inputs and outputs of hypocretin neurons, based on data reported by E. ARRIGONI, M.J.S. CHEE, P.M. FULLER, *To eat or to sleep: That is a lateral hypothalamic question.*

metabotropic: when appropriate ligands bind, they initiate intracellular signaling employing second messengers to initiate various metabolic activities, including the expression of targeted genes. The information processing in which these neurons engage is comparable to the chemical information processing in the endocrine system, with the difference that electrical transmission along axons and dendrites also enables delivering neuropeptides as well as more traditional neurotransmitters to neurons at distal locations.

Some hypothalamic nuclei contain cell populations that appear to respond to specific conditions in the body. The arcuate nucleus, for example, contains neurons that respond to different nutrient needs. Among other inputs, POMC neurons respond to leptin, a hormone synthesized in adipose and related cells in proportion to fat mass in the small intestine and then transmitted in blood to the hypothalamus.¹⁴ Accordingly, POMC neurons signal satiety. A second population, consisting of AgRP neurons, responds to ghrelin synthesized in the stomach and duodenum, especially when no food is being digested (transmission from the stomach to the hypothalamus is likely via the vagus nerve, which has ghrelin receptors, and processed in the brainstem, with new ghrelin being synthesized in the arcuate nucleus). AgRP neurons also respond to other inputs indicating lack of food and so signal hunger.

Both POMC and AgRP neurons send outputs

widely to other nuclei in the hypothalamus and locations elsewhere in the brain. One target is the lateral hypothalamic area (LHA). Neurons in the LHA promote feeding behavior and in 1998 two groups of researchers identified a neuropeptide synthesized in the LHA that specifically promotes feeding behavior. One group named it *orexin*,¹⁵ the other *hypocretin*.¹⁶ The name *orexin* is derived from the Greek word for appetite, signaling that it was assumed to be principally involved in initiating feeding behavior. Shortly after, however, other researchers revealed that these neurons also fire maximally before sleep-to-wake transitions and demonstrated that they promote these transitions. *Figure 1* indicates some of the known inputs and outputs of hypocretin neurons.

Hypocretin neurons represent the norm, not an exception, in the hypothalamus. They can be distinguished from one another in terms of their inputs, the peptides they synthesize, and patterns of connections, but in general hypothalamic neurons are not single purpose processors. Rather, each nucleus acts as a hub, integrating different information about the state of the organism and disseminating it to many other regions of the hypothalamus and elsewhere in the brain. As Saper and Lowell discuss,¹⁷ there are hundreds, perhaps thousands of such cell populations. Such an information processing system is challenging to understand but plays a central role in coordinating vertebrate behavior both through interaction with

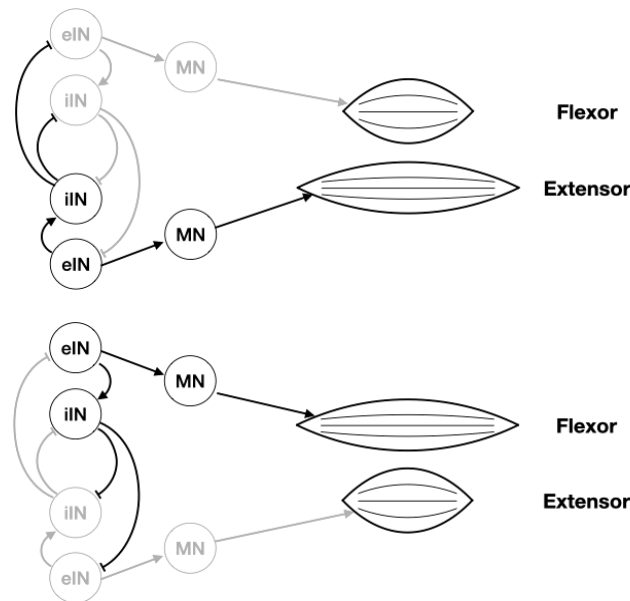


Figure 2. Local pattern generator controlling first contraction of the flexor muscle followed by contraction of the extensor muscle. In the top frame, the neurons in red are active, and the flexor inhibitory interneuron (iIN) inhibits both the extensor iIN and the excitatory interneuron (eIN) as well as driving the flexor motor neuron (MN) which initiates contraction of the flexor muscle. When the activity of the flexor iIN neuron eventually wanes, the extensor iIN, eIN, and MN become active, and cause the extensor muscle to contract.

other brain regions and by directing the synthesis of hormones that directly regulate biochemical processes and behavioral activities. Among these activities are those involved in eating.

Before leaving the hypothalamus, I briefly note one nucleus that serves an important function in regulating nearly all activities of the organism. The suprachiasmatic nucleus (SCN, so named because of its location above the optic chiasm where neurons from the two eyes come together before distributing information to other brain regions) maintains an endogenous circadian (approximately 24 hour) rhythm. Although the rhythm is generated endogenously in individual SCN neurons, since the rhythm is only approximately 24-hours, their activity need to be regularly updated. Accordingly, they receive inputs both from the retina and neuropeptides released in other hypothalamic areas. SCN neurons then distribute this signal broadly, relying sometimes of electrical transmission along neurons and sometimes on neuropeptide dissemination, to tissues throughout the body. The peripheral cells use their circadian oscillation to regulate gene expression, expressing genes appropriate for the activities the cells need to perform at specific times of day. Among the areas regulated by circadian rhythms is the neocortex. As a result, performance of high-level cognitive activities varies over the course of a day.

The nuclei of the hypothalamus act as hubs that extend the information processing capacities of the endocrine system, integrating information from multiple sources and disseminating the resulting activity widely. In many cases the outputs

consist of peptides that disseminate back into the bloodstream and function in the same manner as endocrines. In other cases, outputs are directed to other brain regions, including regions that can elicit motoric responses. Together, these nuclei provide an effective means of informing the various response capacities of the organism with information about when the conditions of the organism necessitate a response.

3 Pattern generators and locomotor regions: Controlling skeletal muscle

Many physiological activities (e.g., blood circulation, digestion) and all motor activities of animals require muscles. Individual muscle fibers can exert force by contracting, but a single muscle fiber acting on its own exerts insufficient force to perform the needed activity. Rather, many muscles must contract in a coordinated fashion. In recognition of this, Keijzer, van Duijn, and Lyon¹⁸ argue that muscle coordination, not processing sensory inputs, was the activity for which neurons first evolved. Keijzer and colleagues appeal to the jellyfish as a model. In the jellyfish, locomotion results from rhythmic contractions of the bell made possible by two layers of contractile epithelial tissues (proto muscles). For the bell to contract and exert force, which allows the jellyfish to swim upwards, the proto muscles in these sheets must contract and then relax in unison. This is achieved by a network of neurons located between two layers of epithelial cells and pattern generating neurons organized in a ring that surrounds the bell.

The neurons surrounding the bell generate a

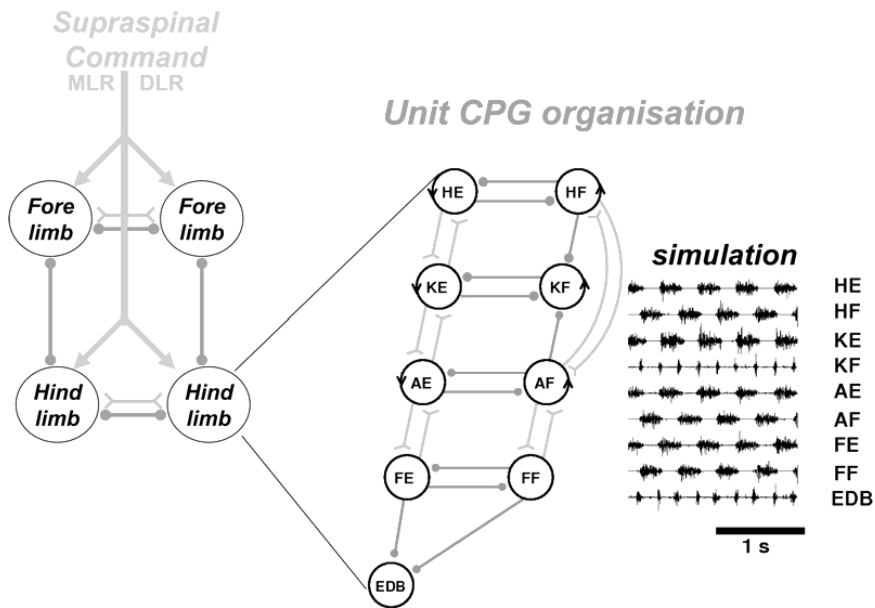


Figure 3. Interconnection of CPGs in a hindlimb of a vertebrate: hip (H), knee (K), ankle (A), and foot (F) extensors (E) or flexors (F). Extensor digitorum brevis (EDB) has a distinctive arrangement. Circles indicate inhibition, and forks/triangles excitation. Reprinted from S. GRILLNER, *Biological pattern generation: The cellular and computational logic of networks in motion*, p. 758. With permission from Elsevier.

rhythmic signal that is passed through the network, eliciting responses in all the proto-muscles at once. A similar arrangement of a nerve net controlling muscles is found in many vertebrate internal organs such as the lungs, heart, and those of the digestive tract.¹⁹ In these networks, like that in the jellyfish, the oscillation is endogenously generated in the local circuits. It can be, and is, modulated by other neural signals, but the neural system regulating these muscles is not dependent on sensory input to initiate activity. In the case of vertebrate internal organs such as the lungs, heart, and those of the digestive tract, modulatory signals are transmitted from relevant nuclei in the hypothal-

amus through the autonomic nervous system, enabling, for example, increased breathing, circulation of blood, or eating when conditions in the organism require it.

The skeletal muscles that make motor activity possible in limbed vertebrates require more fine-grained regulation than just coordinated contraction: individual skeletal muscles must contract in a specific sequence to generate the needed limb movement. Nonetheless, similar organizing processes are employed, beginning with highly local control mechanisms. Skeletal muscles typically occur in pairs: flexor muscles contract the limb while extensor muscles restore the limb. These pairs are

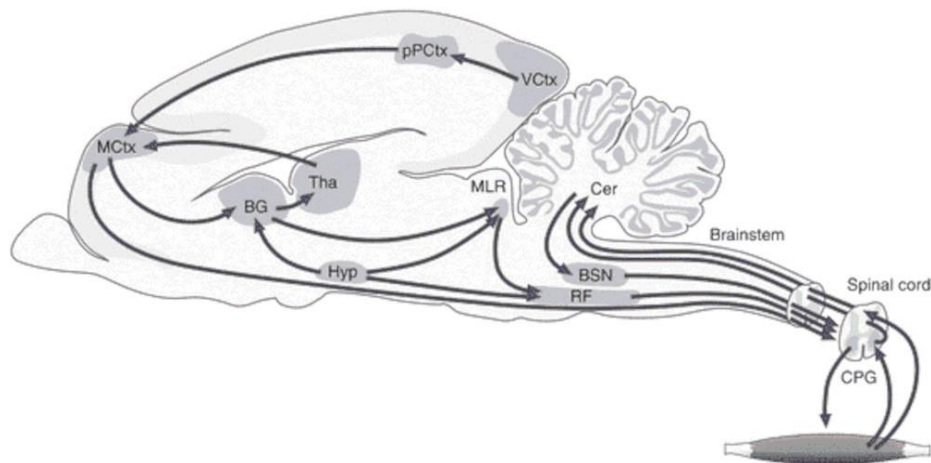


Figure 4. Major brain regions involved in controlling skeletal muscles in rodents. The most immediate control is exercised by central pattern generators (CPG) in the spinal cord, which are in turn integrated by a network in the reticular formation (RF). The mesencephalic local motor region (MLR) sends projections via the RF that direct specific actions. The MLR in turn receives inputs from both the hypothalamus (Hyp) and the basal ganglia (BG). Reprinted from O. KIEHN, K. DOUGHERTY, *Locomotion: Circuits and physiology*, p. 1212. With permission from Springer Nature.

controlled locally to produce successive contractions and relaxations by central pattern generators (CPGs). These act on motor neurons that directly elicit contraction of a specific muscle fibers. In some cases, CPGs rely on intrinsically rhythmic neurons, but in many cases the rhythmic behavior results from the interaction between pairs of mutually inhibitory neurons, as in Figure 2. When one of the excitatory interneurons (eIN) is active, it activates both its motor neuron (MN) and an inhibitory interneuron (iIN) that suppresses both the excitatory and inhibitory interneurons controlling the other muscle. As its activity wanes, the other interneuron is released from inhibition, allowing it to activate its muscle and inhibit the first excitatory neuron.²⁰

CPGs not only create rhythmic muscle contraction but serve as hubs for coordinating other inputs that affect muscle response. Locally, they receive somatosensory feedback from the muscles being controlled. This enables a CPG to, for example, register whatever resistance the muscle confronts and adjust muscle contraction in response.²¹ In limbed organisms it is necessary to coordinate the movement of each limb in the pair. Accordingly, neuronal projections between CPGs on opposite sides of the body enable such coordination. Movement of whole limbs requires precise coordination between different muscle groups, which is typically achieved via interactions between multiple CPGs (Figure 3). These individual CPGs are not as rigidly connected as this suggests; different patterns of activity in the higher-level CPG can initiate different patterns in subordinate CPGs, and their activity can be further modified by neuromodulators (see section 5).

For control of skeletal muscles, peptidergic transmission is too slow. Transmitters such a glutamate and acetylcholine that elicit fast, ionotropic, responses are employed instead. Moreover, specific connectivity patterns are required to insure muscles contract in the right sequence. Networks within the spinal cord achieve these ends.²² These networks are capable of coordinating muscle activity without any higher-level input. When Shik and Orlovsky²³ sectioned the spinal cord from the brain stem in kittens and placed them on a moving treadmill, they generated coordinated limb movements that adjusted appropriately as the speed of the treadmill was changed (e.g., transitioning between walking and running). This coordinated movement presumably reflected the coupling of individual CPGs. This is typically achieved through networks of neurons in the reticular formation that project down the reticulospinal tract. The name “reticular” means network and true to its name, the nuclei within the reticular formation form an extremely complex web. These networks are not well understood, but researchers have established that stimulating neurons in the

lateral anterior reticular formation elicits coordinated movements, with different neurons eliciting different behaviors.²⁴

As shown in Figure 4, the mesencephalic locomotor region (MLR) is a major source of input to the reticular formation.²⁵ When lesions are made above the MLR, stimulating neurons in it elicits coordinated walking or running behavior.²⁶ Roseberry and colleagues²⁷ determined that the MLR consists of multiple groups of neurons differentiated by reliance on different neurotransmitters. Serotonergic and glutamatergic neurons act to initiate muscle activity whereas GABAergic neurons slow muscle responses. Finally, stimulating cholinergic neurons modulates motor responses but does not initiate or inhibit them. The MLR thus appears as an integrating hub whose outputs result in specific motor activity and whose inputs serve to initiate, terminate, or modulate those activities.

In this section I have identified components of the architecture that regulates muscles. Critical to muscle action are the pattern generators responsible for each muscle and the network in which these are interconnected. Neurons at higher levels in the reticular formation and the MLR neurons are connected to specific CPGs and are able to elicit specific coordinated muscle contractions.

4 The basal ganglia: Selecting motor behaviors

In experimental protocols, nuclei in the MLR must be stimulated to produce motor responses. This is in significant part due to the activity of the output nuclei of the basal ganglia, the substantia nigra pars reticularis (SNr) and the globus pallidus interior (GPi). Neurons in these regions are tonically active and send inhibitory outputs not just to the MLR but to many brain regions, including those of the neocortex (Figure 5). Only when these inhibitory outputs are inhibited by processing in the basal ganglia, are neurons in the MLR (and other brain regions) released from inhibition and able to process information. The basal ganglia are thus an important control center that regulates information processing throughout the brain.

The basal ganglia are a set of interconnected nuclei located in the midbrain and base of the forebrain. They form a feedforward network from input nuclei to output nuclei. At a larger scale, the basal ganglia complete processing loops with other brain regions by sending outputs back to the same regions from which they receive inputs, either maintaining inhibition or releasing these areas from inhibition. (These loops also involve regions in the thalamus, which I will not discuss here.)

I focus on two pathways through which activity is directed through the different nuclei of the basal ganglia, one of which, the direct pathway, inhibits selected inhibitory neurons in the SNr and

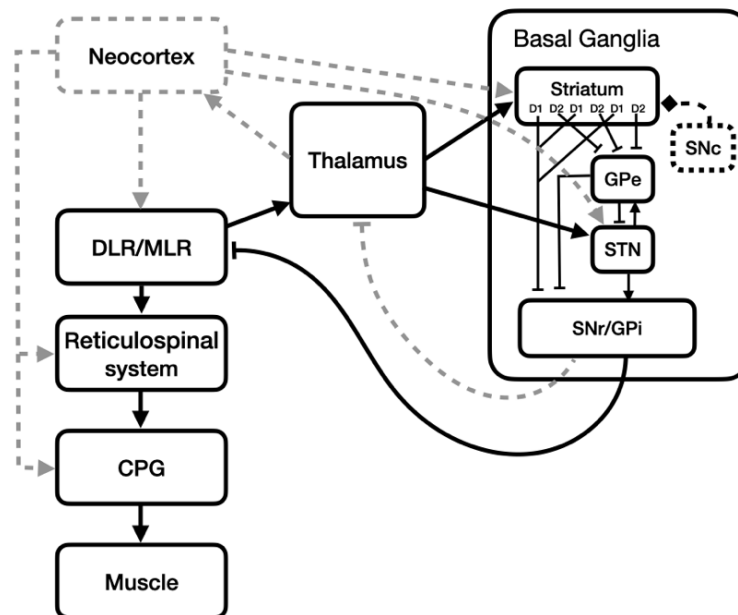


Figure 5. Organization of the Basal Ganglia and their role in control of processing in other brain areas. Components in black are active even without the neocortex, which, together with its connections to other regions, is shown in grey.

the GPi, while the other, the indirect pathway, further enhances their inhibitory outputs. The input neurons to both pathways reside in the striatum and are laid out topographically so that individual neurons receive inputs from specific regions in the brain. These input neurons also have one of two types of dopamine receptors, D1 or D2. I will return to the role of dopamine in the next section. For now, what is important is that D1 striatal neurons send their inhibitory output directly to specific SNr or GPi neurons (hence, this is known as the *direct pathway*); the result of inhibiting these neurons is to relax the inhibition of target areas. Thus, when a D1 neuron that receives input from a specific population in the MLR is activated, the inhibition of the MLR neurons is blocked and they become active and initiate muscle contraction. When D2 neurons in the striatum are activated, they send an inhibitory signal to neurons in the globus pallidus external (GPe), which in turn sends inhibitions to SNr or GPi output neurons (this is known as the *indirect pathway*). Processing in this pathway serves to reinforce the default inhibitory action of these output neurons. In the case of outputs that project to the MLR, activity in the indirect pathway serves to keep the muscles inactive.²⁸

The two pathways in the basal ganglia provide a vehicle for selecting which other brain areas are allowed to function and which are inhibited. Which output neurons are active is in part determined by the relative strength of inputs to the D1 and D2 neurons. The competition is carried out through inhibitory neurons projecting between D1 and D2 neurons. The inputs to the striatum have limited bandwidth. Accordingly, the basal ganglia do not receive the information processed in these

other areas; rather, the inputs it receives reflect the strength of the activity in the source/target regions. The basal ganglia constitute the site at which the competition between alternative neural processing is carried out. They enable the MLR neurons with the greatest activation to initiate muscle movement while those that would initiate competing muscle movements are suppressed.

The brain regions discussed so far provide an evaluation of needs of the organism and a means to select and control muscle activity. I have not considered how these regions are connected. The mediation is often accomplished through neuromodulators, to which I turn in the next section.

5 Neuromodulation: Setting agendas

The striatal neurons that are the inputs to the two pathways in the basal ganglia are differentiated by the type of dopamine receptor each possesses. Dopamine and other monoamines such as serotonin, are referred to as *neuromodulators* due to the roles they play in modulating neural circuits. Theorists sometimes assume that, as a result of a particular pattern of connections, neural circuits always process inputs in the same way. However, research on neural circuits in invertebrates challenged this view. Marder and her collaborators demonstrated that, depending on what neuromodulators are present, neurons in the lobster stomatogastric ganglion, responsible for the rhythmic contraction required for digestion of food, are organized into different coalitions realizing different behaviors.²⁹ In performing their roles as modulators, these molecules function differently than ionotropic neurotransmitters such as glu-

tamate. In many respects these neuromodulators are similar to the neuropeptides described in section 2, which are sometimes also described as neuromodulators. They are released not just at synapses but from many regions in a neuron, including dendrites, and disseminate through the extracellular matrix to any neuron with an appropriate receptor. For a given neuromodulator there are multiple different receptor types that result in different responses by these target neurons to the same transmitter release. Katz offered a broad characterization of neuromodulation: «any communication between neurons, caused by release of a chemical, that is either not fast, or not point-to-point, or not simply excitation or inhibition».³⁰

Before considering the role of neuromodulators more generally, I examine first how dopamine functions in the striatum of the basal ganglia. Dopamine is released into the dorsal striatum by neurons with cell bodies in the substantia nigra pars compactus (SNc). The D1 and D2 neurons are not only connected differently to neurons in other regions of the basal ganglia, but they respond to dopamine in opposite ways – D1 receptors respond to dopamine by bringing their resting potential closer to threshold, thereby making them more likely to generate action potentials and release their target from inhibition. D2 receptors respond in the opposite manner; in response to dopamine, they drive the resting potential further from threshold, making them less likely to respond. Dopamine thus biases the competition towards D1 neurons, making activity more likely.³¹ In addition to regulating the likelihood of response, dopamine in the striatum is thought to play at least two other roles. Phasic dopamine release is hypothesized to constitute a reward signal that alters connectivity so that D1 neurons are more likely to respond to the same input in the future. When dopamine is increased tonically, on the other hand, it is hypothesized to enable activity in the circuit that is currently dominant, curtailing exploration of alternatives.³²

Dopamine clearly modulates the activity of striatal neurons. Although the term *neuromodulator* captures this, it understates the systemic importance of neuromodulators. By determining how target circuits process information, neuromodulators configure the information-processing capacities of organisms. Accordingly, they might better be understood as setting the information-processing agenda. This role has been most clearly demonstrated in invertebrates. In the medicinal leech, serotonin increases activity in circuits leading to swimming behavior, whereas dopamine enhances activity in circuits supporting crawling.³³ There are receptors for dopamine, serotonin, and other neuromodulators, throughout vertebrate brains, including in the neocortex; they thus are situated in locations in which they are able to set

the agenda for neural processing. Adopting such a broad perspective on the roles of dopamine on behavior, and tracing its role across phylogeny, Hills and his collaborators³⁴ have interpreted dopamine as functioning, across phylogeny, to promote search activities (search in new physical as well as in conceptual spaces).

While having widespread effects, neurotransmitters such as dopamine and serotonin are synthesized by neurons in only a few nuclei in the brain. In addition to the SNc, dopaminergic neurons are found in the ventral tegmental area (VTA) and several regions of the hypothalamus, including the arcuate nucleus discussed above. In the brain, serotonin is synthesized primarily in the nine raphe nuclei (it is also synthesized within the digestive tract). These various centers receive inputs from many other brain regions, including areas of the hypothalamus. Their axons then extend widely through the brain, from which the neuromodulator diffuses even more broadly. For example, dopaminergic neurons in the VTA project to the ventral striatum, nucleus accumbens, the amygdala, the hippocampus, the olfactory bulb, as well as the prefrontal cortex.³⁵ Neurons in the more rostral raphe nuclei project to areas in the midbrain and forebrain, while those in the more caudal raphe nuclei project towards the brain stem and spinal cord (areas directly involved in initiating motor movements). The ways in which neuromodulators are synthesized and distributed, and their effects in determining processing in targeted neural circuits, enables them to set the information-processing agenda for much of the brain.

6 From the tectum to the neocortex: Coordinating sensory information

So far I have characterized how subcortical areas support assessing the condition of the organism, provide for and select between a range of motoric responses, and set the agenda for both neural activity and behavior through neuromodulators. With these resources, organisms are equipped to perform actions in the world in response to their current state. But for many actions to be successful the organism needs information about the layout of the world around it. Except for feedback to CPGs, I have ignored the neural processes that acquire this information. That is purposeful – when sensory processing is viewed as the starting point for cognitive processing that leads to action, it leaves fundamental questions unanswered: what moves organisms to act? and to act in one way rather than another? On the account I have advanced, processing components such as the hypothalamus and nuclei that synthesize neuromodulators motivate action. Sensory information is needed not to initiate action. Often its role is to refine already initiated subcortical processing directed at

action as needed given what is available in the environment and how the organism is situated in its environment.³⁶ Identifying objects in the environment, their location, and the activities they afford are major functions of the neocortex. As I noted earlier, however, mammals like cats can live autonomously without a neocortex. They must still acquire and process sensory information and identify affordances. They do so using the optic tectum (commonly referred to as the superior colliculus in mammals).

The primary inputs to the optic tectum are from the retina. The tectum is organized in anatomical layers, each of which represents features of the environment in a topographic map constituting the animal's subjective perspective of the space in which it resides. These maps are aligned, but neurons in more superficial layers of the tectum have smaller receptive fields and respond solely to visual inputs, while those in deeper layers have larger receptive fields and integrate that information with other sensory information (auditory, somatosensory) and inputs from motor systems representing impending movements. Neurons in the tectum also receive inputs about the current needs of the organism, from which they can generate an assessment of the importance of different stimuli. The optic tectum is the hub of the midbrain stimulus selection network. Through its connections with other components of the network, it selects the identity and location of the stimulus of the greatest momentary importance.

The optic tectum sends outputs both to motor areas and to the thalamus and indirectly to the neocortex. The motor outputs suffice both to orient the organism with respect to selected stimuli and to engage in feeding or defensive activities, including fleeing.³⁷ These are presumably what enable decorticated mammals to respond to visual and other sensory input.³⁸ The pathways from the tectum to the neocortex suggests that once the neocortex evolved, the tectum and neocortex could operate together. The neocortex is in an important respect organized in the same manner as the optic tectum with successive maps of stimulus space, each with larger receptive fields, aligned with each other. Instead of being stacked on top of each other, they are laid out as different cortical areas. The maps in the neocortex are more expansive than those in the tectum, allowing for greater processing capacity. In addition, through interactions between the thalamus and regions of neocortex, different areas can be invoked when their information-processing capacities are relevant.

In neural network modeling, the neocortex is treated as a single, comprehensive processing system that processes sensory inputs to arrive at outputs that categorize information or, if models include frontal areas of the brain, direct motor activity. This corresponds to how, in the actual neocor-

tex, visual stimuli are first processed in V1, then in two different streams, one emphasizing identification of objects, the other location in space and potential for action.³⁹ As insightful as this approach is to understanding processing of visual stimuli, it fails to take into account how each of the regions of the neocortex is integrated with other brain regions. This interconnection is seen in the loops through which areas of the neocortex are connected to the basal ganglia and thalamus as well as by the projection of neuromodulators to various areas in the neocortex. This suggests that, in addition to the forward and recurrent projections between cortical areas, each area is coordinated in its own way with subcortical areas. Like the tectum, it carries out this processing through interconnections with subcortical regions, including those discussed in earlier sections. These processes can only be understood by addressing the interconnection of cortical areas with subcortical ones. From such a perspective, one can appreciate both how specialized processing in neocortical areas contributes to processing elsewhere in the brain and how processing in those areas affects processing in the neocortex.

While I am emphasizing subcortical information processing and its integration with cortical processing, it is important to note that the architecture of the neocortex allows for distinctive information processing capacities. The ability to categorize objects and apply knowledge acquired about categories to individual instances is extremely important in human life and appears to be a distinctive capacity of the neocortex. A further distinctive ability is to learn natural languages and deploy them in reasoning. This enables coordinated cognitive investigations between humans, such as scientific inquiries that exceed what members of other species can accomplish. It also enables humans to think about their own thinking, engaging in what Karmiloff-Smith⁴⁰ describes as *representational redescription* and is more generally characterized as *metacognition*. One capacity this makes possible is for humans to construct narratives about themselves and invoke such narratives in regulating their lives.⁴¹ Recognizing and investigating the types of cognition the neocortex makes possible, however, does not require viewing it as operating autonomously from the rest of the brain. Rather, once one recognizes the role processing in other brain areas figures in identifying states of the organism (e.g., hunger, thirst, fear, etc.) and coordinating behavioral activities, one can explore how higher-level cognitive activities coordinate with these. Moreover, insofar as a great deal of relevant information processing is carried out by other brain structures, one can look for and localize in the neocortex the specialized information processing it performs while localizing other information processing that contributes to overall cognitive activities elsewhere in the brain.

7 Conclusion: Vertebrate cognition employs a heterarchy of different cognitive architectures

Once we recognize that there are many different information processing systems in the vertebrate brain and that their different activities are integrated in regulating behavior, we should rethink the types of cognitive architectures invoked in cognitive science. Instead of identifying a single architecture, modeled on the neocortex, we should seek to understand the different architectures that process information in the vertebrate brain and how they are coupled together.

I started with the subcortical processes that evaluate an organism's state and that direct muscle action. These employ different architectures for information processing. The hypothalamus extends the chemical processing architecture of the endocrine system and brainstem areas such as the nucleus of the solitary tract, providing a network of hubs that integrate signals specifying different conditions of the organism that may require responses. CPGs and the MLR are configured in a different manner, one suited to generate a temporal ordering of processing both in individual muscles and between muscles. The basal ganglia provide yet another architecture, one suited to carrying out a competition in which circuits with the strongest activation are released from inhibition and allowed to proceed with processing information. Attending to the role of dopamine receptors in the striatum of the basal ganglia reveals the more general role neuromodulators play in determining how different brain areas work on different occasions.

Only after introducing these regions, I turned to the information processing of sensory stimuli. In early vertebrates, this involved the optic tectum/superior colliculus mapping stimulation from the environment onto motor responses. In doing this, the tectum both receives inputs and sends outputs to these other subcortical areas. In higher mammals the neocortex takes on much of analysis of sensory inputs, but it contributes in much the same manner as the optic tectum, supplying information about the layout of the environment and objects in it that can support activity often initiated by subcortical components.

Accounts of cognitive architectures, whether symbolic or neural network, have generally focused on the resources thought to be needed to perform high-level information processing such as abstract reasoning and problem solving. These accounts support the view that cognition is simply engaged in these high-level, abstractly specified tasks. The subcortical components on which I have focused support a different perspective – the main task for which information processing is required is the maintenance of the organism. This

requires assessing the state of the organism, selecting between possible actions, and coordinating muscles so as to carry out selected actions. The processing of sensory information in the neocortex allows for abstract reasoning that exceeds the immediate needs of the organism. Even that activity, however, is carried out in an integrated system in which, for example, decisions as to what information to consider are made using the basal ganglia, and the neural processes are themselves subject to agendas set by neuromodulators.

Having identified multiple information-processing architectures in the brain, a major question emerges: how are these different systems organized so as to work together? In designing systems in which multiple components need to work together effectively, theorists often turn to hierarchical designs in which multiple lower-level units report to a single higher-level unit, and ultimately one top-level unit oversees the whole system. The different units are organized into a pyramid. Hierarchical organization is often implemented in social systems and in human-made artifacts. It is common to characterize the brain as similarly hierarchically organized, with sensory areas processing input information, passing it to higher processing centers, ultimately reaching a central executive thought to be located in the prefrontal cortex. Once the prefrontal cortex arrives at a plan of action, commands are directed down the hierarchy until they are implemented by muscles. On this scheme, the information processing architectures I have identified are ultimately subordinate to and controlled by the neocortex, and ultimately by the prefrontal cortex.

Hierarchy is not the only possible way of organizing component systems. I use the term *heterarchy* to describe organization that violates one or more features of hierarchical organization.⁴² For purposes here, the crucial departure from hierarchy is that different units are not subordinate to a common higher-level unit. Rather than forming a hierarchical pyramid, the components constitute a network in which different components, operating by their own principles, carry out their activity while also influencing each other.⁴³ Coherent behavior of the whole organism results not from a central executive making all the decisions, but from coordination between the components. The hypothalamus, among other areas, registers needs of the organism, regions releasing neuromodulators such as serotonin and dopamine set the agendas for other areas, the basal ganglia allow for decisions between alternatives, and motor areas determine how actions are carried out. The neocortex carries out specific processing, especially of stimuli, and constructs memories and accesses stored information, but these processes are not in charge of other processing. Rather, they participate in the network in which behavioral is ultimately generated.

One reason that theorists often default to assuming hierarchical organization is that they assume that allowing components each to operate according to its own procedures will lead to conflict and overall dysfunction. It is certainly the case that non-hierarchical systems can engage in conflict and even be dysfunctional, although hierarchical systems can also be dysfunctional. One factor that can mitigate concerns about dysfunctionality is that the organization of brain networks is a product of evolution. Only those organisms in which brain networks enabled the organisms to maintain themselves in their environments long enough to reproduce were inherited. Considering evolution provides a further reason to expect heterarchical organization. The basic organization of the vertebrate brain was in place in the last common ancestor of all of today's vertebrates. What has evolved since then are a large number of variants on this architecture, each deriving from another by "tinkering" with earlier designs.⁴⁴ A process of tinkering with extant designs is not likely to generate hierarchy and may undercut existing hierarchical design. If developed gradually over a prolonged period of evolution, however, it may produce a highly robust information-processing system.⁴⁵

The structure that has undergone the most change in the vertebrate lineage is the pallium, a three-layer structure that expanded into the six-layer neocortex, which has in turn diversified into a host of different regions. As I have emphasized throughout, each of these has maintained interactions with the thalamus, basal ganglia, and other subcortical structures. It is common to construe neocortical areas as processing informationally sequentially (e.g., from primary visual cortex along streams in the inferotemporal cortex and the parietal cortex). These streams continue into the prefrontal cortex, with inferotemporal areas supplying inputs to areas involved in evaluation and parietal areas supplying inputs to areas engaged in associations between stimuli and actions and pragmatic, social, and moral norms affecting action.⁴⁶ But, as Felleman and van Essen showed,⁴⁷ there are also many interconnections between the streams and there are typically more recurrent than forward projections. Recent research has revealed complex dynamical behavior within the cortex such as traveling waves that affect how information is processed in specific brain regions.⁴⁸ These and other cortical oscillations are partly driven by interactions between cortical regions and regions of the thalamus. These findings all point to a heterarchical organization involving multiple interacting components arrived at via tinkering with earlier designs.

Approaching information processing in the brain in the way I have builds on the view that cognition is in the service of the organism, regulating activities it needs to perform to maintain itself.

From this perspective, the forms of information processing that have been the focus of cognitive science – categorization, language production and comprehension, memory encoding and retrieval, problem solving, metacognitive processing, etc. – remain important. The neocortex plays a critical role in these processes. But these processes are integrated into a larger system, one that is more directly tied to the maintenance of the organism. Higher cognitive processes do not exhaust the information processing vertebrates employ to control their behavior. And they are not carried out independently of these other information processing activities that employ brain structures exhibiting different architectures. To understand the different types of information processing that vertebrates perform, cognitive science needs to characterize the diverse information-processing architectures found in the brain and investigate how they coordinate with each other in generating behavior.

Notes

¹ Cf. A. NEWELL, *Physical symbol systems*. In: «Cognitive Science», vol. IV, n. 2, 1980, pp. 135-183; A. NEWELL, *Unified theories of cognition*, Harvard University Press, Cambridge (MA) 1990.

² Cf. D.E. RUMELHART, J.L. MCCLELLAND (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. I, Foundations*.

³ Cf. D.L.K. YAMINS, J.J. DI CARLO, *Using goal-driven deep learning models to understand sensory cortex*; C. ZHUANG, S. YAN, A. NAYEBI, M. SCHRIMPF, M.C. FRANK, J.J. DI CARLO, D.L.K. YAMINS, *Unsupervised neural network models of the ventral visual stream*.

⁴ Cf. H.R. MATURANA, F.J. VARELA, *Autopoiesis and cognition: The realization of the living*.

⁵ Cf. W. BECHTEL, L. BICH, *Grounding cognition: Heterarchical control mechanisms in biology*.

⁶ Cf. P. LYON, F. KEIJZER, D. ARENDT, M. LEVIN, *Reframing cognition: Getting down to biological basics*.

⁷ Cf. S. GRILLNER, *Evolution of the vertebrate motor system - from forebrain to spinal cord*.

⁸ Cf. L.M. BJURSTEN, K. NORRSELL, U. NORRSELL, *Behavioural repertory of cats without cerebral cortex from infancy*.

⁹ Cf. L.T.-L. HUANG, L. BICH, W. BECHTEL, *Model organisms for studying decision-making: A phylogenetically expanded perspective*.

¹⁰ Cf. J.A. FODOR, *The language of thought*.

¹¹ Cf. A. NEWELL, H.A. SIMON, *Human problem solving*.

¹² Cf. D.C. DENNETT, *Real patterns*.

¹³ Cf. B. KLEINE, W.G. ROSSMANITH, *Hormones and the endocrine system: Textbook of endocrinology*.

¹⁴ For a historical review of the discovery and research on leptin, cf. D.L. COLEMAN, *A historical perspective on leptin*. Leptin also acts as an endocrine, producing effects on blood glucose levels similar to those of insulin (CF. A.A. DA SILVA, J.M. DO CARMO, J.E. HALL, *CNS regulation of glucose homeostasis: Role of the leptin-melanocortin system*).

¹⁵ Cf. T. SAKURAI, A. AMEMIYA, M. ISHII, I. MATSUZAKI, R.M. CHEMELLI, H. TANAKA, S.C. WILLIAMS, J.A. RICH-

ARDSON, G.P. KOZLOWSKI, S. WILSON, J.R. ARCH, R.E. BUCKINGHAM, A.C. HAYNES, S.A. CARR, R.S. ANNAN, D.E. McNULTY, W.S. LIU, J.A. TERRETT, N.A. HILSHOURBAGY, D.J. BERGSMA, M. YANAGISAWA, *Orexins and orexin receptors: A family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior.*

¹⁶ Cf. L. DE LECEA, T.S. KILDUFF, C. PEYRON, X.B. GAO, P.E. FOYE, P.E. DANIELSON, C. FUKUHARA, E.L.F. BATTENBERG, V.T. GAUTVIK, F.S. BARTLETT II, W.N. FRANKEL, A.N. VAN DEN POL, F.E. BLOOM, K.M. GAUTVIK, J.G. SUTCLIFFE, *The hypocretins: Hypothalamus-specific peptides with neuroexcitatory activity.*

¹⁷ Cf. C.B. SAPER, B.B. LOWELL, *The hypothalamus.*

¹⁸ Cf. F. KEIJZER, M. VAN DUIJN, P. LYON, *What nervous systems do: Early evolution, input-output, and the skin brain thesis.*

¹⁹ Keijzer and colleagues focus on the locomotive activities of jellyfish. But other theorists, focusing on Hydra, which belong to the same phylum as jellyfish, *Cnidaria*, emphasize digestive activities and interpret the primary function of the nerve net as directing nutrients through the digestive system of the hydra much as the enteric nervous system does in us. As a result, Furness and Stebbing argue that the enteric nervous system, often viewed as a second brain, is in fact the descendent of the first brain (cf. J.B. FURNESS, M.J. STEBBING, *The first brain: Species comparisons and evolutionary implications for the enteric and central nervous systems*).

²⁰ Cf. D. BUCHER, G. HASPEL, J. GOLOWASCH, F. NADIM, *Central pattern generators.*

²¹ Cf. I. STEUER, P.A. GUERTIN, *Central pattern generators in the brainstem and spinal cord: an overview of basic principles, similarities and differences.*

²² Cf. O. KIEHN, *Decoding the organization of spinal circuits that control locomotion.*

²³ Cf. M.L. SHIK, G.N. ORLOVSKY, *Neurophysiology of locomotor automatism.*

²⁴ Cf. R.M. BROWNSTONE, J.W. CHOPEK, *Reticulospinal systems for tuning motor commands.*

²⁵ Cf. U. SŁAWIŃSKA, L.M. JORDAN, *Serotonergic influences on locomotor circuits.*

²⁶ Cf. I. STEUER, P.A. GUERTIN, *Central pattern generators in the brainstem and spinal cord: An overview of basic principles, similarities and differences.*

²⁷ Cf. T.K. ROSEBERRY, A.M. LEE, A.L. LALIVE, L. WILBRECHT, A. BONCI, A.C. KREITZER, *Cell-type-specific control of brainstem locomotor circuits by basal ganglia.*

²⁸ Cf. T.K. ROSEBERRY, A.M. LEE, A.L. LALIVE, L. WILBRECHT, A. BONCI, A.C. KREITZER, *Cell-type-specific control of brainstem locomotor circuits by basal ganglia.*

²⁹ For a succinct review, cf. E. MARDER, *Neuromodulation of neuronal circuits: Back to the future.*

³⁰ P.S. KATZ, *What are we talking about? Modes of neuronal communication.* In: P.S. KATZ (ed.), *Beyond neurotransmission: Neuromodulation and its importance for information processing*, p. 3.

³¹ Cf. T.K. ROSEBERRY, A.M. LEE, A.L. LALIVE, L. WILBRECHT, A. BONCI, A.C. KREITZER, *Cell-type-specific control of brainstem locomotor circuits by basal ganglia.*

³² Cf. T.T. HILLS, P.M. TODD, D. LAZER, A.D. REDISH, I.D. COUZIN, THE COGNITIVE SEARCH RESEARCH GROUP, *Exploration versus exploitation in space, mind, and society.*

³³ Cf. J.G. PUHL, K.A. MESCE, *Dopamine activates the motor pattern for crawling in the medicinal leech.*

³⁴ Cf. T.T. HILLS, P.M. TODD, R.L. GOLDSTONE, *Search*

in external and internal spaces: Evidence for generalized cognitive search processes; T.T. HILLS, P.M. TODD, D. LAZER, A.D. REDISH, I.D. COUZIN, THE COGNITIVE SEARCH RESEARCH GROUP, *Exploration versus exploitation in space, mind, and society.*

³⁵ Dopaminergic neurons from the arcuate nucleus as well as the periventricular nucleus of the hypothalamus project to the pituitary and act to inhibit the secretion of prolactin, a protein that enables milk production.

³⁶ This is not to deny that sensory signals sometimes initiate action. An anonymous reviewer pointed to the role of sensory inputs generating aesthetic evaluation. So far most research on neural processing of aesthetic stimuli has focused on areas of the neocortex, although there has been some research addressing the role of dopamine processing in the striatum (cf. S. LACEY, H. HAGTVEDT, V.M. PATRICK, A. ANDERSON, R. STILLA, G. DESHPANDE, X. HU, J.R. SATO, S. REDDY, K. SATHIAN, *Art for reward's sake: Visual art recruits the ventral striatum*).

³⁷ Cf. E. COMOLI, P. DAS NEVES FAVARO, N. VAUTRELLE, M. LERICHE, P.G. OVERTON, P. REDGRAVE, *Segregated anatomical input to sub-regions of the rodent superior colliculus associated with approach and defense.*

³⁸ Cf. B.E. STEIN, M.A. MEREDITH, *The merging of the senses.*

³⁹ Cf. D.C. VAN ESSEN, J.L. GALLANT, *Neural mechanisms of form and motion processing in the primate visual system.*

⁴⁰ Cf. A. KARMILOFF-SMITH, *Beyond modularity.*

⁴¹ Cf. W. BECHTEL, *Mental mechanisms. Philosophical perspectives on cognitive neuroscience*, pp. 239-268.

⁴² The term was introduced by McCulloch (cf. W.S. MCCULLOCH, *A heterarchy of values determined by the topology of nervous nets*) for preferences that form circles (A is preferred to B, B to C, and C to A). Pattee extended the concept to control mechanisms when they deviate from a hierarchy (cf. H.H. PATTEE, *Measurement-control heterarchical networks in living systems*). Heterarchical organization can differ from hierarchy in multiple ways – there can be more units at higher levels than at lower levels, inverting the pyramid, and when units exercise control over others, multiple ones may control the same component and controlled components can, in some respects, exercise control over those taken to be in control (cf. W. BECHTEL, *Levels in biological organisms: Hierarchy of production mechanisms, heterarchy of control mechanisms*).

⁴³ Cf. W. BECHTEL, *Levels in biological organisms: Hierarchy of production mechanisms, heterarchy of control mechanisms.*

⁴⁴ Cf. F. JACOB, *Evolution and tinkering.*

⁴⁵ Cf. L. BICH, W. BECHTEL, *Control mechanisms: Explaining the integration and versatility of biological organisms.*

⁴⁶ Cf. R.W. CARLSON, M.J. CROCKETT, *The lateral prefrontal cortex and moral goal pursuit.*

⁴⁷ Cf. D.J. FELLEMAN, D.C. VAN ESSEN, *Distributed hierarchical processing in the primate cerebral cortex.*

⁴⁸ Cf. Z.W. DAVIS, L. MULLER, J. MARTINEZ-TRUJILLO, T. SEJNOWSKI, J.H. REYNOLDS, *Spontaneous travelling cortical waves gate perception in behaving primates.*

Literature

ARRIGONI, E., CHEE, M.J.S., FULLER, P.M. (2019). *To eat or to sleep: That is a lateral hypothalamic question.*

- In: «Neuropharmacology», vol. CLIV, pp. 34-49.
- BECHTEL, W. (2008). *Mental mechanisms. Philosophical perspectives on cognitive neuroscience*, Routledge, London.
- BECHTEL, W. (2022) *Levels in biological organisms: Hierarchy of production mechanisms, heterarchy of control mechanisms*. In: «The Monist», vol. CV, n. 2, pp. 156-174.
- BECHTEL, W., BICH, L. (2021). *Grounding cognition: Heterarchical control mechanisms in biology*. In: «Philosophical Transactions of the Royal Society of London: B – Biological Sciences», vol. CCCLXXVI, n. 1820, Art. Nr. 20190751 – doi: 10.1098/rstb.2019.0751.
- BICH, L., BECHTEL, W. (2022). *Control mechanisms: Explaining the integration and versatility of biological organisms*. In: «Adaptive Behavior», vol. XXX, n. 5, pp. 389-407.
- BJURSTEN, L.M., NORRSELL, K., NORRSELL, U. (1976). *Behavioural repertory of cats without cerebral cortex from infancy*. In: «Experimental Brain Research», vol. XXV, n. 2, pp. 115-130.
- BROWNSTONE, R.M., CHOPEK, J.W. (2018). *Reticulospinal systems for tuning motor commands*. In: «Frontiers in Neural Circuits», vol. XII, art. nr. 30 – doi: 10.3389/fncir.2018.00030.
- BUCHER, D., HASPEL, G., GOLOWASCH, J., NADIM, F. (2015). *Central pattern generators*. In: «eLS» - doi: 10.1002/9780470015902/a0000032.pub2.
- CARLSON, R.W., CROCKETT, M.J. (2018). *The lateral prefrontal cortex and moral goal pursuit*. In: «Current Opinion in Psychology», vol. XXIV, pp. 77-82.
- COLEMAN, D.L. (2010). *A historical perspective on leptin*. In: «Nature Medicine», vol. XVI, n. 10, pp. 1097-1099.
- COMOLI, E., DAS NEVES FAVARO, P., VAUTRELLE, N., LERICHE, M., OVERTON, P.G., REDGRAVE, P. (2012). *Segregated anatomical input to sub-regions of the rodent superior colliculus associated with approach and defense*. In: «Frontiers in Neuroanatomy», vol. VI, Art.Nr. 9 – doi: 10.3389/fnana.2012.00009.
- DAVIS, Z.W., MULLER, L., MARTINEZ-TRUJILLO, J., SEJNOWSKI, T., REYNOLDS, J.H. (2020). *Spontaneous travelling cortical waves gate perception in behaving primates*. In: «Nature», vol. DLXXXVII, n. 7834, pp. 432-436.
- DE LECEA, L., KILDUFF, T.S., PEYRON, C., GAO, X.B., FOYE, P.E., DANIELSON, P.E., FUKUHARA, C., BATTENBERG, E.L.F., GAUTVIK, V.T., BARTLETT II, F.S., FRANKEL, W.N., VAN DEN POL, A.N., BLOOM, F.E., GAUTVIK, K.M., SUTCLIFFE, J.G. (1998). *The hypocretins: Hypothalamus-specific peptides with neuroexcitatory activity*. In: «Proceedings of the National Academy of Sciences of the United States of America», vol. XCV, n. 1, pp. 322-327.
- DENNETT, D.C. (1991). *Real patterns*. In: «The Journal of Philosophy», vol. LXXXVIII, n. 1, pp. 27-51.
- FELLEMAN, D.J., VAN ESSEN, D.C. (1991). *Distributed hierarchical processing in the primate cerebral cortex*. In: «Cerebral Cortex», vol. I, n. 1, pp. 1-47.
- FODOR, J.A. (1975). *The language of thought*, Crowell, New York.
- FURNESS, J.B., STEBBING, M.J. (2018). *The first brain: Species comparisons and evolutionary implications for the enteric and central nervous systems*. In: «Neurogastroenterology and Motility», vol. XXX, n. 2, Art.Nr. e13234 - doi: 10.1111/nmo.13234.
- GRILLNER, S. (2021). *Evolution of the vertebrate motor system - from forebrain to spinal cord*. In: «Current Opinion in Neurobiology», vol. LXXI, pp. 11-18.
- GRILLNER, S. (2006). *Biological pattern generation: The cellular and computational logic of networks in motion*. In: «Neuron», vol. LII, n. 5, pp. 751-766.
- HILLS, T.T., TODD, P.M., GOLDSTONE, R.L. (2008). *Search in external and internal spaces: Evidence for generalized cognitive search processes*. In: «Psychological Science», vol. XIX, n. 8, pp. 802-808.
- HILLS, T.T., TODD, P.M., LAZER, D., REDISH, A.D., COUZIN, I.D., THE COGNITIVE SEARCH RESEARCH GROUP (2015). *Exploration versus exploitation in space, mind, and society*. In: «Trends in Cognitive Science», vol. XIX, n. 1, pp. 46-54.
- HUANG, L.T.-L., BICH, L.L., BECHTEL, W. (2021). *Model organisms for studying decision-making: A phylogenetically expanded perspective*. In: «Philosophy of Science», vol. LXXXVIII, n. 5, pp. 1055-1066.
- JACOB, J. (1977). *Evolution and tinkering*. In: «Science», vol. CXCVI, n. 4295, pp. 1161-1166.
- KARMILOFF-SMITH, A. (1992). *Beyond modularity*, MIT Press, Cambridge (MA).
- KATZ, P.S. (1999). *What are we talking about? Modes of neuronal communication*. In: P.S. KATZ (ed.), *Beyond neurotransmission: Neuromodulation and its importance for information processing*, Oxford University Press, Oxford/New York, pp. 1-28.
- KEIJZER, F., VAN DUIJN, M., LYON, P. (2013). *What nervous systems do: Early evolution, input-output, and the skin brain thesis*. In: «Adaptive Behavior», vol. XXI, n. 2, pp. 67-85.
- KIEHN, O. (2016). *Decoding the organization of spinal circuits that control locomotion*. In: «Nature Reviews Neuroscience», vol. XVII, n. 4, pp. 224-238.
- KIEHN, O., DOUGHERTY, K. (2013). *Locomotion: Circuits and physiology*. In: D.W. PFAFF (ed.), *Neuroscience in the 21st century: From basic to clinical*, Springer, Berlin/New York/Heidelberg, pp. 1209-1236.
- KLEINE, B., ROSSMANITH, W.G. (2016). *Hormones and the endocrine system: Textbook of endocrinology*, Springer, Cham.
- LYON, P., KEIJZER, F., ARENDT, D., LEVIN, M. (2021). *Reframing cognition: Getting down to biological basics*. In: «Philosophical Transactions of the Royal Society of London: B – Biological Sciences», vol. CCCLXXVI, n. 1820, Art. Nr. 20190750 – doi: 10.1098/rstb.2019.0750.
- MARDER, E. (2012). *Neuromodulation of neuronal circuits: Back to the future*. In: «Neuron», vol. LXXVI, n. 1, pp. 1-11.
- MATURANA, H.R., VARELA, F.J. (1980). *Autopoiesis and cognition: The realization of the living*, Reidel, Dordrecht.
- MCCULLOCH, W.S. (1945). *A heterarchy of values determined by the topology of nervous nets*. In: «The Bulletin of Mathematical Biophysics», vol. VII, n. 2, pp. 89-93.
- NEWELL, A. (1980). *Physical symbol systems*. In: «Cognitive Science», vol. IV, n. 2, pp. 135-183.
- NEWELL, A. (1990). *Unified theories of cognition*, Harvard University Press, Cambridge (MA) 1990.
- NEWELL, A., SIMON, H.A. (1972). *Human problem solving*, Prentice-Hall, Englewood Cliffs (NJ).
- PATTEE, H.H. (1991). *Measurement-control heterarchical networks in living systems*. In: «International Journal of

- General Systems», vol. XVIII, n. 3, pp. 213-221.
- PUHL, J.G., MESCE, K.A. (2008). *Dopamine activates the motor pattern for crawling in the medicinal leech*. In: «Journal of Neuroscience», vol. XXVIII, n. 16, pp. 4192-4200.
- ROSEBERRY, T.K., LEE, A.M., LALIVE, A.L., WILBRECHT, L., BONCI, A., KREITZER, A.C. (2016). *Cell-type-specific control of brainstem locomotor circuits by basal ganglia*. In: «Cell», vol. CLXIV, n. 3, pp. 526-537.
- RUMELHART, D.E., MCCLELLAND, J.L. (eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. I: Foundations*, MIT Press, Cambridge (MA).
- SAKURAI, T., AMEMIYA, A., ISHII, M., MATSUZAKI, I., CHEMELLI, R.M., TANAKA, H., WILLIAMS, S.C., RICHARDSON, J.A., KOZLOWSKI, G.P., WILSON, S., ARCH, J.R., BUCKINGHAM, R.E., HAYNES, A.C., CARR, S.A., ANNAN, R.S., MCNULTY, D.E., LIU, W.S., TERRETT, J.A., HLSHOUBAGY, N.A., BERGSMAN, D.J., YANAGISAWA, M. (1998). *Orexins and orexin receptors: A family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior*. In: «Cell», XCII, n. 4, pp. 573-585.
- SAPER, C.B., LOWELL, B.B. (2014). *The hypothalamus*. In: «Current Biology», vol. XXIV, n. 23, pp. R1111-R1116.
- SHIK, M.L., ORLOVSKY, G.N. (1976). *Neurophysiology of locomotor automatism*. In: «Physiological Reviews», vol. LVI, n. 3, pp. 465-501.
- SŁAWIŃSKA, U., JORDAN, L.M. (2019). *Serotonergic influences on locomotor circuits*. In: «Current Opinion in Physiology», vol. VIII, pp. 63-69.
- STEIN, B.E., MEREDITH, M.A. (1993). *The merging of the senses*, MIT Press, Cambridge (MA).
- STEUER, I., GUERTIN, P.A. (2019). *Central pattern generators in the brainstem and spinal cord: An overview of basic principles, similarities and differences*. In: «Reviews in the Neurosciences», vol. XXX, n. 2, pp. 107-164.
- VAN ESSEN, D.C., GALLANT, J.L. (1994). *Neural mechanisms of form and motion processing in the primate visual system*. In: «Neuron», vol. XIII, n. 1, pp. 1-10.
- YAMINS, D.L.K., DI CARLO, J.J. (2016). *Using goal-driven deep learning models to understand sensory cortex*. In: «Nature Neuroscience», vol. XIX, n. 3, pp. 356-365.
- ZHUANG, C., YAN, S., NAYEBI, A., SCHRIMPF, M., FRANK, M.C., DI CARLO, J.J., YAMINS, D.L.K. (2021). *Unsupervised neural network models of the ventral visual stream*. In: «Proceedings of the National Academy of Sciences», vol. CXVIII, n. 3 - doi:10.1073/pnas.2014196118.

RICERCHE

Distributed loci of control: Overcoming stale dichotomies in biology and cognitive science

Daniel C. Burnston^(a) & Antonella Tramacere^{(b), (c)}

Ricevuto: 23 marzo 2022; accettato: 12 gennaio 2023

Abstract We argue that theoretical debates in biology and cognitive science often are based around differences in the posited *locus of control* for biological and cognitive phenomena. Internalists about locus of control posit that specific causal control over the phenomenon is exerted by factors internal (to the relevant subsystem) of an organism. Externalists posit that causally specific influence is due to external factors. In theoretical biology, we suggest, a minimal agreement has developed that the locus of control for heritable variation is *distributed* – that is, both internal and external factors exert specific, non-redundant causal influence on evolved traits. We suggest that debates in cognitive science, particularly surrounding “enactivism”, should also embrace a distributed locus of control. We show how both internal and external factors contribute non-redundantly to psychological capacities and behavior. We further suggest that embracing a distributed locus of control provides a basis for a revisionary, but substantive account of “mental representation”.

KEYWORDS: Locus of control; Internalism/Externalism; Evolutionary Biology; Cognitive Science; Mental Representation

Riassunto *Loci di controllo distribuiti: superare le vecchie dicotomie nella biologia e nella scienza cognitiva* – Numerose discussioni in biologia e nella scienza cognitiva vertono spesso su differenze nella individuazione del *locus di controllo* dei fenomeni biologici e cognitivi. Posizioni internaliste rispetto al *locus di controllo* ritengono che il controllo causale specifico su un fenomeno biologico o cognitivo venga esercitato da fattori interni a (un sottosistema rilevante) di un organismo. Al contrario, posizioni esternaliste assumono che specifiche influenze causali siano dovute a fattori esterni. In questo articolo mostriamo che nell’ambito della biologia teorica si è giunti a un minimo accordo rispetto al fatto che il *locus di controllo* per l’emergere di variazioni ereditarie è *distribuito* – ossia dovuto sia a fattori interni che esterni all’organismo che esercitano influenze specifiche e non-ridondanti sui tratti che si sono evoluti. Riteniamo che tale accordo debba essere raggiunto anche nell’ambito della scienza cognitiva, in particolare per quel che concerne l’*enattivismo*. In questo articolo mostriamo che sia i fattori interni che quelli esterni contribuiscono in maniera non-ridondante alle capacità psicologiche e al comportamento di un individuo. Inoltre sosteneremo che l’adozione di un *locus di controllo* distribuito possa costituire un buon punto di partenza per una revisione sostanziale della nozione di “rappresentazione mentale”.

PAROLE CHIAVE: Locus di controllo causale; Internalismo/Esternalismo; Biologia evolutiva; Scienze cognitive; Rappresentazione mentale

^(a)Department of Philosophy, Tulane University, Newcomb Hall, 105 - 70118 New Orleans, Louisiana (USA)

^(b)Dipartimento di Filosofia e Comunicazione, Università degli Studi di Bologna, via Zamboni, 38 - 40122 Bologna (IT)

^(c)Max Planck Institut für Geoanthropologie, Abteilung Sprach- und Kulturevolution, Kahlaische Strasse, 10 - 07745 Jena (DE)

Email: dburnsto@tulane.edu (✉); antonella.tramacere2@unibo.it



1 Introduction

Suppose a group of children that, for whatever reason, wants to explain the movements of dolls and puppets. Suppose they take two systems of interest to study – a marionette, and a more new-fangled doll that moves and talks on its own. The behaviors of these two systems will require very different explanations. In the marionette, the motive force for its movements originates *outside* of its boundaries; namely, with the strings and the puppeteer. In the modern doll, however, the explanation of its movements will posit accounts about its *internal* organization – i.e., its motor, its microphone, etc. In the terminology we will employ in this paper, the marionette’s movement has an *external* locus of control. The doll’s behavior has an *internal* locus of control.

We will argue that long-running debates in evolutionary biology and cognitive science have taken place between *internalists* and *externalists* about the locus of control. While these terms have a long history in philosophy, for instance with regards to semantics and epistemic justification, we only employ them here for thinking about the relative causal contributions of components internal to a system and those external to it to that system’s behavior, as described in the case above (i.e., we do not mean to imply anything about those other debates). In this dialectical setting, “internalists” argue that the causal source of the phenomenon of interest is internal to the organism, and “externalists” that it is external to it. In evolutionary biology, one phenomenon of interest is phenotypic evolution. Internalists have historically focused on genetic modification as the causal source of phenotypic change, while externalists focus on the environment and developmental organization of an organism. In cognitive science, the phenomenon of interest is intelligent behavior. Internalists have generally focused on explaining behavior by positing internal representations or models, while externalists deny the importance of these constructs, instead positing interaction with the environment as the primary *explanans*.

Interestingly, in the evolutionary biology debates there is an emerging consensus towards what we call a *distributed* locus of control. Modern theorists have begun to recognize that both genes and environmental factors exert specific, but non-redundant, causal influence on the heritable variations. No such consensus has developed in the cognitive science case, however. We suggest that such a move is needed. We further suggest that moving to a distributed locus of control has upshot for the construct of “mental representation”. In evolutionary biology, the concept of “gene” has been modified with the progress of genetic and molecular investigation. Although a plurality of gene concepts can be recognized today in evolu-

tionary biology, they are not anymore seen as a causally sufficient internal program for producing traits. In contrast, a view based on a distributed locus of control interprets genes as a non-sufficient causal contributor that interacts reciprocally with environmental factors.¹ Again, we suggest that a similar concept change can productively posit mental representation as the internal contributor to a distributed locus of control for behavior.

At the outset, we note that positing a distributed locus of control does not end debate. In any specific case, it is still important to determine the relative causal contributions of internal and external factors for the phenomenon of interest. Moreover, individual scientists may reasonably focus, both methodologically and theoretically, on primarily one kind of causal factor. What it does do, and what we suggest has occurred in evolutionary biology, is produce a *minimal agreement*, that in many or most cases the causal source of the phenomenon will involve, non-eliminably, both internal and external factors. On a distributed-locus view, it is thus *methodologically* important to decompose systems into internal and external factors, and to study their interactions and their relative contribution in different contexts.

In Section 2, we begin by defining a locus of control, and explaining the difference between *internal*, *external*, and *distributed* loci. In Section 3, we assess the history of debates in evolutionary biology, arguing that the Modern Synthesis primarily posited an internal locus of control for genetic variation (3.1), that critics of the modern synthesis proposed alternatives based on external loci (3.2), and that contemporary evolutionary theory is moving towards a minimal agreement that the locus of control for evolutionary change is distributed. In Section 4 (4.1), we introduce the debate surrounding “*enactivism*” in cognitive science, and argue that, here, the debate is similarly driven by different sides positing different loci of control. Furthermore, we propose how a distributed-locus view accounts for behavior and describe the role of mental representation in these explanations, giving examples from psychology (4.2) and neuroscience (4.3). Section 5 concludes.

2 Locus of control

A locus of control is a combination of *causal* and *locational* factors. It is causal in the sense that manipulating the locus of control changes the phenomenon of interest in a specific, fine-grained way. It is locational because it occupies some position, either spatial or topological, in the system under study. Suppose that the phenomenon of interest is the long-term fiscal strategy of a corporation. In this case, the locus of control is the board and the CFO. Variations in the decisions made by

these factors will change the fiscal policy in a fine-grained way, and these factors are located at a specific location in the organization, namely at the top of a decision hierarchy.

We can further flesh out the notion of causal control of the phenomenon with the notion of *causal specificity*, familiar from the literature on causation.² According to interventionist accounts of causation, causes are specific to effects when fine-grained manipulations of the cause result in fine-grained changes to the effect. Woodward's example is a radio: while the power switch will exert coarse control over the noise emitted from the radio – either sound or no sound – the tuning dial will exert fine-grained control by determining the specific frequency to which the radio is tuned. On our view, a locus of control comprises components that exhibit fine-grained causal influence on the phenomenon of interest.

The notion of “location” we care about in this paper is whether the locus of control is *internal* or *external*. Internalists localize the locus of control within an organism, externalists outside of it. This needs immediate further specification, though, since organism-environment boundaries are often blurry, and organisms are not unitary. Positing an internal or external locus of control depends on having *some* extant division between a system and its environment, and often which system within the organism is posited varies with the phenomenon of interest. Candidates for an internal locus of control thus might be the genome, the brain, the circulatory system, the immune system, etc. External environments might include the organism's ecological niche, its conspecifics, its family, and so on. When we talk about internal versus external locus of control, we talk about whether factors internal or external to the relevant system drive the phenomenon of interest.

Let's return to our doll examples. The explanandum is the movement of the dolls. In the marionette case, it is the movements of the puppeteer and the strings that explains the way the doll moves. If the puppeteer moved even slightly differently, the marionette would move differently as well. In the doll, however, it is the movement of the motor, as well as the linkages to the doll's “limbs” that explains the doll's movement.

It is important to note that internalists are not committed to *never referencing* external factors in their explanations; nor are externalists committed to never referencing internal factors. What movements the marionette can perform will depend, to some degree, on the structure of the puppet's joints, what material it is made of, etc. The behavior of the doll will depend on whether anyone has turned it on. What is important, however, is that these references to factors not in the locus of control are not causally specific. For any token explanandum, the fine-grained control of that ex-

planandum will be posited to reside in the locus of control – external for the marionette, internal for the doll. Or, to take an even simpler example: the radio will not work if I unplug it. But the presence or absence of electricity is a coarse-grained influence, compared to the manipulation of the dial.³

In a distributed locus of control, alternatively, internal and external factors are, at least in principle, equally specific in the grain of control they exert on the phenomenon, and thus manipulating either would change the outcome in a fine-grained way. Systems with distributed loci of control are harder to characterize, and uncontroversial examples are therefore a bit harder to come by. One explanandum that would be a candidate for a distributed-locus explanation is policy enactment in a representative democracy. Actors in the government make policy in such a system, but which actors are in office, as well as which policies they feel they can safely support, depend on public opinion. Another example may be commodities exchange in a regulated market. While the exchange of goods is driven by (amongst other things) supply and demand, these factors occasionally produce inefficiencies or exploitation, which are then regulated by governments. The exchange of commodities thus depends on both factors internal to and external to the market.

Vitality, in a distributed locus of control, the contributions of internal and external factors are independently characterizable. While both may exert fine-grained influence on the phenomenon, their contributions are not the *same*. The regulatory body does not exert the same control over commodities exchanges as firms do, for instance. Hence, any explanation of a system with a distributed locus of control must characterize independent, if interacting, causal forces internal and external to the system. The notion of a distributed locus of control is therefore different from the idea of a *holistic* system. Holists argue that there are no important divisions either within systems or between systems and their environments. This is because individual, internal components of the system make no specific causal contribution, aside from their interactions with other components and the environmental context.⁴ Since holists deny the division between a system and its environment, positing holism is logically distinct from positing any of internal, external, or distributed loci.

As a matter of fact, though, many holists in the debates we will discuss ally their views with externalism. This is because holists deny that internal components make any independent causal contribution to the phenomenon of interest. At best, on this kind of view, internal components are *coupled to* or *driven by* external ones, and thus the fine-grained control in the system is due to external factors. We will thus consider holist views to be examples of externalism, and we will discuss several examples in the paper.

3 Internalism vs externalism in evolutionary biology

3.1 Internalism in the Modern Synthesis

The Modern Synthesis at the heart of 20th century evolutionary theory posited an internal locus of control for evolutionarily significant phenotypic change. On this view, the genotype of an individual produces its specific phenotype. Hence, differences in the genetics of different individuals drive phenotypic variations, which in turn are the basis of selection and evolutionary change. The Modern Synthesis thus combined a Darwinian perspective on evolutionary change with emerging knowledge of genetic mechanisms. Consider the following quote from Thomas Hunt Morgan:

All the genes are instrumental in producing each organ of the body. This may only mean that they all produce chemical substances essential for the normal course of development. If now one gene is changed so that it produces some substance different from that which it produced before, the end-result may be affected, and if the change affects one organ predominantly it may appear that one gene alone has produced this effect. In a strictly causal sense this is true, but the effect is produced only in conjunction with all the other genes.⁵

In this quote, Morgan evinces an internalist notion of the drivers of phenotypic variation. Note the implicit interventionist attitude taken towards genes – it is posited that changes to the genome are causally responsible for phenotypic outcomes. As such, explanations for variations in phenotype are to be given by citing causes in individuals' genes. The quote, importantly, does not evince a naïve atomism about genetic influence. It is admitted here that individual genes can influence specific traits at a variety of degrees of specificity, and further that an individual gene's contribution can only operate in conjunction with both other genes. Nonetheless, the locus of control for an individuals' traits is posited to be in its genes – the genetic makeup of the individual as a whole exerts specific causal influence on its phenotype.

The role of the genes is to determine the “normal course of development”. This is a tacit admission that external factors can of course affect how phenotype comes out – removing oxygen from an aerobic organism's environment, for instance, will prevent it from developing. This is however a *non-specific* influence on evolved phenotypic variation, much akin to removing the power from the radio. In some cases, the phenotype reactivity to different environmental factors has been studied, but considered as a global and dynamic property of genotypes and proposed in the concept of *norm of*

reaction.⁶ Also in the concept of norm of reaction however, fine-grained control over heritable changes resides in the individual's genes, because genes are causal factors that are differentially inherited across generations and control phenotypic expressions in different environments. Therefore, the locus of control for phenotypic variations is internal.

As molecular biology developed in the mid-20th century, the idea of an internal locus of control for traits was taken up in the famous “central dogma” of molecular biology. On this view, genetic “information” is transmitted unidirectionally from genes (the sequence of nucleic acids in the DNA) to RNA to protein, and never in reverse. This shaped the molecular notion of genes as a conserved syntax-like sequence of chemical bases, the genome as a *program*, which encodes a set of instructions for development.⁷ On this view, an individual inherits a genetic program, encoded in the syntax of its genetic code – DNA has often been referred to as a *code of life*⁸ or a *blueprint*.⁹ This view fits naturally with the internal locus of control, since such analogies posit fine-grained influence of genotype on phenotype. Fine-grained changes to the code/blueprint will result in fine-grained changes in the organism's traits.

Let's consider a classic example, which we will use to contrast the internalist view with externalism in the next section. During the industrial revolution in Great Britain, certain moths, *Biston betularia*, changed their colors. While the moths, prior to the industrial revolution, were light-colored and “peppered”, air pollution from industrialization produced a layer of soot on the trees that peppered moths perched upon. Over time, a larger number of peppered moths began to exhibit a distinct phenotype, turning from peppered to almost solidly black.

On the internalist story, change occurs through random mutation in the genome. Peppered moths must have either already possessed different alleles for different colors,¹⁰ or undergone a mutation which produced, in some individuals, the black phenotype.¹¹ The new phenotype was then available for natural selection, which operated on it due to the camouflage advantage experienced by darker moths in soot-covered trees. So, the specific change in phenotype is caused by the specific change in genotype. Concomitant with this approach is a search for genetic modifications that are implicated in producing the phenotypic change. The environment applies selection pressure, but it does not influence the genetic variation that determines phenotypic outcomes. That is, the Modern Synthesis conception explains traits by positing an internal locus of control.

3.2 Critiques of the internalist conception

Major critiques of the Modern Synthesis, we suggest, are based on criticisms of the internal lo-

cus of control that it posits. Concomitantly, they are often pitched as critiques about the preformist notion of genes, arguing that phenotypes are in fact determined by the particular ways in which organisms interact with their environments. As such, critics of the Modern Synthesis propose as an alternative the idea that phenotypic variation has an external locus of control.

There are a variety of positions that fall under this rubric. For an early statement, consider the following quote:

Contrary to the neo-Darwinian view, we point out that the variations of the phenotype, on which natural selection could act, do not arise at random; they are produced by interactions between the organism and the environment during development.¹²

Building on an analogous proposal from Konrad Waddington,¹³ the authors here state that phenotypic change is *not* due to random change in the genotype. Instead, it is driven by specific environmental variation. Change the individual's environment in particular ways, the quote suggests, and you will change its phenotype. Hence, the quote posits an external locus of control.

On the views proposed by externalists, the syntax-based concepts of genes and the genome as program are untenable. Critics have emphasized that the concept of "genetic program" misrepresents the role of development and environmental contexts and promotes a reductionistic view of evolution. On these views, genes are better described as participating in regulatory networks or as subroutines in developmental operating systems.¹⁴ The gene concept was thus progressively based on views including extended regulatory networks, as to include epigenetic interactions and cellular products.

Externalist views thus focus on the causally specific influence on heritable phenotypic traits by environments. As noted above, externalists do not have to deny that genetic changes are causally relevant to variations in phenotype. In particular, externalists often posit that variations in genotype are *driven* by changes in environment, through epigenetic interactions.¹⁵ What is important, however, is that the directionality of this process is distinct from the directionality on internalist views. Genes, on their own, do not produce specific phenotypic variation on externalist views. Instead, interactions with the environment both drives genetic expression and "unlocks" extant genetic variation, causing it to produce specific phenotypic changes in specific environments.¹⁶ So, the locus of control for the variation is in external factors.

To see the contrast here, consider how an externalist would explain the peppered moth case. On this view, the interaction of peppered moths with their environments produces variation in

their phenotype that was not already present. Again, this process may partially run through already extant genetic variation, but the genetic variation that was present is non-specific to black coloring. Black coloring is the result of specific interactions with an environment. This view, like its internalist alternatives, comes along with a methodological/explanatory prescription, namely to search for how environmental information is recognized by the organism and results in phenotypic plasticity.¹⁷

A variety of positions in evolutionary biology take up this alternative, externalist notion of the locus of control. On these views, it is an environment with a particular structure that explains heritable phenotypic variations through epigenetic regulations of genes, and not the organisms' genes per se. The heritable changes in the phenotype produced by the environmental influences on the organisms (eco-phenotypes) can in turn be *stabilized* across evolutionary time.¹⁸ Epigenetic interactions between organisms and environments can become standard and produce canalized developmental pathways leading to consistent phenotypes across individuals. For instance, the post-industrialization environment of the peppered moth could exert developmental/epigenetic effects on successive generations of moths, underlying the stability of the black coloring phenotype across generations.

Another position to discuss in this tradition, but one that requires some care in exposition, is developmental systems theory (DST).¹⁹ DST proponents, like other critics of the Modern Synthesis, reject the idea of genes as privileged causal factor in development, as well as the idea of the genome as a program for producing phenotypes. DST theorists, though, often end up embracing a kind of *holism* about genetic/environment interactions.²⁰ On this view, there is one continuous systemic interaction between genetic and environmental features, and the unit of analysis is the developmental system. As we noted, holist positions are not the same as externalist ones, because of the emphasis on the organism, rather than on internal or external factors. However, there are aspects of DST that fit with an external locus of control. Specifically, in DST phenotypic change is often taken to be the result of continuous dynamic interaction between a system and its environment.²¹ Hence environmental changes produce phenotypic change by coupling with internal factors in a non-decomposable way.

3.3 Contemporary Evolutionary Synthesis

We have argued that the genetic Darwinism of the Modern Synthesis was an internalist conception of the locus of control for evolved variations, and that criticisms of this view have taken an externalist position. We think that, in evolutionary

biology, the field is trending towards a minimal agreement that the locus of control for evolved variation is distributed. On a distributed locus of control view, both genetic and environmental/developmental factors exhibit specific, non-redundant causal effects on heritable phenotypic changes during evolutionary processes. The phenotypic outcome for organisms is thus due to both kinds of factors.

On this view, genetic variation may be partially responsible for development of new traits, but the development of these traits requires specific contributions from developmental processes and environment. There are no *primum movens* or in principle directions of causal primacy. Genes are indeed heritable factors in producing phenotypes. But often, developmental and environmental factors will influence genes through epigenetic regulation, and traits will develop within, and take advantage of, environmental niches.²² The environment can thus influence heredity in a variety of ways. For example, environmental influences (or different environmental niches) can affect genetic expression through epigenetic regulation, therefore varying the rate and type of heritable mutations.²³

In contemporary evolutionary biology, a variety of gene concepts have been proposed and are used in different experimental contexts and sub-disciplines. Genes are understood in a pluralist fashion²⁴ and both scientists and philosophers recognize that the concept of gene has evolved and will continue to evolve with scientific discoveries, explanatory practices, and goals. However, no matter how genes are conceptualized, the recognized causal effects of genes reflect the distributed locus of control view we are defending here. This point is well expressed by Waters in his *Pluralist interpretation of gene-centered biology*. While commenting on the criticisms of environmentalists and developmental system theorists to the gene-centrism and preformation of the Modern Synthesis, he states:

Biological research, much of it gene – centered, has indeed shown that inheritance is systematic and involves the interaction of genes, accessory molecules, cellular structures, and the surround. Furthermore, [critics have] a valid point about those who would want to identify nature with genes and nurture with environment. What I want to emphasize is that this argument [ndr against causal privilege of genes] does not show, and apparently does not purport to show, that genes cannot be “properly contrasted” with environment in certain contexts.²⁵

What Waters is saying here is that, irrespective of the many definitions and conceptions of genes in various scientific fields and in philosophical re-

flection, scholars acknowledge that genes are only one of the many causal factors having effects on phenotypes and that it is important to disambiguate the specific causal role of the environment and of genes in producing inherited variations. As such, this claim is useful to capture the minimal agreement we have outlined here, which suggests that, rather than either an internalist or externalist view, the best approach posits a distributed locus.

As noted in Section 2, it is vital on a distributed locus of control view that internal and external components be independently characterizable. Understanding, for instance, how epigenetic regulation works requires distinguishing the genetic component from the environmental factors that regulate it, and in turn understanding their interaction. While the result may be a kind of “reciprocal causation”,²⁶ it is a reciprocal causation between functionally distinct elements.²⁷

Let’s apply the distributed control perspective to the peppered moth example. This perspective suggests that there will be both internal and external factors that will contribute non-redundantly to the variation in phenotype. So, while it is important to understand if there are genetic variations specific to possession of the phenotype, this in no way undermines the need to look for environmental and epigenetic processes that also may be required. Both kinds of factors will affect the evolution of color distributions across the moth population. Indeed, many current models attempt to account for and differentiate the role of these distinct factors, as well as the reciprocal interactions between them, in a bidirectional eco-evolutionary feedback system.²⁸

It is important to note the sense in which the kind of agreement we’ve outlined here is “minimal”. What the minimal agreement suggests is that there is no *in principle* primary causal factor or causal directionality. This is compatible with different specific traits being more-or-less under the control of genetic or environmental factors. It is also compatible with a range of positions about the nature of evolutionary theorizing. Some, for instance proponents of the “Extended Evolutionary Synthesis”, have suggested that the recognition of reciprocal causation is both an expansion and a foundational revision of Darwinian evolutionary theory.²⁹ Others think that the recognition is compatible with and simply adds to traditional Darwinism – i.e., that the “Extended” in “Extended Synthesis” should be lowercase.³⁰

Finally, the notion of a distributed locus of control is also compatible with the methodological focus of a specific scientist or group being primarily on one kind of factor or another. Indeed, the distributed locus view explains the need for this differentiation, since both internal and external factors need to be independently characterizable on that view. So, the minimal agreement we’ve

outlined does not refer to a monolithic evolutionary theory or methodology. What it does suggest is the recognition of the field that the domain is not structured in a way that supports either internalism or externalism about the locus of control. In the remainder of the paper, we suggest that a similar progress is needed for theoretical debates in cognitive science.

4 Locus of control in cognitive science

4.1 *The debate*

In the previous sections, we argued that traditional debates in evolutionary biology have been driven by disagreements about the locus of control for heritable phenotypic variation. We also suggested that a minimal agreement has emerged that the locus of control for evolved traits is, in fact, distributed. In this section we argue that debates in cognitive science are at a stage similar to the traditional debate in biology, and that they need to move in a similar direction. That is, theorists are debating between internal and external loci of control, when the consensus should be that the locus of control is distributed.

We further suggest that, in the cognitive science debates, the idea of mental and neural “representations” should be viewed similarly to that of “genes” in the biology debates. That is, it became a lightning rod for contention in the traditional debates, only to evolve to serve a more modest explanatory role in the agreement about distributed locus of control. Again, we suggest that a similar development is both possible and desirable in the cognitive science case.

In cognitive science, we presume that the phenomenon of interest is intelligent behavior (we offer no analysis of “intelligent” here). From its foundation, “classical” cognitive science was dedicatedly internalist. Founded on the analogy to digital computers, classical cognitive science posited rich internal data structures underlying thought and behavior.³¹ Mental processes were equated with manipulations of syntactic strings in a computational language according to internalized rules. The semantic interpretation of these strings is the foundation for the idea of a mental representation. Further, classical cognitive science viewed the relationship between the mind and the body in a similar way as the modern synthesis viewed the relationship between genes and traits – on the classical view, the mind encodes a set of instructions that the body follows, hence producing observable behavior.

On the classical view, factors internal to the organism’s mind are the locus of control. Changing the mental “program” being run by the organism – or, in its folk-psychological guise, changing the organism’s beliefs and desires, which are presumed

to be internal states – will exert fine-grained changes on the behavior of the subject. The idea of an internal program or “model” has also been foundational for cognitive neuroscience, which, traditionally, has taken as its goal the localization of the representations and mental functions posited in cognitive theories.³²

In the last 25 years, a variety of alternative research programs have developed, focusing on embodiment and dynamics as opposed to internal representational structures. We focus on the position often referred to as “enactivism”.³³ This view is often pitched explicitly against the internalism of the classical position. Enactivists are fond of pointing out that, according to classical cognitive science, an organism could be a brain in a vat and their mind would be unchanged – since the locus control for their thoughts and behaviors is internal, the facts of their external environment are at best minimally relevant in determining what goes on in their mind. Enactivists propose that this view of the mind is radically false. The mind, they suggest, can only be understood as embedded in an environment; we misunderstand the mind by talking about it solely in terms of an internal program performed over mental representations. Instead, they posit that the very nature of the mind is determined by the environments with which it interacts.

In our reading, enactivism embraces an external locus of control for behavior, which is sometimes combined with a holism about mental systems. Concomitant to these moves is a view about the traditional explanatory posits and methodological aims of both classical cognitive science and cognitive neuroscience. In particular, enactivists systematically deny the existence and explanatory usefulness of mental representations, and they deny the possibility or desirability of functionally decomposing the brain. Rather, they suggest, one can only understand the mind and brain as holistic systems coupled to the environment. This view is externalist because it denies the independent contribution of factors internal to the mind, and thus claims that fine-grained control of behavior is due to the environment in which cognitive agents act.

As such, enactivism comes along with the standard externalist set of methodological and theoretical prescriptions – it suggests limited utility to characterizing internal factors independently of the external ones with which they interact. Specifically, enactivists deny that cognitive science ever benefits from positing internal informational structures, even in cases such as memory or imagery. This methodological stricture applies to the brain as well. While enactivists admit that the brain makes *some* contribution to behavior, perhaps by synaptic or genetic changes in the case of memory,³⁴ or by physiological response to the environment during behavior,³⁵ they insist that (i)

there is no independent representational characterization of that contribution, and, concomitantly, (ii) the contribution can only be characterized as part of a dynamic coupling driven by the environment. In other words, the contribution of neural or genetic factors outside of environmental interaction is *non-specific*, in the sense outlined above. While some genetic or physiological processes may be necessary for memory or behavior, the explanation of their contributions depends asymmetrically on external factors.

We wish to note at the outset that we think enactivist approaches have made significant contributions to our understanding of the mind. We agree with enactivism that the internal locus of control posited by classical cognitive science is untenable. But we disagree with their externalist approach just as much. We suggest that, on a parallel with the emerging minimal agreement in evolutionary biology, cognitive and neuroscience need to embrace a distributed locus of control. And, while it is logically separable from this main point, we suggest that this makes room for the concept of mental and neural representation. Similar to how the concept of *gene* has evolved from a causally specific program to a contributing causal factor in a distributed locus of control, we suggest that the internal *aspect* of the distributed locus of control for cognition is plausibly characterized in representational terms.

There is no universally agreed upon set of conditions for what makes something a mental representation, although there are a number of proposals for what the “core,” minimal notion of representation comprises.³⁶ There have also been explicitly revisionary proposals for how to construe mental representations outside of the classical picture.³⁷ Enactivists are fond of denying that these alternative proposals *really* offer an account of mental representations, rather than simply re-describing enactivist points in representational language.³⁸ One might take this situation as sufficient reason to just abandon the notion of representation altogether – not as false *per se*, but as not theoretically useful.³⁹

In our view, there is no pre-theoretically agreeable set of necessary and sufficient conditions for what counts as a representation, and therefore the notion is likely to evolve with time. Moreover, there is nothing objectionable to this evolution so long as the particular theoretical commitments being undertaken by particular theorists are clear; indeed, this kind of evolution mirrors the evolution of the concept “gene” and the interpretation of its causal role in evolutionary biology.⁴⁰ We suggest that locating “representations” within the internal components of a distributed locus of control is one productive way for the discussion to advance.

Some recent advancements from Rowlands are

helpful in this regard.⁴¹ First, Rowlands suggests that at the core of the idea of representation are the intertwined ideas of *decouplability* and *normativity*. If a mental state or process can be decoupled from the environment, then its function is understandable independently of its active interaction with aspects of that environment. Further, and partially *because* of the decouplability of the function, it has a *normative* aspect. That is, it can fulfill or not fulfill its function. Importantly, Rowlands notes that these are not offered as necessary and sufficient conditions – other theorists can and do add other notions, or attempt to abandon these. But they are particularly useful in our discussion, for two reasons: because they help clarify the opposition with enactivism, in arguing for decouplability where enactivists deny it, and because they presume no particular account of mental *content*. Particular notions of what mental content must amount to often inform debates about whether representation exists at all. But there are different accounts of what semantic content might be,⁴² so it seems, in our view, unwise to insist on one as constitutive of representation.

Certain instances of learning, on our view, are clear evidence of the distributed locus of control.⁴³ We argue that particular patterns in how agents learn new behaviors, shown at both the psychological and the neural level, evidence functioning of brain systems in a way that is not strictly definable as coupling with the environment. Enactivists are often vague about what they think learning does, saying that learning allows the agent to become attuned to new aspects of the environment, without representing it. We think this is insufficient – the outcome of learning is an internal structure that is not fully explicable in terms of coupling with specific environments. At the same time, however, they do not constitute programs or complete internal models of the environment. Both the nature of the learning process, and the success of the learned behavior, depend on a coordination between the learned internal/mental structure and the environmental or task context. Hence, the locus of control for behavior is distributed.

Here is how we think the story goes. In a learning context, agents develop internal structures and processes that track the structure of the task situation. Consider this as a kind of statistical learning about the task situation, which is neither necessarily nor strictly determined top-down by agents’ prior knowledge or assumptions.⁴⁴ Hence, the nature of the internal process depends, etiologically, on the structure of the environment. A different environment would produce a different learning outcome. However, once the structure is learned, the structure enables generalization to novel examples and instances. Change the internal structure, and you would get a different pattern of generalization, and a different range of environments

in which the behavior would be successful. So, both internal and external factors exhibit fine-grained causal contributions to successful behavior, but those are not the same contributions.

We suggest that the internal causal factors here, which we will refer to as “internal structures” to avoid begging the question, are good candidates for a modern conception of mental representation, because they meet decouplability and normativity. Decouplability is evidenced by the fact that the contribution of internal structures is not fully explained by interacting with specific environments. It is only by talking about the internalized structure that one can explain patterns of generalization. But normativity is explained by the fact that whatever environment is generalized to must share a structure with the learned environment in order for behavior to be successful. In what remains of the paper, we discuss examples meeting this description of internal structure at both the psychological and neural levels.

■ 4.2 Psychological examples

We start with a very simple example about how environmental factors contribute to the allocation of attention. By “factors” here we simply mean aspects of the environment; this can include, for instance, regularities amongst the kind of stimuli the organism encounters, or in the kinds of rewards available in the environment. A simple kind of investigation in this vein looks for how regular structures in the environment are used to shape attentional patterns. For instance, in a study by Zang and colleagues,⁴⁵ subjects were instructed to look for a shape amongst distractors. In some experimental groups, the shapes were distributed around a task-irrelevant display structure, such as a large cube. Despite the task-irrelevance of this structure, subjects learn to use it to guide their attentional search. When it is removed in a subsequent task epoch, subjects’ search success goes down, even if the layout of shapes and distractors remains the same.

We suggest that this simple case exhibits the structure of a distributed locus of control. What subjects learn depends on the task environment – if you changed the shape of the cube, for instance, subjects would learn a different search pattern. Once learned, however, the internalized structure shapes behavior, even in different environments (e.g., the ones without the cube). If you changed the internal structure, the search behavior would be different. The success of the behavior requires that the two independently characterizable components – the environment and the learned structure – to be aligned. So, while the internal component is decouplable from the environment, it needs to be employed in the right circumstances for behavioral success, hence underlying normativity.

Even clearer examples come in cases where generalization is explicitly investigated. In studies of motor control, for instance, subjects learn a task situation and then perform actions in a range of conditions to test what they’ve learned.⁴⁶ These studies employ a motor control element and a visual feedback element. So, subjects might have to move a joystick, which controls a visually presented cursor, until the cursor reaches a target. In some conditions, the visual feedback is “perturbed” – the cursor, for instance, may move to the left or right – and subjects have to adjust to the perturbation by manipulating the joystick. Learning is shown by quicker and more efficient responses to perturbations.

A huge range of studies have shown that subjects learn to respond not only to individual perturbations, but to ranges of perturbations, which subsequently allows them to generalize.⁴⁷ So, if the perturbations occur across a range of angles and magnitudes, subjects will not only learn to respond to any of those perturbations, but will respond to a novel perturbation just as well, if it falls in the range.

Here we see the same kind of explanatory structure playing out. Subjects’ learned abilities depend on a structure in the environment – in this case, the variational structure in the perturbations they experience. However, they also internalize that structure in a way that allows for generalization. It is thus not coupling to specific environments that determines behavior, but an internalization of task structure. Again, while a different task structure would produce a different internal structure, this does not mean that the internal structure is not an independent causal contributor. Of course, successful generalized behavior depends upon the new task condition falling within the generalized range. So, the locus of control is distributed, and the internal component is both decouplable and underlies normativity.

A last psychological example comes from studies of categorical perception.⁴⁸ Studies in this field start with construction of a “space” of examples, such as the one shown below. The space is constructed by “morphing” two pairs of faces towards each other, generating new examples along the way. Each face is thus a combination of the four “parents”. In the studies, experimenters determine what the categories will be. In the examples below, “A faces” are on the left side of the (arbitrary) line in the middle, and “B faces” are on the right. Subjects are trained with feedback, and after training become good at immediately recognizing a face as an A or a B. This involves discriminating along the horizontal “X axis” in the space.

Importantly, subjects then generalize this learned structure in a variety of ways. For instance, if a new space of faces is constructed by keeping the parents comprising the X axis the

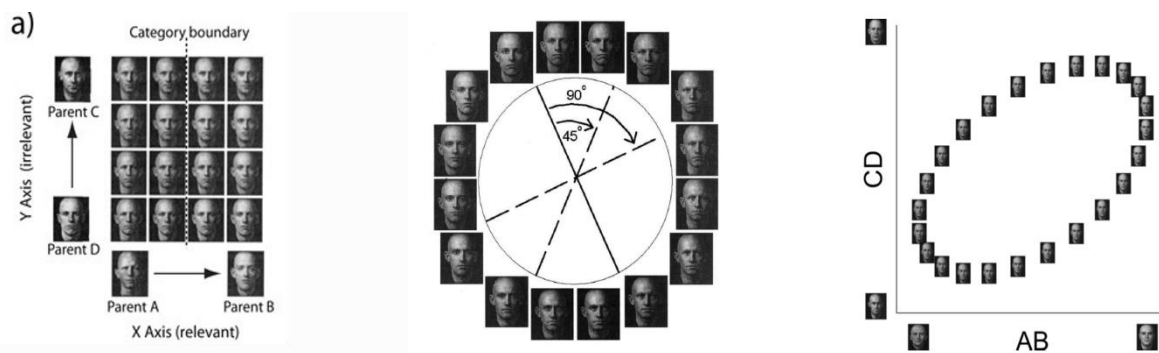


Figure 1. Exemplar spaces. Left panel after J.R. FOLSTEIN, I. GAUTHIER, T.J. PALMERI, *How category learning affects object representations: Not all morphospaces stretch alike*. In: «Journal of Experimental Psychology: Learning, Memory, and Cognition», vol. XXXVIII, n. 4, 2012, pp. 807-820. Middle panel after R.L. GOLDSTONE, M. STEYVERS, *The sensitization and differentiation of dimensions during category learning*. In: «Journal of Experimental Psychology: General», vol. CXXX, n. 1, 2001, pp. 116-139. Right panel after M. JONES, R.L. GOLDSTONE, *The structure of integral dimensions: Contrasting topological and Cartesian representations*. In: «Journal of Experimental Psychology: Human Perception and Performance», vol. XXXIX, n. 1, 2013, pp. 111-132. All images courtesy of Rob Goldstone.

same, while pairing them with a new set of parents, Z, subjects can still recognize the faces.⁴⁹ That is, they learn to recognize the “X” pattern of variation even in new examples. Other kinds of generalization are also relevant. For instance, if in a subsequent task epoch the category boundary “turns” 90 degrees, subjects can transfer their ability almost immediately. But such a transfer is not shown if the category boundary turns 45 degrees (cf. middle panel, *Figure 1*). This is explained by the posit that subjects not only learn the dimension that defines the category, but the dimension orthogonal to it, as a way of fully parsing the space.

Importantly, there is no particular set of exemplars that is necessary to produce this result. Subjects can be trained on a range of exemplar sets, so long as those exemplars vary in the way that is relevant for subsequent category recognition.⁵⁰ This is shown in the middle and right panels of *Figure 1*. Again, we have the same explanatory structure we have discussed above. What structure subjects learn depends on the variation that is present in the exemplar set and the feedback they receive. Once they have learned that structure, they can generalize it to new instances. But successful generalization requires the new instances to have the right kind of structure.

4.3 Neural dynamics

Enactivists regularly espouse a dynamic approach to the brain. Rather than thinking of the brain in terms of a functionally decomposable system, they argue, we should understand it as dynamically coupled to the environment.⁵¹ This claim rarely is pursued at anything other than an abstract level, however, and historical debates about the relationship between representation and dynamics have similarly taken place at a far remove from neuroscientific practice.⁵² There is no doubt that neural dynamics are vitally important for under-

standing brain function, but we should ask whether that fact is best read as supporting enactivism. In this section, we suggest that investigations into neural dynamics are in fact supportive of a distributed locus of control.

We will focus specifically on frontal cortex, a part of the brain that is involved in organizing complex behavior. Importantly, recent physiological and modeling studies of this area suggest that function is widely distributed, and that individual units in the system are multifunctional. Particular variables cause physiological responses in widely distributed groups of cells, and individual cells are “multiplexed”, showing significant physiological responses to an array of task-related variables.⁵³ Moreover, the same cells respond to distinct variables in distinct conditions, suggesting that function is highly context-sensitive. The functionality of a system like this is often determinable only at the population level, and that functionality consists in patterns of dynamic activity over time. Despite this, we think, studies in this area do not support an externalist view.

Instead, a vital part of understanding these systems involves *decomposing* the dynamics into distinct regimes that correspond to variations in task context. Again, learning is an important part of the story. We discuss only one highly-cited example here, but we take this as an exemplar for how to understand neural dynamics on a distributed locus of control-based view.

Mante and colleagues attempted to discern the function of a distributed population of prefrontal cortical cells in complex task situations.⁵⁴ Their question was the following: how can a prefrontal population sensitive to multiple task-relevant factors change *which* factor should drive behavior in a particular context? To analyze this, they modified a traditional “dot-motion” task paradigm. In traditional dot motion studies, a subject, in this case a monkey, views an array of moving dots and has to

judge the direction of their predominant motion. With no correlation in the motion of individual dots, there is no predominant motion. With small degrees of correlation monkeys remain close to chance at detecting it, but at greater degrees of correlation they can generally report the direction in which the motion of the dots is correlated.

Mante and colleagues' modification was to also vary *color* along with motion. So, the color of the dots could be either red or green. The level of predominance – i.e., the proportion of the colors – could vary just as the degree of correlated motion did, and across trials both the color and motion predominance varied independently. When given a context cue telling them which was the relevant stimulus parameter, monkeys would indicate their choice via a saccade to the right or to the left. The results showed that in the color context, the monkeys could accurately judge the predominance of color independently of motion, and vice versa in the motion context. The question was how their brains separated and organized the information.

Importantly, widely dispersed prefrontal cell populations showed overlapping physiological selectivity for color and motion – they were distributed and multiplexed in the ways described above. Equally importantly, color perception *continued to influence* physiological responses in this population in the motion-judging context, and vice-versa. So, the story is not a simple one of the monkey attending to the output of a distinct, independent color response in the color context, and a separate, independent motion response in the motion context. Somehow, the very same population that was responding to the task parameters extracted the relevant information in the correct context.

The full story of Mante and colleagues' explanation for how this works is complex, but here are the basics.⁵⁵ First, they analyzed population responses in terms of *task axes*, represented in a principal components space. Principal components analysis (PCA) measures covariation in a high-dimensional system, in this case the physiological responses of a large population of individual neurons, in terms of orthogonal dimensions, small numbers of which are sufficient to describe the variation of the entire population. Mante and colleagues performed PCA, then constructed a description of the populations' specific responses to variation in motion, color, and context in terms of those components. That is, the population's response to each task parameter was construed as a particular subset of the overall population response. Lastly, they analyzed the axis of *choice* – i.e., which behavioral indication the monkey made as indicated by the direction of their saccade response. The outcome of the initial analysis was that the task context changed the relationship between the task parameters and choice. In effect, in motion contexts motion cues would vary the pop-

ulation along the choice axis, while color cues wouldn't, and the opposite in the color context.

To analyze this further, they trained a recurrent neural network to mimic the population response, and analyzed the dynamics of this network. They showed that the system exhibited two distinct "line attractors", corresponding to the motion context and the color context, and that the context cue would push the system into one or the other. On the line attractor for the motion context, motion cues would drive the choice, and color would drive the choice on the attractor for the color context. Details aside, what is important here is that the system learned to *modify its own dynamics* given the task context. What the system does is learn to implement different dynamic regimes for the same set of stimuli, as the task context varies. Learning in this case allows the system to extract the right cues from the environment in the right context.

We suggest that the study shows decouplability *in the dynamics* of the neural responses, and that this case exhibits the same pattern of explanation as the examples in the previous subsection. In particular, the nature of the learning depends on the environment in which learning occurs – it is the fact that color and motion vary in the environment that leads the PFC to implement selectivity for those variables. If you varied the training history, the system would track something different. What learning does, however, is establish a way of using or manipulating those responses that is functional for the task context. The task requires separating motion from color cues, and employing the right response in the right context. The solution that the system hits upon, if the analysis is correct, is the development of distinct line attractors, such that the context cue pushes the system to the appropriate one for the task context.

At the beginning of the training, monkeys are not able to respond successfully to the relevant cues. After the training, they are. What explains the difference is the way that learning organized internal neural resources so as to drive population dynamics. So, we have the same schema proposed in the previous section. Environmental variation and training history determine what is learned, but once the learning has taken place, the learned structure can generalize – the monkey's frontal cortex can respond, presumably, in a similar way to a degree of correlated motion it hasn't seen before, for instance. Of course, as we saw in the previous section, this generalization will only work so long as the task situation is appropriate. The ability to separate motion from color will not necessarily carry over to the ability to separate, say, timbre from pitch. While decoupling occurs, the success of behaviors guided by the internal changes will depend on novel task environments having a similar enough structure to those learned. Given

the specific causal contributions of both internal and external components, even neural dynamics are best understood in terms of a distributed locus of control.

5 Conclusion

We have argued that theoretical debates in both evolutionary biology and cognitive science have, at least implicitly, taken place between theorists positing different loci of control. Internalists place the locus of control internal to the (relevant subsystem of) the organism. Externalists place the locus of control outside of the (relevant subsystem of) the organism. We have contended that a productive trend is emerging in evolutionary biology which posits instead a distributed locus of control. Our conclusion in this paper is a normative one for debates in theoretical cognitive science: we should move beyond the internalism/externalism dichotomy by positing a distributed locus of control, and we should reconstrue key notions such as “mental representation” along those lines.

In closing, we stress two further advantages of positing a distributed locus of control, both of which stem from the fact that it is an *explanatory strategy*. It suggests that a phenomenon of interest will be best explained as the result of specific, non-redundant contributions from both internal and external components. Read as an explanatory strategy, positing a distributed locus of control retains a large amount of both methodological and metaphysical *flexibility*.

A view based on a distributed locus of control is *methodologically* flexible because it suggests that both internal and external components need to be independently identified and characterized. As such, if particular researchers (say, as in our case of the peppered moth in Section 3) are primarily interested in a particular internal or external component, a distributed locus of control view supports their primary methodological focus on that component. What it denies, however, is that that methodological focus comes along with *explanatory* primacy. The necessity of individuating and studying both internal and external components, on our view, does not equate to situating the locus of control with one or the other. In cognitive science, a distributed locus of control view thus supports the investigation of cognitive systems, as well as views of cognition that focus on the cultural or material contributions to cognition.⁵⁶

A view based on a distributed locus of control is *metaphysically* flexible because it is ultimately neutral with how to draw boundaries around the phenomenon of interest. What it requires is functional decomposability – i.e., that *components* have distinct causal contributions. This is compatible with a metaphysical view on which biological mechanisms extend into the environment, or even

one on which it is ultimately processes, and not mechanisms, that are at work (so long as the processes themselves are decomposable).⁵⁷ It is also compatible with a view of the mind on which the mind itself extends into the body or the environment.⁵⁸ So long as the explanatory suggestions are respected, a distributed locus of control view can accommodate each of these alternatives.

Hence, we suggest, views based on distributed loci produce the best overall result for theorizing in cognitive science. They avoid falsely locating the locus of control either internally or externally to the system. They support a range of methodological and theoretical approaches. And they allow for those approaches to be pursued independently of broader metaphysical issues. This is the best approach for a functioning science of organism and mind.

Notes

¹ Cf. C.K. WATERS, *Pluralist interpretation of gene-centered biology*; D. NOBLE, *Genes and causation*.

² Cf. J. WOODWARD, *Causation in biology: Stability, specificity, and the choice of levels of explanation*; C. KLEIN, *Brain regions as difference-makers*.

³ One way to put this difference is that an internal locus of control view will view external factors as *resources*, and vice versa for externalist views (cf. C. KLEIN, *Mechanisms, resources, and background conditions*). Resources are non-specific, not-persistent causal influences on a phenomenon. The “mechanism” versus “resource” distinction, however, does not line up precisely with positing internal versus external loci of control, since there might be internal (e.g., metabolic) resources necessary even on an internalist conception.

⁴ For a further exposition of holism, cf. D.C. BURNSTON, *Getting over atomism: Functional decomposition in complex neural systems*.

⁵ T.H. MORGAN, *The theory of the gene*, p. 306.

⁶ Cf. S. SARKAR, *From the Reaktionsnorm to the adaptive norm: The norm of reaction. 1909-1960*.

⁷ Cf. P. PORTIN, *The concept of the gene: Short history and present status*.

⁸ Cf. E. BOREK, *The code of life*.

⁹ Cf. H. GAFFRON, *The origin of life*.

¹⁰ Cf. L.M. COOK, I.L. SACCHERI, *The peppered moth and industrial melanism: Evolution of a natural selection case study*.

¹¹ Cf. A.E. VAN’T HOF, P. CAMPAGNE, D.J. RIGDEN, C.J. YUNG, J. LINGLEY, M.A. QUAIL, N. HALL, A.C. DARBY, I.J. SACCHERI, *The industrial melanism mutation in British peppered moths is a transposable element*.

¹² M.W. HO, P.T. SAUNDERS, *Beyond neo-Darwinism - An epigenetic approach to evolution*, here p. 573.

¹³ Cf. C.H. WADDINGTON, *The evolution of an evolutionist*.

¹⁴ Cf. R. FALK, *What is a gene?*.

¹⁵ Cf. E. JABLONKA, M.J. LAMB, *The inheritance of acquired epigenetic variations*.

¹⁶ Cf. B.K. HALL, *Genetics, epigenetics and environment*.

¹⁷ Cf. A. EACOCK, H.M. ROWLAND, A.E. VAN’T HOF, C.J. YUNG, N. EDMONDS, I.J. SACCHERI, *Adaptive colour change*

and background choice behaviour in peppered moth caterpillars is mediated by extraocular photoreception.

¹⁸ Cf. T.D. JOHNSTON, G. GOTTLIEB, *Neophenogenesis: A developmental theory of phenotypic evolution*.

¹⁹ Cf. D.H. FORD, R.M. LERNER, *Developmental systems theory: An integrative approach*; S. OYAMA, *The ontogeny of information: Developmental systems and evolution*.

²⁰ Cf. U. DEICHMANN, *Hierarchy, determinism, and specificity in theories of development and evolution*; K. STERELNY, K.C. SMITH, M. DICKISON, *The extended replicator*; P.E. GRIFFITHS, K. STOTZ, *Developmental systems theory as a process theory*.

²¹ Cf. K. STOTZ, *With "genes" like that, who needs an environment? Postgenomics's argument for the "ontogeny of information"*.

²² Cf. A. TRAMACERE, J. BICKLE, *Neuroepigenetics in philosophical focus: A critical analysis of the philosophy of mechanisms*.

²³ Cf. A. TRAMACERE, F. MAFESSONI, *Cognitive twists: The coevolution of learning and genes in human cognition*.

²⁴ Cf. I. BRIGANDT, *Explanation in biology: Reduction, pluralism, and explanatory aims*.

²⁵ C.K. WATERS, *Pluralist interpretation of gene-centered biology*, p. 193.

²⁶ Cf. K.N. LALAND, J. ODLING-SMEE, W. HOPPITT, T. ULLER, *More on how and why: Cause and effect in biology revisited*.

²⁷ Cf. A. BUSKELL, *Reciprocal causation and the extended evolutionary synthesis*; E.I. SVENSSON, *On reciprocal causation in the evolutionary process*.

²⁸ Cf. M. ALBERTI, J. MARZLUFF, V.M. HUNT, *Urban driven phenotypic changes: Empirical observations and theoretical implications for eco-evolutionary feedback*.

²⁹ Cf. K. LALAND, T. ULLER, M. FELDMAN, K. STERELNY, G.B. MÜLLER, A. MOCZEK, E. JABLONKA, J. ODLING-SMEE, G.A. WRAY, H.E. HOEKSTRA, D.J. FUTUYAMA, R.E. LENSKI, T.F.C. MACKAY, D. SCHLUTER, J.E. STRASSMANN, *Does evolutionary theory need a rethink? Point: Yes, urgently*.

³⁰ Cf. G.A. WRAY, H.E. HOEKSTRA, D.J. FUTUYAMA, R.E. LENSKI, T.F.C. MACKAY, D. SCHLUTER, J.E. STRASSMANN, *Does evolutionary theory need a rethink? Counterpoint: No, all is well*.

³¹ Cf. J.A. FODOR, *The language of thought*; A. NEWELL, H.A. SIMON, *Computer science as empirical inquiry: Symbols and search*.

³² Cf. W. BOONE, G. PICCININI, *The cognitive neuroscience revolution*; N. SHEA, *Representation in cognitive science*.

³³ Cf. S. GALLAGHER, *Enactivist interventions: Rethinking the mind*; D.D. HUTTO, M.D. KIRCHHOFF, E. MYIN, *Extensive enactivism*; E. THOMPSON, F.J. VARELA, *Radical embodiment: Neural dynamics and consciousness*.

³⁴ Cf. D.D. HUTTO, M.D. KIRCHHOFF, E. MYIN, *Extensive enactivism: Why keep it all in?*

³⁵ Cf. S. GALLAGHER, *Enactivist interventions*.

³⁶ Cf. J.L. BERMÚDEZ, *Thinking without words*; G. PICCININI, *Neurocognitive mechanisms*; W.M. RAMSEY, *Representation reconsidered*; M. ROWLANDS, *Arguing about representation*; R.D. RUPERT, *Representation and mental representation*.

³⁷ Cf., e.g., A.I. GOLDMAN, *A moderate approach to embodied cognitive science*; M. ROWLANDS, *Situated representation*.

³⁸ Cf. S. GALLAGHER, *Enactivist interventions: Rethinking the mind*.

³⁹ Cf. M. ROWLANDS, *Arguing about representation*.

⁴⁰ Cf. C.K. WATERS, *Pluralist interpretation of gene-centered biology*.

⁴¹ Cf. M. ROWLANDS, *Arguing about representation*.

⁴² Cf. R.D. RUPERT, *Representation and mental representation*.

⁴³ Cf. K. STOTZ, C. ALLEN, *From cell-surface receptors to higher learning: A whole world of experience*.

⁴⁴ Cf. D.C. BURNSTON, *Anti-intellectualism for the learning and employment of skill*; D.C. BURNSTON, *Contents, vehicles, and complex data analysis in neuroscience*.

⁴⁵ Cf. X. ZANG, T. GEYER, L. ASSUMPÇÃO, H.J. MÜLLER, Z. SHI, *From foreground to background: How task-neutral context influences contextual cueing of visual search*.

⁴⁶ Cf. D.M. WOLPERT, J.R. FLANAGAN, *Computations underlying sensorimotor learning*.

⁴⁷ For a summary, cf. D.C. BURNSTON, *Anti-intellectualism for the learning and employment of skill*.

⁴⁸ Cf. R.L. GOLDSTONE, B.J. ROGOSKY, R. PEVTZOW, M. BLAIR, *The construction of perceptual and semantic features during category learning*. For in depth philosophical analysis of these and similar cases, cf. D.C. BURNSTON, *How to think about higher-level perceptual contents*; D.C. BURNSTON, *Perceptual learning, categorical perception, and cognitive permeation*.

⁴⁹ Cf. R.L. GOLDSTONE, M. STEYVERS, *The sensitization and differentiation of dimensions during category learning*.

⁵⁰ Cf. M. JONES, R.L. GOLDSTONE, *The structure of integral dimensions: Contrasting topological and Cartesian representations*.

⁵¹ Cf. M. SILBERSTEIN, A. CHEREMO, *Constraints on localization and decomposition as explanatory strategies in the biological sciences*.

⁵² Cf., e.g., W. BECHTEL, *Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science*. For a more up-to-date discussion, cf. O. SHAGRIR, *Structural representations and the brain*.

⁵³ Cf. M. RIGOTTI, O. BARAK, M.R. WARDEN, X.-J. WANG, N.D. DAW, E.K. MILLER, S. FUSI, *The importance of mixed selectivity in complex cognitive tasks*.

⁵⁴ Cf. V. MANTE, D. SUSSILLO, K.V. SHENOY, W.T. NEWSOME, *Context-dependent computation by recurrent dynamics in prefrontal cortex*.

⁵⁵ For a fuller description, cf. D.C. BURNSTON, *Contents, vehicles, and complex data analysis in neuroscience*.

⁵⁶ Cf., e.g., R. HEERSMINK, J. SUTTON, *Cognition and the Web: Extended, Transactive, or Scaffolded?*; R. MENARY, *Mathematical cognition*; K. STOTZ, *Human nature and cognitive-developmental niche construction*.

⁵⁷ Cf. D.J. NICHOLSON, J. DUPRÉ, *Everything flows: Towards a processual philosophy of biology*; S.M. ROE, B. BAUMGAERTNER, *Extended mechanistic explanations*; D.C. BURNSTON, *Perceptual learning, categorical perception, and cognitive permeation*.

⁵⁸ Cf. M. KIRCHHOFF, *Extended cognition and constitution*; M. SPREVAK, *Extended cognition*.

Literature

ALBERTI, M., MARZLUFF, J., HUNT, V.M. (2017). *Urban driven phenotypic changes: Empirical observations and theoretical implications for eco-evolutionary feedback*. In: «Philosophical Transactions of the Royal Society B: Biological Sciences», vol.

- CCCLXXII, n. 1712, 20160029 – doi: 10.1098/rstb.2016.0029.
- BECHTEL, W. (1998). *Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science*. In: «Cognitive Science», vol. XXII, n. 3, pp. 295-317.
- BERMÚDEZ, J.L. (2017). *Thinking without words*, Oxford University Press, Oxford.
- BOONE, W., PICCININI, G. (2016). *The cognitive neuroscience revolution*. In: «Synthese», vol. CXCIII, n. 5, pp. 1509-1534.
- BOREK, E. (1969). *The code of life*, Columbia University Press, New York/London.
- BRIGANDT, I. (2013). *Explanation in biology: Reduction, pluralism, and explanatory aims*. In: «Science & Education», vol. XXII, n. 1, pp. 69-91.
- BURNSTON, D.C. (2021). *Anti-intellectualism for the learning and employment of skill*. In: «Review of Philosophy and Psychology», vol. XII, n. 3, pp. 507-526.
- BURNSTON, D.C. (2021). *Contents, vehicles, and complex data analysis in neuroscience*. In: «Synthese», vol. CXCIX, n. 1, pp. 1617-1639.
- BURNSTON, D.C. (2021). *Getting over atomism: Functional decomposition in complex neural systems*. In: «British Journal for the Philosophy of Science», vol. LXXII, n. 3, pp. 743-772.
- BURNSTON, D.C. (2022). *How to think about higher-level perceptual contents*. In: «Mind and Language», online first: 2022, November, 24th – doi: 10.1111/mila.12446.
- BURNSTON, D.C. (forthcoming). *Perceptual learning, categorical perception, and cognitive permeation*. In: «Dialectica».
- BUSKELL, A. (2019). *Reciprocal causation and the extended evolutionary synthesis*. In: «Biological Theory», vol. XIV, n. 4, pp. 267-279.
- COOK, L.M., SACCHERI, I.J. (2013). *The peppered moth and industrial melanism: Evolution of a natural selection case study*. In: «Heredity», vol. CX, n. 3, pp. 207-212.
- DEICHMANN, U. (2017). *Hierarchy, determinism, and specificity in theories of development and evolution*. In: «History and Philosophy of the Life Sciences», vol. XXXIX, n. 4, Art.Nr. 33 – doi: 10.1007/s40656-017-0160-3.
- EACOCK, A., ROWLAND, H.M., VAN'T HOF, A.E., YUNG, C.J., EDMONDS, N., SACCHERI, I.J. (2019). *Adaptive colour change and background choice behaviour in peppered moth caterpillars is mediated by extraocular photoreception*. In: «Communications Biology», vol. II, Art.Nr. 286 – doi: 10.1038/s42003-019-0502-7.
- FALK, R. (1986). *What is a gene?*. In: «Studies in History and Philosophy of Science Part A», vol. XVII, n. 2, pp. 133-173.
- FODOR, J.A. (1975). *The language of thought*, Harvard University Press, Cambridge (MA).
- FORD, D.H., LERNER, R.M. (1992). *Developmental systems theory: An integrative approach*, Sage, London / New York.
- GAFFRON, H. (1960). *The origin of life*. In: «Perspectives in Biology and Medicine», vol. III, pp. 163-212.
- GALLAGHER, S. (2017). *Enactivist interventions: Rethinking the mind*, Oxford University Press, Oxford.
- GOLDMAN, A.I. (2012). *A moderate approach to embodied cognitive science*. In: «Review of Philosophy and Psychology», vol. III, n. 1, pp. 71-88.
- GOLDSTONE, R.L., ROGOSKY, B.J., PEVTZOW, R., BLAIR, M. (2017). *The construction of perceptual and semantic features during category learning*. In: H. COHEN, C. LEFEBVRE (ed.), *Handbook of categorization in cognitive science*, Elsevier, Amsterdam, pp. 851-882, 2nd edition.
- GOLDSTONE, R.L., STEYVERS, M. (2001). *The sensitization and differentiation of dimensions during category learning*. In: «Journal of Experimental Psychology: General», vol. CXXX, n. 1, pp. 116-139.
- GRIFFITHS, P.E., STOTZ, K. (2018). *Developmental systems theory as a process theory*. In: D.J. NICHOLSON, J. DUPRE (eds.), *Everything flows: Towards a processual philosophy of biology*, Oxford University Press, Oxford/New York, pp. 25-245.
- HALL, B.K. (1992). *Genetics, epigenetics and environment*. In: B.K. HALL (ed.), *Evolutionary developmental biology*, Springer, Berlin/Dordrecht/Amsterdam, pp. 150-182.
- HEERSMINK, R., SUTTON, J. (2020). *Cognition and the web: Extended, transactive, or scaffolded?*. In: «Erkenntnis», vol. LXXXV, n. 1, pp. 139-164.
- HO, M.W., SAUNDERS, P.T. (1979). *Beyond neo-Darwinism - An epigenetic approach to evolution*. In: «Journal of Theoretical Biology», vol. LXXVIII, n. 4, pp. 573-591.
- HUTTO, D.D., KIRCHHOFF, M.D., MYIN, E. (2014). *Extensive enactivism: Why keep it all in?*. In: «Frontiers in Human Neuroscience», vol. VIII, Art.Nr. 706 – doi: 10.3389/fnhum.2014.00706.
- JABLONKA, E., LAMB, M.J. (1989). *The inheritance of acquired epigenetic variations*. In: «Journal of Theoretical Biology», vol. CXXXIX, n. 1, pp. 69-83.
- JOHNSTON, T.D., GOTTLIEB, G. (1990). *Neophenogenesis: A developmental theory of phenotypic evolution*. In: «Journal of Theoretical Biology», vol. CXLVII, n. 4, pp. 471-495.
- JONES, M., GOLDSTONE, R.L. (2013). *The structure of integral dimensions: Contrasting topological and Cartesian representations*. In: «Journal of Experimental Psychology: Human Perception and Performance», vol. XXXIX, n. 1, pp. 111-132.
- KIRCHHOFF, M. (2014). *Extended cognition and constitution: Re-evaluating the constitutive claim of extended cognition*. In: «Philosophical Psychology», vol. XXVII, n. 2, pp. 258-283.
- KLEIN, C. (2017). *Brain regions as difference-makers*. In: «Philosophical Psychology», vol. XXX, n. 1-2, pp. 1-20.
- KLEIN, C. (2018). *Mechanisms, resources, and background conditions*. In: «Biology and Philosophy», vol. XXXIII, n. 5, Art.Nr. 36 – doi: 10.1007/s10539-018-9646-y.
- LALAND, K., ULLER, T., FELDMAN, M., STERELNY, K., MÜLLER, G.B., MOCZEK, A., JABLONKA, E., ODLING-SMEE, J., WRAY, G.A., HOEKSTRA, H.E., FUTUYMA, D.J., LENSKI, R.E., MACKAY, T.F.C., SCHLUTER, D., STRASSMANN, J.E. (2014). *Does evolutionary theory need a rethink? Point: Yes, urgently*. In: «Nature», vol. DXIV, n. 7521, pp. 161-164.
- LALAND, K.L., ODLING-SMEE, J., HOPPITT, W., ULLER, T. (2013). *More on how and why: Cause and effect in biology revisited*. In: «Biology and Philosophy», vol. XXVIII, n. 5, pp. 719-745.
- MANTE, V., SUSSILLO, D., SHENOY, K.V., NEWSOME, W.T. (2013). *Context-dependent computation by recurrent dynamics in prefrontal cortex*. In: «Nature»,

- vol. DIII, n. 7474, pp. 78-97.
- MENARY, R. (2015). *Mathematical cognition: A case of enculturation*. In: T.M.J. WINDT (ed.), *Open Mind*, Open Project, Frankfurt a.M.
- MORGAN, T.H. (1926). *The theory of the gene*, Yale University Press, New Haven.
- NEWELL, A., SIMON, H.A. (1976). *Computer science as empirical inquiry: Symbols and search*. In: «Communications of the ACM», vol. XIX, pp. 113-126.
- NICHOLSON, D.J., DUPRÉ, J. (2018). *Everything flows: Towards a processual philosophy of biology*, Oxford University Press, Oxford.
- NOBLE, D. (2008). *Genes and causation*. In: «Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences», vol. CCCLXVI, n. 1878, pp. 3001-3015.
- OYAMA, S. (1986). *The ontogeny of information: Developmental systems and evolution*, Cambridge University Press, Cambridge.
- PICCININI, G. (2020). *Neurocognitive mechanisms: Explaining biological cognition*, Oxford University Press, Oxford.
- PORTIN, P. (1993). *The concept of the gene: Short history and present status*. In: «The Quarterly Review of Biology», vol. LXVIII, n. 2, pp. 173-223.
- RAMSEY, W.M. (2007). *Representation reconsidered*, Cambridge University Press, Cambridge.
- RIGOTTI, M., BARAK, O., WARDEN, M.R., WANG, X.-J., DAW, N.D., MILLER, E.K., FUSI, S. (2013). *The importance of mixed selectivity in complex cognitive tasks*. In: «Nature», vol. CDXCVII, n. 7451, p. 585-590.
- ROE, S.M., BAUMGAERTNER, B. (2017). *Extended mechanistic explanations: Expanding the current mechanistic conception to include more complex biological systems*. In: «Journal for General Philosophy of Science», vol. XLVIII, n. 4, pp. 517-534.
- ROWLANDS, M. (2009). *Situated representation*. In: M. AYDEDE, P. ROBBINS (eds.), *The Cambridge handbook of situated cognition*, Cambridge University Press, Cambridge, pp. 117-133.
- ROWLANDS, M. (2017). *Arguing about representation*. In: «Synthese», vol. CXCIV, n. 11, pp. 4215-4232.
- RUPERT, R.D. (2018). *Representation and mental representation*. In: «Philosophical Explorations», vol. XXI, n. 2, pp. 204-225.
- SARKAR, S. (1999). *From the Reaktionsnorm to the adaptive norm: The norm of reaction. 1909-1960*. In: «Biology and Philosophy», vol. XIV, n. 2, pp. 235-252.
- SHAGRIR, O. (2012). *Structural representations and the brain*. In: «British Journal for the Philosophy of Science», vol. LXIII, n. 3, pp. 519-545.
- SHEA, N. (2018). *Representation in cognitive science*, Oxford University Press, Oxford.
- SILBERSTEIN, M., CHEMERO, A. (2013). *Constraints on localization and decomposition as explanatory strategies in the biological sciences*. In: «Philosophy of Science», vol. LXXX, n. 5, pp. 958-970.
- SPREVAK, M. (2019). *Extended cognition*. In: E. CRAIG (ed.), *Routledge encyclopedia of philosophy*, Routledge, London/New York – doi: 10.4324/9780415249126-V049-1.
- STERELNY, K., SMITH, K.C., DICKISON, M. (1996). *The extended replicator*. In: «Biology and Philosophy», vol. XI, n. 3, pp. 377-403.
- STOTZ, K. (2006). *With “genes” like that, who needs an environment? Postgenomics’s argument for the “ontogeny of information”*. In: «Philosophy of Science», vol. LXXIII, n. 5, pp. 905-917.
- STOTZ, K. (2010). *Human nature and cognitive-developmental niche construction*. In: «Phenomenology and the Cognitive Sciences», vol. IX, n. 4, pp. 483-501.
- STOTZ, K., ALLEN, C. (2012). *From cell-surface receptors to higher learning: A whole world of experience*. In: K.S. PLAISANCE, T.A.C. REYDON (eds.), *Philosophy of behavioral biology*, Springer, Berlin/Dordrecht/The Hague, pp. 85-123.
- SVENSSON, E.I. (2018). *On reciprocal causation in the evolutionary process*. In: «Evolutionary Biology», vol. XLV, n. 1, pp. 1-14.
- THOMPSON, E., VARELA, F.J. (2001). *Radical embodiment: Neural dynamics and consciousness*. In: «Trends in Cognitive Sciences», vol. V, n. 10, pp. 418-425.
- TRAMACERE, A., BICKLE, J. (2023). *Neuroepigenetics in philosophical focus: A critical analysis of the philosophy of mechanisms*. In: «Biological Theory», online first: 2023, March 23th – doi: 10.1007/s13752-023-00435-3.
- TRAMACERE, A., MAFESSONI, F. (2022). *Cognitive twists: The coevolution of learning and genes in human cognition*. In: «Review of Philosophy and Psychology», online first: 2022, December 21th – doi: 10.1007/s13164-022-00670-w.
- VAN’T HOF, A.E., CAMPAGNE, P., RIGDEN, D.J., YUNG, C.J., LINGLEY, J., QUAIL, M.A., HALL, N., DARBY, A.C., SACCHERI, I.J. (2016). *The industrial melanism mutation in British peppered moths is a transposable element*. In: «Nature», vol. DXXXIV, n. 7605, pp. 102-105.
- WADDINGTON, C.H. (1977). *The evolution of an evolutionist*. In: «Journal of the History of Biology», vol. X, n. 2, pp. 369-370.
- WATERS, C.K. (2006). *Pluralist interpretation of gene-centered biology*. In: S.H. KELLERT, H.E. LONGINO, C.K. WATERS (eds.), *Scientific pluralism*, Minnesota University Press, Minneapolis, pp. 190-214.
- WOLPERT, D.M., FLANAGAN, J.R. (2016). *Computations underlying sensorimotor learning*. In: «Current Opinion in Neurobiology», vol. XXXVII, n. 1, pp. 7-11.
- WOODWARD, J. (2010). *Causation in biology: Stability, specificity, and the choice of levels of explanation*. In: «Biology and Philosophy», vol. XXV, n. 3, pp. 287-318.
- WRAY, G.A., HOEKSTRA, H.E., FUTUYAMA, D.J., LENSKI, R.E., MACKAY, T.F.C., SCHLUTER, D., STRASSMAN, J.E. (2014). *Does evolutionary theory need a rethink? Counterpoint: No, all is well*. In: «Nature», vol. DXIV, n. 7521, pp. 161-164.
- ZANG, X., GEYER, T., ASSUMPÇÃO, L., MÜLLER, H.J., SHI, Z. (2016). *From foreground to background: How task-neutral context influences contextual cueing of visual search*. In: «Frontiers in Psychology», vol. VII, Art.Nr. 825 - doi: 10.3389/fpsyg.2016.00852.

RICERCHE

The mind-body problem in philosophy and the cognitive sciences

Sandro Nannini^(a)

Ricevuto: 14 maggio 2022; accettato: 7 aprile 2023

Abstract Here, I examine the main philosophical solutions to the mind-body problem distinguishing between “historicist” solutions that (more or less clearly) separate philosophy from science and solutions that instead result from a double “cognitive turn”, and see “continuity” between philosophy of mind and the cognitive sciences. The “historicist” solutions include ontological dualism (together with “skepticism” and “new mysterianism”), epistemological dualism, subjective idealism, and absolute idealism. In this group, transcendental idealism, phenomenology, and neutral monism are the solutions most open to a dialogue between philosophy and science. The “naturalistic” solutions can be divided into four groups: (1) behaviorism (psychological, logical, philosophical-analytical behaviorism); (2) materialism (identity theory, physicalism); (3) “weak naturalism” (functionalism, anomalous monism, “biological naturalism”, liberal naturalism, emergentism); (4) “strong naturalism” (“cognitive neo-evolutionism”, eliminativism). These offer a physicalist-eliminative solution to the mind-body problem (here called “soft physicalistic eliminativism”) that allows for more continuity between philosophy of mind and the cognitive sciences.

KEYWORDS: Mind/Body-Problem; Cognitive Science; Cognitive Turn; Naturalism; Eliminativism

Riassunto *Il problema mente-corpo in filosofia e le scienze cognitive* - Vengono esaminate le principali soluzioni filosofiche al problema mente-corpo al fine di distinguere le soluzioni di origine storicistica che separano (più o meno chiaramente) filosofia e scienza da quelle che invece, nate da una duplice “svolta cognitiva”, favoriscono la “continuità” tra filosofia della mente e scienze cognitive. Le soluzioni del primo gruppo sono anzitutto il dualismo ontologico (insieme allo “scetticismo” e al “neo-misterianismo”), il dualismo epistemologico, l’idealismo soggettivo e l’idealismo assoluto. L’idealismo trascendentale, la fenomenologia e il monismo neutrale sono invece soluzioni più aperte al dialogo tra filosofia e scienza. Le soluzioni del secondo gruppo, quelle “naturalistiche”, vengono suddivise in quattro grandi gruppi: (1) il comportamentismo (psicologico, logico e filosofico-analitico); (2) il materialismo (teoria dell’identità e fisicalismo); (3) il “naturalismo debole” (funzionalismo, monismo anomalo, “naturalismo biologico”, naturalismo liberalizzato e emergentismo); (4) il “naturalismo forte” (“neo-evoluzionismo cognitivo” ed eliminativismo). In conclusione, una soluzione di tipo fisicalistico-eliminativo (qui chiamata “eliminativismo fisicalistico moderato”) al problema mente-corpo è quella che più di altre favorisce la continuità tra filosofia della mente e scienze cognitive.

PAROLE CHIAVE: Problema mente/corpo; Scienza cognitiva; Svolta cognitiva; Naturalismo; Eliminativismo

^(a)Dipartimento di Scienze Sociali, Politiche e Cognitive, Università degli Studi di Siena, Palazzo San Niccolò, via Roma, 56 - 53100 Siena (IT)

E-mail: sandro.nannini@unisi.it (✉)



1 Philosophy and the cognitive sciences

IN *THE MIND'S NEW SCIENCE*, Howard Gardner called philosophy, psychology, artificial intelligence, linguistics, anthropology, and neuroscience «the cognitive sciences».¹ Gardner argued that an intense and fruitful dialogue was developing between philosophy and the cognitive sciences, despite their differences.² This dialogue has now been going on for decades and has led to good results. For instance, new cognitive sciences such as neuroethics³ and neuroaesthetics⁴ have been developed from traditional philosophical disciplines such as ethics and aesthetics. However, some basic theoretical issues still remain unclear. First of all, it is evident that philosophy, despite often cooperative relationships with all sciences, is neither science nor history, art, literature, myth or religion. Philosophy is simply... philosophy! Philosophy does not study a part of reality, for example the human world or human history. Philosophy is rational reflection on all other forms of knowledge or intellectual activity.⁵

As regards the relationship between philosophy and the cognitive sciences, it must first be noted that the attitude of philosophers towards the scientific study of cognitive processes is distributed along a continuum between two poles established by the “historicists” and positivists in the 19th century. At one pole, are the heirs of “historicism”⁶ (idealists, Marxists, existentialists, “hermeneuts” or postmodernists). They consider philosophy to be a humanistic discipline outside or even, following G.W.F. Hegel (1770-1831), superior to science. At the other pole, are the “naturalists”, the heirs of positivism and supporters of “scientific naturalism”⁷ today, who see philosophy as a kind of knowledge destined to either be replaced by or at least deeply connected to the sciences.

The first attitude does not allow for any real and intense collaboration between philosophers and cognitive scientists. It is typical of those philosophers, theologians, and more generally, scholars in the humanities who see the reduction of all mental phenomena (including the products of intellectual activity such as art, literature and science) to simple neuronal processes or to activities that could even be performed by a robot as an attack on the dignity and creativity of man and a rejection of the moral and cultural values that constitute human civilization.⁸

The second attitude, inaugurated in the 19th century by the positivists, is today supported by “naturalists” in the sense indicated above. It has proved to be more productive in view of fomenting collaboration between philosophers and cognitive scientists. In particular, in the 1960s, some philosophers of mind and many cognitive scientists began an intense dialogue aimed at founding, to use Gardner’s words, «the mind’s new sci-

ence».⁹ This cooperation between philosophy of mind and the cognitive sciences (particularly, artificial intelligence and cognitive psychology) brought about the “first cognitive turn”.¹⁰ The 1990s saw a “second cognitive turn”, this time based on cooperation between philosophy of mind and cognitive neuroscience.¹¹

Today’s solutions to the mind-body problem, heirs to these “cognitive turns”, have led to two kinds of “cognitive naturalism”.¹² Cognitive naturalism, understood in this way, still follows the model offered by W.v.O. Quine’s “naturalized epistemology”, an ontological and epistemological framework of a philosophical kind within which the various cognitive sciences can “naturalize” different aspects of mental activity by explaining them in a rigorously scientific-empirical way.¹³

Both these basic conceptions on the nature of mind – the former anti-reductionist and anti-naturalistic, the latter reductionist and naturalistic – are reflected in the orientations still prevalent nowadays with regard to solutions given to the so called “mind-body problem”.¹⁴ In order to clarify the relationship between philosophy and the cognitive sciences it is worthwhile providing a brief systematic review of the solutions philosophers have given to the mind-body problem from antiquity to the present day¹⁵ and the way in which these solutions have been accepted or criticized by cognitive scientists.

In other words, this review has a double purpose: on the one hand, to offer a survey of the main solutions that have been given and are still given today to the mind-body problem; on the other hand, to highlight how the last of these solutions, that is, eliminativism (or better, “soft physicalistic eliminativism”) is the most convincing solution and the most suitable ontological-epistemological framework for promoting interdisciplinary and multidisciplinary studies between the various cognitive sciences.

2 Ontological dualism, skepticism and new mysterianism

According to ontological *dualism*,¹⁶ the soul and the body are two substances or two independent and incompatible properties. Ontological dualism, although not so called, was clearly formulated by R. Descartes in the 17th century. The mind (whose concept replaces the concept of soul in Descartes) is *res cogitans* [thinking thing] and therefore it is thinking and not extended. The body is *res extensa* [extended thing]¹⁷ and therefore it is instead extended and not thinking.¹⁸

The Cartesian dualism can also be called *interactionism* because it admits causal mind-body interaction. It has met with enormous success in modern philosophy, in the 20th century it was accepted by important philosophers and neuroscien-

tists such as K.R. Popper and J. Eccles¹⁹ and is still supported by several philosophers today.²⁰ However, nowadays interactionism enjoys very little success both among cognitive scientists (in particular, neuroscientists)²¹ and among naturalistic philosophers of mind²² because it is difficult to reconcile it with contemporary physics. A causal interaction between the mind and the body, when considered a causal interaction between something physical and something not physical, would violate the principle of energy conservation and therefore would violate the principle of the Causal Closure of the Physical World according to which only physical events can cause other physical events.²³ A mental event external to the physical world would belong to a dimension of reality extraneous to physical reality. Therefore, if a mental event of this kind were the cause of brain processes it would inject a certain amount of energy into the physical world and thereby violate the principle of energy conservation. Moreover, not even the theory of relativity can avoid this consequence since it too accepts the principle of conservation of mass-energy.²⁴

Quantum mechanics instead has offered support to those who want to justify interactionism. In fact, at least according to the Copenhagen Interpretation, quantum mechanics maintains that there is no observer-independent physical reality. Those who accept this interpretation are free to believe that the intervention of an observer determines those aspects of experimentally observable physical reality that the probabilistic equations of quantum mechanics leave undetermined.²⁵ This has brought about a lively discussion²⁶ for or against «a quantum mechanical argument for mind-body dualism».²⁷ However, although its promoters are serious and sometimes distinguished scientists, this hypothesis remains highly controversial.²⁸

In any case, the argument against interactionism based on the violation of the Causal Closure of the Physical World does not apply to the other two kinds of ontological dualism: epiphenomenalism and “parallelism”. *Epiphenomenalism*²⁹ argues that certain brain processes in addition to causing motor responses or other physical events also produce states of consciousness through a kind of causation called “mental causation” distinct from physical causation. According to the supporters of epiphenomenalism, these states are causally inert and cannot retroact on the brain that produced them. Their state of inertia implies that hypothesizing their existence is completely useless from a scientific point of view. Therefore, epiphenomenalism can be criticized in the light of Ockham’s Razor: “entities should not be multiplied beyond necessity”.

Parallelism was hypothesized in different forms by B. Spinoza and G.W. Leibniz in the 17th century. Parallelism argues that there is no causal interaction between minds and bodies but that

this does not prevent mental events from being perfectly synchronized with what happens in the body (particularly in the brain). The theological-speculative justifications for mind-body parallelism given by Spinoza and Leibniz are devoid of solid evidence. Even if the prominent neuroscientist A.R. Damasio recently championed Spinoza’s views,³⁰ they are generally considered of little interest by today’s cognitive scientists. However, in the 19th and 20th centuries, an empirical reformulation of Spinoza’s hypothesized parallelism of thought and extension met with considerable success under the name of “neutral monism”.³¹

Furthermore, the basic thesis of ontological dualism in all its forms, namely the ontological irreducibility of mental phenomena to physical phenomena, is still indirectly acknowledged by two groups of philosophers and neuroscientists. Although philosophers in the first group (let’s call them supporters of *skepticism* in the philosophy of mind) do not maintain mind-body dualism, they nevertheless argue (as did, for example, T. Nagel) that states of consciousness are subjective and therefore cannot be investigated in a scientific and objective way.³² But the results of the empirical investigations of behaviorists, physicalists, and naturalists militate against this thesis.³³

Philosophers in the second group (first and foremost, C. McGinn)³⁴ together with some neuroscientists (such as A. Benini)³⁵ have instead supported the so-called *new mysterianism* (a denomination they do not themselves accept). According to the “new mysterians”, there is certainly something in brain activities that could explain the emergence of consciousness but unfortunately the human mind is not capable of grasping this explanation. New mysterianism therefore appears to result from a renunciation of the advances made by the cognitive sciences in the last thirty years.

3 Epistemological dualism and liberal naturalism

Ontologically distinguishing the soul (or the mind) from the body implies that knowing such different “objects” would require sufficiently different methodologies. This is the fundamental thesis of *epistemological dualism*,³⁶ a thesis based on the distinction between *Geistes- und Naturwissenschaften* (human sciences and natural sciences) proposed by W. Dilthey in opposition to positivism in 1881.³⁷ This distinction is still present in the contemporary philosophical debate. Philosophers such as H. Putnam³⁸ and J.R. McDowell³⁹ have implicitly exploited W.v.O. Quine’s “naturalized epistemology”⁴⁰ to support a thesis contrary to Quine’s naturalism.⁴¹ They have stated that different sciences (or more generally different disciplines, including philosophy) have different and independent “ontological commitments” that are the basis for independent

world views which while at least partly incompatible are all equally legitimate.⁴²

Putnam in particular admitted that a «mentalist» description of the mind could account for the «qualitative character of sensations»,⁴³ claiming this would not be possible within the reductionist approaches of the natural sciences. Therefore, in regard to the mind-body problem, unlike interactionism, Putnam's epistemological dualism did not violate the Causal Closure of the Physical World because his mentalistic description of conscious states was independent of their naturalistic description and therefore did not violate any laws of nature. However, Putnam's epistemological dualism (or pluralism) led to cognitive relativism; his epistemological dualism (or pluralism) sacrificed the central thesis of "scientific realism"⁴⁴ that science must try to provide a true description of the real world that is unique and ontologically independent of the way in which it is described. While this thesis has been and continues to be the center of a heated dispute between realist and anti-realist philosophers,⁴⁵ it remains, at least implicitly, accepted by the majority of scientists.

The epistemological pluralism proposed in M. De Caro and D. Macarthur's *liberal naturalism*⁴⁶ owes much to Putnam but takes a more moderate stance on cognitive relativism. It aspires to a form of naturalism, albeit a form of naturalism that differs from scientific naturalism. It can be classified as a form of "weak naturalism"⁴⁷ which rejects Quine's «continuity» between science and philosophy. However, it instead requires «compatibility» between philosophical and scientific theories: mentalistic descriptions of psychological states and human actions must not violate any laws of nature.⁴⁸ Can De Caro and Macarthur keep this commitment to the principle of energy conservation while and still maintain that conscious states can produce effects on voluntary behavior despite being irreducible to brain processes that should respect all laws of nature? This seems doubtful.

4 Idealism and neutral monism

Among the monistic solutions to the mind-body problem that oppose dualism, the most akin to dualism itself is *idealism*⁴⁹ because of its anti-naturalistic and anti-reductionist character. Idealism was already present in antiquity under the form of Plotinus' neoplatonic "immaterialism" and in the modern age as G.W. Leibniz's "Monadology" but first developed thanks to classical empiricism, which claimed it was nonsense to speak of an external reality independent of the representation it is given. This empiricist approach to the external world has given rise to three distinct kinds of idealism: subjective idealism, transcendental idealism and phenomenology, and absolute idealism.

Subjective idealism, in the version presented by

G. Berkeley in the 18th century, argues that bodies as autonomous substances distinct from minds (or better, from "spirits") do not exist. Bodies exist only insofar as they are the inner objects of mental acts.⁵⁰ Something similar to subjective idealism has reappeared today among interpretations of quantum mechanics as a form of "anti-realism" similar to that one already examined and criticized which attempted to recover ontological dualism through microphysics.⁵¹ But despite this partial retrieval of actuality through quantum mechanics, subjective idealism remains a highly speculative and non-scientific conception of mind.

Absolute idealism, instead of considering bodies as the contents of perceptions and thoughts, considers nature to be a manifestation of spirit. Especially in Hegel's version, absolute idealism met with enormous success during the 19th century and the first half of the 20th century but it, too, is a speculative theory on a par with subjective idealism that lacks any empirical-scientific basis.

Transcendental idealism, the solution given by I. Kant to the mind-body problem at the end of the 18th century, fared and still fares better among psychologists and cognitive scientists than subjective idealism or absolute idealism. However, Kant's philosophy is still a form of idealism because according to Kant the external reality that we perceive through the senses or know through science (in particular through I. Newton's physics) is not reality in-itself but reality as it appears to us. We do not know the "thing-in-itself" but only the "phenomenon".⁵² More in detail, Kant thought that our "Gemüt" ("animus", "mens"), unlike Berkeley's "spirit", does not determine the content of sensitive intuitions and empirical concepts. Through the "pure intuitions" of space and time and the "pure concepts of intellect" (that is, the "categories"), it determines only their forms.⁵³

At the beginning of the 20th century, "Gestalt psychology" (C. von Ehrenfels, etc.) took the idea that empirical data are transformed into a coherent image of the world only if they are organized according to a priori forms provided by the human mind from Kant's transcendental idealism.⁵⁴ Furthermore, at the beginning of the 20th century a kind of epistemology similar in part to Kant's epistemology was proposed by E. Husserl, the founder of phenomenology.⁵⁵

Phenomenology, a current of thought that is still central in contemporary philosophy, nowadays arouses new interest among many cognitive scientists.⁵⁶ Moreover many phenomenologists have opened themselves up to a dialogue with cognitive psychology and cognitive neuroscience.⁵⁷ However, although the results of this dialogue are very interesting, Husserl's staunch anti-naturalism⁵⁸ is still an obstacle to the compatibility of phenomenology with a scientific approach to the study of mind (or better to the study of mind/brain).

*Neutral monism*⁵⁹ is a kind of monism opposed to dualism but like idealism it also differs from materialism.⁶⁰ In the 17th century, this solution of the mind-body problem was proposed by B. Spinoza to justify “mind-body parallelism”.⁶¹ Neutral monism and its variant – known as “double aspect theory” – were transformed into an empirical hypothesis in accordance with modern science by G.T. Fechner, R. Ardigò, E. Mach and W. James in the 19th century. In 1921, neutral monism had its best-known formulation in B. Russell’s *The Analysis of Mind*. According to Russell, minds and bodies are theoretical constructs obtained by combining the same sense data in two different ways.⁶² However, if minds and bodies are theoretical constructs obtained by combining the same sense data (therefore psychological data) in two different ways, Russell’s monism ceases to be neutral and instead leans towards subjective idealism.

The same criticism cannot be directed at the version of empiricist neutral monism proposed more recently by D. Chalmers. He uses quantum mechanics⁶³ to support his “naturalistic dualism of properties” and to hypothesize that the states of phenomenal consciousness and their neuronal correlates are two sides of the same coin and that their perfect correspondence is guaranteed by a psycho-physical law that according to Chalmers is an essential part of the fundamental laws of nature.⁶⁴ Therefore Chalmers’ monism is truly neutral and exhibits no inclination towards idealism. However, it too runs into the central criticism of neutral monism in general: there is little point in saying that mind and body are two sides of the same coin if one is not able to say in empirically well-founded terms what the coin is! Neutral monists run the risk of explaining “*obscurum per obscurius*” (that is, they risk providing an explanation that is even more obscure than the thing it purports to explain).

5 Psychological behaviorism, logical behaviorism, philosophical-analytical behaviorism

All the solutions to the mind-body problem examined so far have their roots in the history of philosophy from antiquity to the middle of the 19th century. *Psychological behaviorism*⁶⁵ (together with psychoanalysis)⁶⁶ is an exception. It finds its roots in a scientific turning point on the conception of mind that emerged among psychologists in the second half of the 19th century and then rose to prominence in the first half of the 20th century as Darwin’s *The descent of man*⁶⁷ pushed us to consider *Homo sapiens* as just another animal species among others and the study of man as falling within the natural sciences.

Contrary to all previous philosophical solutions to the mind-body problem, behavioral psychologists maintained and still maintain (insofar as they

are still present) that one must renounce all belief in the existence of presumed internal mental causes of manifest behavior and instead explain such behavior in scientific terms. To come to understand human behavior in scientific terms it is necessary to use empirical generalizations based on the model: “(observable) distal stimulus → (observable) motor response”.⁶⁸ According to psychological behaviorists, it is misleading to consider the mind as a mediator between sensory stimuli and voluntary actions as do introspective psychologists; on the contrary, it must be seen as a “black box” strictly placed between parentheses.

Psychological behaviorism experienced immense popularity in the first half of the 20th century (first, in the United States) but subsequently entered a profound crisis, although it has not completely disappeared. The first insurmountable objection to psychological behaviorism was raised by N. Chomsky in 1959. He pointed out that Skinner’s interpretation of language as «verbal behavior»⁶⁹ (that is, as a set of conditioned reflexes learned after birth) fails because Skinner did not take into account the fact that human beings could neither speak nor understand those who speak to them in early childhood and thus could not learn to do so in such a short time if they were not able to unconsciously follow the rules of an innate universal grammar that forms the basis of all possible natural languages.⁷⁰

Since the 1960s, criticisms directed at psychological behaviorism by the nascent cognitive sciences have moved in the same direction. According to cognitive scientists, there are no directly observable regularities between sensory stimuli and motor responses that can explain human behavior without recourse to hypothesized internal mental states that can act as intermediaries. In short, one needs to open the “black box” and study its functioning if one wants to understand human behavior.⁷¹

However, these criticisms of psychological behaviorism do not directly affect philosophical forms of behaviorism, that is, the logical behaviorism⁷² proposed by logical empiricists and the philosophical-analytical behaviorism⁷³ proposed by the supporters of ordinary language philosophy. These philosophical forms of behaviorism reject dualism like psychological behaviorism but unlike psychological behaviorism do not deny the existence of mental states or mental events, even considering them relevant when explaining for voluntary actions. Philosophical kinds of behaviorism apply a “linguistic turn”⁷⁴ (a turn typical of analytic philosophy with its consequent rejection of metaphysics) to the philosophy of mind and suggest that psychological concepts, far from referring to “ghostly entities”, refer to publicly observable behavioral dispositions of flesh and bone agents. However, the way in which logical behaviorism and philosophical-analytical behaviorism

have introduced the concept of behavioral disposition into their respective conceptions of the mind is quite different.

According to the “*logical behaviorism*”⁷⁵ of the first logical empiricists in the 1930s, for example the version given by R. Carnap in 1932,⁷⁶ mental states are reducible to behavioral dispositions only in an indirect way. According to Carnap, if Mr. *A* is excited, his state of excitement is identical to a state of his central nervous system. However, since brain states are not directly observable,⁷⁷ it is convenient to identify Mr. *A*’s brain state with a behavioral disposition to excitability which in turn can be implicitly defined by observable effects on Mr. *A*’s behavior: his hands shake and he responds aggressively to whatever is said to him, etc.⁷⁸ Therefore, Carnap’s logical behaviorism is a sort of “*physicalist behaviorism*”⁷⁹ according to which mental states are brain processes (the physicalist aspect of Carnap’s definition of mental states). However, brain processes can only be described indirectly as behavioral dispositions by observing the behavior they determine (the behavioral aspect of Carnap’s definition).

Supporters of *philosophical-analytical behaviorism* offer a definition of mental states that resembles Carnap’s definition in terms of its behavioral but not its physicalist aspect. This is particularly true of Ryle’s conception of the mind⁸⁰ (although he refused to call it a kind of behaviorism). While Ryle proposes a non-physicalist conception of mind, he shares Carnap’s thesis that mental states are knowable only if they are redescribed as behavioral dispositions.

For example – as Ryle has himself clarified – if you ask me why the person sitting next to me at the table passed me the salt and I reply that she did it out of courtesy, my explanation is dispositional and not causal. Her courtesy is not an occult mental cause that would have mysteriously caused a certain movement of her hand. Her courtesy is her tendency to be a kind person, that is, a person who under certain circumstances behaves in a certain way. More generally, the reason or motive for which one acts (or the intention with which one acts) is not a mental cause of the action but a behavioral disposition of the agent.⁸¹ However, Ryle’s behaviorism runs the risk of falling into a vicious circle. Saying that my neighbor passed me the salt out of courtesy is equivalent, in Ryle’s dispositional explanation, to saying that she behaved politely because she is a polite person. Therefore, if this is the first time I have ever seen her in my life, then the only proof I have of her being a polite person is that on this occasion she behaved politely!

As H. Putnam made clear, if I say for example that a person groans because of the pain she is experiencing, do I mean that, since pain can be defined as a disposition to groan, that person groans because she has the disposition to groan? Or do I

not rather mean that an internal state in her body is causing her to groan?⁸² Mentalism,⁸³ that is, the theory according to which there are always unobservable internal causes of observable behavior is undeniable.⁸⁴ Therefore philosophical-analytical behaviorists make a serious mistake when, in order to reject dualism, they reject mentalism as well. In fact, one can be a mentalist without being a dualist: it is sufficient to be a materialist!

6 Materialism: The mind-brain identity theory and physicalism

Both the *mind-brain identity theory*⁸⁵ and *physicalism*⁸⁶ escape Putnam’s objection to behaviorism because both these solutions to the mind-body problem admit the existence of unobservable mental causes for manifest behavior without falling back on dualism. Their proponents think that mental causes of behavior, even if they exist as required by mentalism, need not be “ghostly entities” extraneous to the physical world as dualists maintain. They are simply brain processes. Therefore, one can be a mentalist without being a dualist if one at least implicitly accepts the basic thesis of *materialism*: everything that is real is physical.⁸⁷ According to materialism, mental phenomena are real and, by causing bodily movements, they have effects in the physical world. However, in materialism, unlike dualistic interactionism, mental phenomena can cause voluntary actions without violating the principle of Causal Closure of the Physical World because they are themselves physical processes.

More specifically, according to the first version of mind-brain identity theory proposed by U.T. Place in 1954⁸⁸ and 1956,⁸⁹ “cognitive concepts” such as “knowing” and “believing” and “volitional” concepts such as “wanting” and “intending” refer to behavioral dispositions, as Ryle thought, but this is not the case for mental events such as being conscious or having sensations. This is because mental events, unlike mental states, are not behavioral dispositions but the inner bodily causes of behavior, that is, they are brain processes.⁹⁰

As Place reiterated in a later essay, «materialism as applied to mental events is a reasonable scientific hypothesis, which cannot be ruled out of court by a priori philosophical argument».⁹¹ In conclusion, Place and other supporters of both mind-brain identity (in particular J.J.C. Smart and D.K. Lewis)⁹² and the similar “materialism of the central state” theory (D. Armstrong),⁹³ led the philosophy of mind from philosophical-analytical behaviorism to materialism.

A similar path was also taken by those logical empiricists who, in the 1930s, extended O. Neurath’s physicalism to the philosophy of mind, claiming it was possible to translate the statements of any science into the language of physics.⁹⁴

These developments led H. Feigl to write, in

1934, that thanks to physicalism it was possible to definitively provide a purely logical-linguistic solution to the “psycho-physical problem”.⁹⁵ Nevertheless, by 1958, Feigl had revised his thinking, publishing an essay entitled *The “Mental” and the “Physical”*, where he recognized that, precisely because the mind-body problem was a scientific not metaphysical problem, it must also be an authentically ontological, not merely logical-linguistic problem.⁹⁶ Feigl, who sympathized with Place on this point, clarified in a 1967 *Postscript* to his 1958 essay that «mind-body problems cannot simply be made to disappear by purely linguistic maneuvers».⁹⁷

The road to scientific naturalism in the philosophy of mind was open. However, even in this second formulation, physicalism found it difficult to account for the emergence of phenomenal consciousness from brain activity or more generally to reconcile the “manifest image” of the world with its “scientific image”.⁹⁸

7 Weak naturalism: Functionalism, emergentism and other hybrid solutions

Feigl’s formulation of physicalism as an ontological-scientific theory was part of a more general trend towards the naturalization of epistemology promoted above all by Quine in the 1950s and 1960s. According to Quine, every scientific theory has its own “ontological commitments”. Therefore, in the name of “scientific realism”,⁹⁹ Quine thought it was the joint task of philosophers and scientists to “naturalize” research fields usually reserved for philosophical speculation by proposing adequate scientific theories.¹⁰⁰

However, neither behavioral psychology nor the brain studies available in the early 1950s were sufficient to give philosophy of mind a decisive turn towards the naturalization of the mental. This turning point took place thanks to the birth and first affirmation of the cognitive sciences in the second half of the 1950s, the 1960s and 1970s.¹⁰¹ Among the emerging cognitive sciences, those that made the greatest contribution to the philosophy of mind were artificial intelligence (AI)¹⁰² and cognitive psychology.¹⁰³ First of all, the cooperation of philosophy of mind with artificial intelligence gave birth to the “mind-computer analogy” and functionalism,¹⁰⁴ a solution to the mind-body problem that was dominant among philosophers of mind and cognitive scientists from the 1960s to the 1980s¹⁰⁵ and still counts numerous authoritative defenders.¹⁰⁶

Functionalism has been divided into multiple versions but all of them more or less directly share the key idea proposed by H. Putnam (who later became a critic of himself) in the essay *Minds and machines* published in 1960.¹⁰⁷ The human mind is the functional organization of brain activity that processes information contained in sensory inputs

and transforms them into motor outputs. According to functionalists, finding a solution to the mind-body problem that avoids the opposite errors of dualism and physicalism requires an intermediate level of analysis between folk-psychology and cognitive neuroscience. At this intermediate level it is possible to consider mental states as functional states that are implemented by but not identical to brain processes.

The key point is that implementation does not mean identity. As clarified in the theory of “multiple realizability”,¹⁰⁸ which marks the detachment of functionalism from physicalism, the same information processing function can be performed by different brain processes (in different animals, in different human beings or in the same person at different times) just as the same arithmetic operation can be performed both by an electronic calculator or by a mechanical calculator although the physical processes that perform it are completely different in the two cases. Because of this “multiple realizability”, functionalists argue that theorists, like physicalists, who identify mental states and mental events with brain processes, must admit that any mental state or mental event can be identified with multiple brain processes. However, this is obviously absurd because in that case a mental state or a mental event would be different from itself! To avoid this contradiction, functionalists argue that a mental state or a mental event is not identical to the different brain processes that implement it in different cases (“token physicalism”)¹⁰⁹ but instead to the information processing function common to all these processes.

Functionalism has met with great success and has produced a large number of scientific studies on various types of mental phenomena. However, firstly functionalism leaves in parentheses the problem of the emergence of consciousness and self-consciousness, since it defines the mental as a kind of information processing that can be performed within certain limits even by machines.¹¹⁰ Secondly, functionalism is a “third way” between dualism and physicalism and therefore creates a sort of epistemological dualism separating cognitive psychologists’ study of the mind from neuroscientists’ study of the brain. For this reason, functionalism can be considered the main form of “weak naturalism”, a kind of naturalism judged insufficient by all supporters of “strong naturalism”.¹¹¹

Similar considerations apply to De Caro and Macarthur’s “liberal naturalism”,¹¹² J.R. Searle’s “biological naturalism”,¹¹³ D. Davidson’s “anomalous monism”,¹¹⁴ “emergentism”¹¹⁵ and “non-reductive physicalism”.¹¹⁶ They are all “*hybrid solutions*” that claim to be both naturalistic and non-reductionist.

Among these hybrid solutions *emergentism*¹¹⁷ deserves separate consideration. It is a traditional solution to the mind-body problem which has enjoyed some success in the philosophy of mind today be-

cause it seems to reconcile a naturalistic conception of the mind with ontological non-reductionism. But in reality, this conciliation only results from an ambiguous definition of the concept of emergent properties. In fact, emergent properties can be understood in three ways:¹¹⁸

- Firstly, they can be understood as systemic properties emerging from the local properties of the elements of a physical system (think of the heat of a gas as the average kinetic energy resulting from the movements of its molecules). If consciousness is a holistic property of this kind and, at least in principle, we know the bridging principles that explain its emergence from neuronal activity, then emergentism is a kind of physicalism.
- Secondly, if you believe that the bridging principles that link consciousness to the physico-chemical processes of individual neurons are not sufficiently understood, then you can more prudently consider mental properties as virtual properties that are implemented by brain dynamics that still remain largely unknown. In this case, emergentism is a kind of functionalism.
- Thirdly, it can be thought that consciousness does emerge from brain activity but after its emergence has an ontological autonomy that allows it to causally retroact on brain activity. In this case, emergentism becomes a kind of ontological dualism and meets the same criticisms. In particular, it violates the *Causal Closure of the Physical World*.¹¹⁹

The trick that gives emergentism a certain success lies in confusing this third case with the first case or the second case. This creates the illusion that emergentism can reconcile the non-reductionism of dualism with the strong naturalism of physicalism or, at least, the weak naturalism of functionalism. The same criticism can be addressed to non-reductive physicalism which differs from emergentism more in detail than in substance.

8 Strong naturalism: Cognitive neo-evolutionism and eliminative materialism

Since the 1980s, four novel events in the cognitive sciences and the theory of biological evolution have promoted the “second cognitive turn in the philosophy of mind” mentioned above in the direction of *strong naturalism*.

Firstly, psychologists, mathematicians and computer scientists at the University of San Diego (CA) were able to use digital computers to simulate virtual machines capable of processing information in a parallel and distributed way,¹²⁰ thereby

founding “connectionism”.¹²¹ These virtual machines are artificial neural networks that simulate brain function and, in turn, offer neurologists a simplified mathematical model to reconstruct and explain brain activity.

Secondly, neuroscientists imitated biologists’ reconstruction of the human genome and developed a research program aimed at the reconstruction of the “connectome”, that is, a comprehensive map of neural connections in the human brain.¹²²

Thirdly, important neuroscientists have dealt with topics once reserved to philosophical speculation such as consciousness, the Self, or free will, giving rise to “cognitive neuroscience”.¹²³ This has brought about a lively discussion between philosophers and humanists in general, who object to reducing spiritual capacities to brain processes.¹²⁴

Fourthly, evolutionists have introduced interesting innovations in neo-Darwinian studies on the origins of social life, language, and culture in human beings.¹²⁵

Among philosophers of mind, some have used these new theories offered by computer scientists, neuroscientists, and neo-Darwinian biologists to promote strong naturalism: the philosophy of mind must be closely linked to the cognitive sciences. The two most important new solutions to the mind-body problem that have arisen in this context are D.C. Dennett’s “cognitive neo-evolutionism”¹²⁶ and eliminative materialism (or eliminativism),¹²⁷ particularly the version presented by Paul M. Churchland and his wife Patricia.¹²⁸

In the 1970s and 1980s, Dennett, starting from an approach pioneered by his mentor Ryle,¹²⁹ offered a solution to the mind-body problem, later called “homuncular functionalism” by W.G. Lycan.¹³⁰ Dennett’s solution is based on the notion that a virtual system can process information in an intelligent way if it is implemented by more stupid virtual subsystems that operate by trial and error. These subsystems are implemented by even more stupid subsystems and so on, until this tower of virtual systems is implemented by a physical system that works automatically according to the laws of nature.¹³¹

In the 1990s, Dennett refined his conception of the mental and applied the connectionist model of artificial neural networks to the human brain, that is, he argued that the human brain, like artificial neural networks, is a parallel self-programming machine that lacks any central processor (in digital computers the C.P.U.). In this way, Dennett has come to argue in particular that there is no Self, in the sense of a homunculus (material or immaterial) on the “inside” who turns his attention to some of our inner states thereby making them conscious to himself (and therefore to us) (“The Myth of the Cartesian Theater”). In reality, each of us is just a bundle of mental states that, like “demons”, compete with each other to find access to the motor system. Those

states that, in coalition with others, win the competition and control our behavior consolidate their victory by inventing an *ex post* “narrative” that justifies the behavior.¹³² This narrative represents the emergence of a phenomenal state of consciousness, while the Self, as the main character in this narrative, is its fictional “Center of Narrative Gravity”.¹³³

Subsequently, Dennett has consolidated his image of the human mind in the light of Dawkins’ neo-Darwinism and has developed a complex conception aimed at reconstructing its phylogeny.¹³⁴ This conception can be called “cognitive neo-evolutionism”. According to Dennett, the biological evolution that took place from prokaryotes to *Homo sapiens* by natural selection can be divided into four fundamental stages:

- “Darwinian creatures” survive only if they have got the right answer in their genes when they are faced with the challenges of external environment. If they do not have it, they die and in the long run their species becomes extinct.
- Some of these Darwinian creatures have become “Skinnerian creatures” by natural selection. Skinnerian creatures are equipped with a fairly large repertoire of possible motor responses to solve the problems that the environment poses to them and are able to learn by trial and error what the most effective responses are, but they run the risk of dying before they have found the right responses even if such responses are present in their repertoire.
- Some of these Skinnerian creatures have become, again by natural selection, “Popperian creatures”. Popperian creatures are able to build an internal model of the external environment and to test in a safe way in advance the effectiveness of the motor responses available to them without running the risks that Skinnerian creatures face.
- Finally (at least, for now) a Popperian creature, *Homo sapiens*, has become, again by natural selection, a “Gregorian creature” capable of enriching her internal model of the external environment by drawing on the massive information deposited there by other human beings (contemporaries or ancestors).¹³⁵

Dennett’s conception of the mind enjoys wide diffusion and great prestige nowadays, but it is sometimes criticized because it is too anthropocentric and assigns an excessive role to language in the evolution of *Homo sapiens*.¹³⁶ Furthermore, according to other naturalists it seems to lack sufficient empirical support from a neuroscientific point of view.¹³⁷

The other most important strong naturalism

solution to the mind-body problem, namely *eliminative materialism*, focuses on the ontogeny of the mind and its functioning in adult human beings. Eliminative materialism owes its name to Paul Churchland, who considers folk psychology concepts to be pre-scientific and advises they be eliminated and replaced with scientific concepts drawn from neuroscience.¹³⁸ Physicalists think that the mental states of folk psychology are reducible to brain processes; functionalists think that they are reducible to functional states. However, according to Paul Churchland both reductions are problematic. The only way to make psychology scientific, transforming it into psycho-neurology, is to replace the old pre-scientific concepts of folk-psychology with new scientific concepts. Something similar was done in chemistry when the concept of phlogiston was abandoned in favor of the concept of oxygen.¹³⁹

More in detail, Paul Churchland has shown how all mental states (perceptions, memories and so on) can be reduced to higher order properties of brain dynamics that can be represented by “vector coding” (or “vector processing”), that is, by means of mathematical matrices that transform the neuronal patterns of sensory input into the neuronal patterns of motor output in a certain vector space according to the model offered by artificial neural networks. In this way, Paul Churchland has shown in his book *The Engine of Reason* how artificial neural networks can simulate perceptions (“taste coding”, “color coding”, “smell coding”), recognize faces (“face coding”) or patterns of any kind,¹⁴⁰ simulate long-term memory and short-term memory,¹⁴¹ and acquire the ability to correctly read words written in English.¹⁴² Moreover, comparing brain functioning to that of artificial neural networks provides a first glimpse of how the human brain implements phenomenal consciousness through the activity of the intralaminar nucleus of the thalamus.¹⁴³

Eliminative materialism has met with many objections.¹⁴⁴ Firstly, that it refers to a cognitive psycho-neurology of the future which is mere wishful thinking so far and, secondly, that the elimination of phenomenal consciousness is nonsense. After all, if I am awake, I necessarily feel awake!

However, both objections are invalid. Firstly, it is not true that the cognitive psycho-neurology proposed by the Churchlands is mere wishful thinking. As clarified above, in *The engine of reason* already in 1995 Paul Churchland provided many concrete examples of how various mental phenomena can be interpreted as properties of brain dynamics and can be simulated, at least in part, through artificial neural networks. In the last forty years, Patricia Churchland has founded and developed a new science, “neurophilosophy”, and has written several books that show how many themes addressed in epistemology, ethics, psy-

chology, as well as those that consider the genetic bases of human societies can be approached “brain-wise”, that is, from the point of view of cognitive neuroscience.¹⁴⁵

Secondly, the widely shared accusation that the Churchlands believe phenomenal consciousness does not exist and we humans are zombies is completely unfounded. As the Churchlands have made clear, they do not deny the fact that «there is a nontrivial difference between being asleep and being awake».¹⁴⁶ They want to eliminate only the philosophical-Cartesian concept of consciousness and replace it with the description of many neurological “mechanisms”. The facts that lead to the philosophical-Cartesian concept of consciousness are real, it is just that the concept groups and describes them inappropriately.¹⁴⁷ With respect to consciousness, the Churchlands are fundamentally physicalists, not eliminativists!

9 Towards a “soft physicalistic eliminativism”

However, as the Churchlands themselves have recognized, this casts a shadow on the adequacy of the expression “eliminative materialism”.¹⁴⁸ Name aside, it is possible to make slight corrections and further develop their eliminativism in order to formulate a solution to the mind-body problem that can be called “*soft physicalistic eliminativism*”. It is based on the following points:¹⁴⁹

- Folk psychology, cognitive psychology, and cognitive neuroscience formulate their theories in three distinct “discourse universes” which have three distinct types of ontological commitments: mental states, psycho-functional states and neurological states.
- To avoid falling into cognitive relativism and epistemological dualism¹⁵⁰ virtual reality must be attributed to mental states only to the extent that they are reduced to psycho-functional states implemented by physico-chemical processes (these are the only “things” that enjoy effective reality as recognized by physicalists).
- Psycho-functional states are a partial and idealized redescription of mental states in the language of scientific psychology. They are virtual functional states completely implemented by brain processes. The aspects of mental states eliminated in a psycho-functional redescription are not neurologically implemented and therefore are not real, they are *ficta* (this is a kind of “soft physicalistic eliminativism”). However sometimes (not always!) *ficta* cannot be eliminated from the scientific description of the mind because, although they are not real, human agents spontaneously do believe that they

are real! These erroneous beliefs about the existence of some *ficta* are neurologically implemented; therefore, they are real in a physicalistic sense and can be functionally effective. For example, let us assume, by hypothesis, that we human beings do not enjoy free will.¹⁵¹ Even in this case, it is a part of our psychic normality to feel that we are free agents, and this fallacious feeling of “free agency” is real because it is neurologically implemented and supports our mental health.¹⁵²

- In accordance with physicalism, each psycho-functional state can be defined in such a way as to be identical to a higher order property of brain dynamics if these dynamics are described in an appropriate abstract way (e.g. by tables, graphs or various kinds of vector coding).¹⁵³ Each of these higher order properties is a holistic property of the brain that can emerge from the activity of different neuronal groups (in accordance with the theory of multiple realizability). Therefore in this way the theory of multiple realizability becomes compatible with eliminative materialism and mind-brain identity theory: each mental state of folk psychology, once purified of its *ficta* and reduced to a psycho-functional state of cognitive psychology, becomes identical to a single higher level property of brain dynamics; this single property can in turn be implemented by different neuronal processes on different occasions or in different persons.

In conclusion, the interdisciplinary ontology of the mental sketched above which corrects physicalism in the light of eliminativism and “homuncular functionalism” appears to be the philosophical solution to the mind-body problem that, more than others, maintains continuity with the cognitive sciences in general and cognitive neuroscience in particular.

It may perhaps be objected that this solution to the mind-body problem presupposes, in the wake of Quine’s naturalized epistemology,¹⁵⁴ a continuity between philosophy and science without offering any a priori philosophical argument in its favor. Now, this is undoubtedly true. But I can reply that if one accepts the continuity between philosophy and science one must also accept that no philosophical argument can a priori decide any dispute that arises on scientific grounds (to think otherwise would mean having more or less explicitly accepted the transcendental idealism of Kant and Husserl).¹⁵⁵ Therefore not even the validity of the very principle of continuity between philosophy and science can be established through an a priori philosophical argument. Even this continuity, applied to the particular case of the search for a fruitful collaborative relationship between the philoso-

phers of mind and cognitive scientists, can prove to be gradually and pragmatically feasible with success only thanks to the search for something like N. Goodman's "reflective equilibrium".¹⁵⁶ This equilibrium requires that philosophers of mind suggest an ontological-epistemological framework capable of favoring interdisciplinary research in the field of "the science of mind" while the cognitive scientists evaluate whether this philosophical suggestion is of any use to them. It is with the intention of fostering a reflexive equilibrium of this kind that I have argued here for the superiority of an eliminativistic-physicalistic solution to the mind-body problem.

Notes

- ¹ H. GARDNER, *The mind's new science: A history of the cognitive revolution*, p. 7 (moreover, see p. 37 for the relations between the six cognitive sciences). According to some interpreters, their process of unification into a single science has already been completed (cf., e.g., A. STEPHAN, S. WALTER (Hrsg.), *Handbuch Kognitionswissenschaft*), according to others not yet (cf., e.g., C.P. SOBEL, *Cognitive sciences: An interdisciplinary approach*).
- ² H. GARDNER, *The mind's new science*, chapter 4.
- ³ A. ROSKIES, *Neuroethics*; E. RACINE, *Pragmatic neuroethics*; P.S. CHURCHLAND, *Braintrust*; P.S. CHURCHLAND, *Conscience*; G. CORBELLINI, E. SIRGIOVANNI, *Tutta colpa del cervello*.
- ⁴ J. ROBSON, G. CURRIE, *Aesthetics and cognitive science*.
- ⁵ T. NAGEL, *What does it all mean?*, p. 4.
- ⁶ P.D. HAMILTON, *Historicism*; P. ROSSI, *Lo storicismo tedesco contemporaneo*.
- ⁷ D. PAPINEAU, *Naturalism*.
- ⁸ Cf., e.g., the theologian Hans Küng, *Der Anfang aller Dinge*.
- ⁹ Cf. H. GARDNER, *The mind's new science*.
- ¹⁰ S. NANNINI, *The second cognitive turn in the philosophy of mind*, pp. 2-4
- ¹¹ *Ibid.*, pp. 4-5.
- ¹² Cf. S. NANNINI, H.J. SANDKÜHLER (eds.), *Cognitive naturalism in the philosophy of mind*; S. NANNINI, *Naturalismo cognitivo*; S. NANNINI, *Mente e cervello*.
- ¹³ Cf. *infra*, §§7-8.
- ¹⁴ Cf. M. ROWLANDS, *The mind-body problem*.
- ¹⁵ For a wider historical introduction to the philosophy of mind cf., e.g., S. NANNINI, *L'anima e il corpo*.
- ¹⁶ H. ROBINSON, *Dualism*.
- ¹⁷ Descartes thought that the vacuum does not exist and confused space and matter under the name of extension.
- ¹⁸ Cf. R. DESCARTES, *Meditations on first philosophy*.
- ¹⁹ K.R. POPPER, J.C. ECCLES, *The self and its brain*; K.R. POPPER, *Knowledge and the body-mind problem*.
- ²⁰ Cf., e.g., E.J. LOWE, *An introduction to the philosophy of mind*; R. SWINBURNE, *From mental/physical identity to substance dualism*; A. LAVAZZA (a cura di), *L'uomo a due dimensioni. Il dualismo mente-corpo oggi*.
- ²¹ Cf. first of all G. EDELMAN, *Wider than the sky*, chapter 7; G. ROTH, N. STRÜBER, *Wie das Gehirn die Seele macht*.
- ²² Cf., e.g., P.S. CHURCHLAND, *Touching a nerve*, pp. 44-53; D.C. DENNETT, *From bacteria to Bach and back*, pp. 20-23.
- ²³ D. SPURRETT, D. PAPINEAU, *A note on the completeness of "physics"*.
- ²⁴ A. EINSTEIN, *Relativity: The special and general theory*, p. 54 ff.
- ²⁵ Cf. first of all J. VON NEUMANN, *Mathematical foundations of quantum mechanics*; E.P. WIGNER, *Remarks on the mind-body question*.
- ²⁶ Cf. about this discussion H. ROBINSON, *Dualism*, §3.1
- ²⁷ J.A. BARRETT, *A quantum mechanical argument for mind-body dualism*. Cf. also H.P. STAPP (ed.), *Mind, matter and quantum mechanics*; M. JIBU, K. YASUE, *Quantum brain dynamics and consciousness: An introduction*.
- ²⁸ Cf., e.g., S. YU, D. NIKOLIĆ, *Quantum mechanics needs no consciousness*.
- ²⁹ W. ROBINSON, *Epiphenomenalism*.
- ³⁰ Damasio maintains that Spinoza was a «proto-biologist» and accepted a materialistic solution to the mind-body problem (cf. A.R. DAMASIO, *Looking for Spinoza*). However, this is not true. Spinoza was a speculative neutral monist influenced by neoplatonism (cf. S. NANNINI, *Alla riscoperta di Spinoza*).
- ³¹ Cf. *infra*, §4.
- ³² T. NAGEL, *What is it like to be a bat?*.
- ³³ Cf. *infra*, §§5-8.
- ³⁴ Cf. C. MCGINN, *The character of mind*; C. MCGINN, *The Problem of Consciousness*; C. MCGINN, *The mysterious flame. Conscious minds in a material world*.
- ³⁵ Cf. A. BENINI, *Che cosa sono io? Il cervello alla ricerca di se stesso*.
- ³⁶ Cf. R.W. SELLARS, *Epistemological dualism vs. metaphysical dualism*.
- ³⁷ Cf. W. DILTHEY, *Introduction to the human sciences*.
- ³⁸ H. PUTNAM, *Reason, truth and history*, p. 79.
- ³⁹ Cf. J.H. MCDOWELL, *Mind and world*.
- ⁴⁰ W.V.O. QUINE, *Ontological relativity and other essays*; W. V. O. QUINE, *Theories and things*, p. 21.
- ⁴¹ On Quine's naturalism cf. *infra*, §7.
- ⁴² H. PUTNAM, *The content and appeal of "naturalism"*; J. MCDOWELL, *Naturalism in the philosophy of mind*.
- ⁴³ H. PUTNAM, *Reason, truth and history*, p. 79.
- ⁴⁴ A. CHAKRAVARTTY, *Scientific realism*.
- ⁴⁵ A. MILLER, *Realism*.
- ⁴⁶ M. DE CARO, D. MACARTHUR (eds.), *Naturalism in question*.
- ⁴⁷ Cf. *infra*, §7.
- ⁴⁸ M. DE CARO, D. MACARTHUR, *Naturalismo scientifico e naturalismo liberalizzato*, p. XV.
- ⁴⁹ P. GUYER, R.P. HORSTMANN, *Idealism*.
- ⁵⁰ G. BERKELEY, *A treatise concerning the principles of human knowledge*.
- ⁵¹ Cf. *supra*, § 2.
- ⁵² I. KANT, *Preface to the second edition (1787)*, in I. KANT, *Critique of pure reason*, pp. 106-124.
- ⁵³ Cf. the first two parts of I. KANT, *Critique of pure reason*.
- ⁵⁴ B. SMITH, *Gestalt theory. An essay in philosophy*.
- ⁵⁵ Cf. E. HUSSERL, *Logical investigations*; E. HUSSERL, *Ideas pertaining a pure phenomenology and a phenomenological philosophy*, 3 voll.
- ⁵⁶ Cf., e.g., A. SETH, *Being you. A new science of consciousness*.
- ⁵⁷ The first link between phenomenology and naturalism in the philosophy of mind can be found in W.S. SELLARS, *Empiricism and the philosophy of mind*. For further developments cf., e.g., J. PETITOT, F.J. VARELA, B. PACHOUD, J.-M. ROY (eds.), *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*; S. GALLAGHER, D. ZAHAVI, *The phenomenological mind*.

- ⁵⁸ E. HUSSERL, *Logical investigations*, p. 2.
- ⁵⁹ L. STUBENBERG, *Neutral monism*.
- ⁶⁰ Cf. *infra*, § 6.
- ⁶¹ Cf. *supra*, § 2.
- ⁶² B. RUSSELL, *The analysis of mind*, pp. 103-105.
- ⁶³ Cf. *supra*, § 2.
- ⁶⁴ D. CHALMERS, *The conscious mind*; D. CHALMERS, *The character of consciousness*.
- ⁶⁵ G. GEORGE, *Behaviorism*.
- ⁶⁶ Psychoanalysis, like psychological behaviorism, has also marked and still marks a profound theoretical turning point in the conception of the human mind and has recently established connections with the cognitive sciences (Cf., e.g., M. MANCIA (ed.), *Psychoanalysis and neuroscience*). However, psychoanalysis has not proposed any new fundamental solution to the mind-body problem: Freud was a materialist in his youth, while in his maturity he arrived at his own original form of functionalism with his distinction between “Es”, “Ich” and “Über-Ich”, (cf., e.g., S. NANNINI, *Mente e cervello*, pp. 157-167). Therefore, psychoanalysis will not be discussed here.
- ⁶⁷ C. DARWIN, *The descent of man, and selection in relation to sex*.
- ⁶⁸ Cf. among the most important behaviorist psychologists: J.B. WATSON, *Behaviorism*; B.F. SKINNER, *Science and human behavior*.
- ⁶⁹ B.F. SKINNER, *Verbal behavior*.
- ⁷⁰ N. CHOMSKY, *Review of Skinner’s “Verbal Behavior”*.
- ⁷¹ G. GEORGE, *Behaviorism*, § 7.
- ⁷² M. BEANEY, *Analysis*.
- ⁷³ M. BEANEY, *Analysis*, § 6.8.
- ⁷⁴ R. RORTY (ed.), *Linguistic turn*.
- ⁷⁵ C.G. HEMPEL, *The logical analysis of psychology*, p. 20 - my italics.
- ⁷⁶ R. CARNAP, *Psychologie in physikalischer Sprache*.
- ⁷⁷ Obviously in the 1930s there was no possibility to obtain brain images.
- ⁷⁸ R. CARNAP, *Psychologie in physikalischer Sprache*, pp. 112-117 and pp. 136-132.
- ⁷⁹ Cf. *infra*, § 6 about physicalism.
- ⁸⁰ G. RYLE, *The concept of mind*.
- ⁸¹ *Ibid.*, chapter IV, § 8.
- ⁸² H. PUTNAM, *Mind, language, and reality*, chapter 16.
- ⁸³ The contrast made in the philosophy of mind between mentalism and behaviorism (the existence or non-existence of internal mental causes of behavior) is similar to the contrast between internalism and externalism made in the philosophy of language. Cf. S. DELLANTONIO, L. PASTORE, *Internal perception: The role of bodily information in concepts and word mastery*; S. DELLANTONIO, *La dimensione interna del significato. Esternismo, internismo e competenza semantica*.
- ⁸⁴ Cf., e.g., S. NANNINI, *The second cognitive turn*, pp. 1-2. More details in S. NANNINI, *Cause e ragioni*, pp. 124-128.
- ⁸⁵ J.J.C. SMART, *The mind/brain identity theory*.
- ⁸⁶ C. STOLJAR, *Physicalism*.
- ⁸⁷ J.J.C. SMART, *Materialism*.
- ⁸⁸ U.T. PLACE, *The concept of heed*.
- ⁸⁹ U.T. PLACE, *Is consciousness a brain process?*.
- ⁹⁰ U.T. PLACE, *Is consciousness a brain process?*, p. 45.
- ⁹¹ U.T. PLACE, *Thirty years on – Is consciousness still a brain process?*, p. 211; cf. also U.T. PLACE, *From mystical experience to biological consciousness. A pilgrim’s progress?*.
- ⁹² J.J.C. SMART, *Sensations and brain processes*; D.K. LEWIS, *An argument for the identity theory*.
- ⁹³ D.M. ARMSTRONG, *A materialist theory of mind*.
- ⁹⁴ Cf. on logical empiricists’ early version of physicalism O. NEURATH, *Physikalismus*; R. CARNAP, *Die physikalische Sprache als Universalsprache der Wissenschaft*.
- ⁹⁵ Cf. H. FEIGL, *Logical analysis of the psychophysical problem*.
- ⁹⁶ Cf. H. FEIGL, *The “Mental” and the “Physical”*; H. FEIGL, *Physicalism*.
- ⁹⁷ Cf. H. FEIGL, *Postscript after ten years*.
- ⁹⁸ About the distinction between “manifest image” and “scientific image” cf. W.S. SELLARS, *Empiricism and the philosophy of mind*.
- ⁹⁹ Cf. *supra*, § 3.
- ¹⁰⁰ Cf. W.V.O. QUINE, *Ontological relativity and other essays*; W.V.O. QUINE, *Word and object*.
- ¹⁰¹ Cf. W. BECHTEL, *Philosophy of science. An overview for cognitive sciences*, Erlbaum, Hillsdale (NJ); H. GARDNER, *The mind’s new science*.
- ¹⁰² The AI literature is endless. Cf., e.g., on the relationship between AI and philosophy M. BODEN (ed.), *The philosophy of artificial intelligence*.
- ¹⁰³ The literature on cognitive psychology is endless as well. Cf., e.g., on its birth and development U. NEISSER, *Cognitive psychology*; L.W. BARSALOU, *Cognitive psychology*.
- ¹⁰⁴ J. LEVIN, *Functionalism*.
- ¹⁰⁵ Cf., among the founders of functionalism, H. PUTNAM, *Minds and machines*; J.A. FODOR, *The language of thought*; Z.W. PYLYSHIN, *Meaning and cognitive structure*.
- ¹⁰⁶ Cf., e.g., S. PINKER, *How the mind works*. However Cf. also the sharp criticism of J.A. FODOR, *Mind doesn’t work that way: The scope and limits of computational psychology*.
- ¹⁰⁷ H. PUTNAM, *Minds and machines*.
- ¹⁰⁸ J. BICKLE, *Multiple realizability*.
- ¹⁰⁹ C. STOLJAR, *Physicalism*, § 2.2.1.
- ¹¹⁰ Cf., first of all with regard to this criticism of functionalism, H.L. DREYFUS; N. BLOCK, *Troubles with functionalism*; J.R. SEARLE, *Minds, brains and programs*.
- ¹¹¹ Cf. *infra*, § 8.
- ¹¹² Cf. *supra*, § 3.
- ¹¹³ Cf. first of all J.R. SEARLE, *The rediscovery of the mind*; J.R. SEARLE, *Mind. A brief introduction*.
- ¹¹⁴ Cf. D. DAVIDSON, *Essays on actions and events*; S. YALOWITZ, *Anomalous monism*.
- ¹¹⁵ Cf. T. O’CONNOR, *Emergent properties*; A. STEPHAN, *Emergenz: von der Unvorhersagbarkeit zur Selbstorganisation*.
- ¹¹⁶ Cf. A. BECKERMANN, H. FLOHR, J. KIM (eds.), *Emergence or reduction? Essays on the prospects of non-reductive physicalism*.
- ¹¹⁷ Cf. T. O’CONNOR, *Emergent properties*.
- ¹¹⁸ Cf., e.g., S. NANNINI, *L’anima e il corpo*, pp. XXI-XXIV.
- ¹¹⁹ Cf. *supra*, § 2.
- ¹²⁰ D.E. RUMELHART, J.L. MCCLELLAND (eds.), *Parallel distributed processing: exploration in the microstructure of cognition*.
- ¹²¹ C. BUCKNER, J.W. GARSON, *Connectionism*.
- ¹²² More in detail: <To understand the functioning of a network, one must know its elements and their interconnections. The purpose of this article is to discuss research strategies aimed at a comprehensive structural description of the network of elements and connections forming the

human brain. We propose to call this dataset the human “connectome”, and we argue that it is fundamentally important in cognitive neuroscience and neuropsychology. The connectome will significantly increase our understanding of how functional brain states emerge from their underlying structural substrate, and will provide new mechanistic insights into how brain function is affected if this structural substrate is disrupted» (O. SPORNS, G. TONONI, R. KÖTTER, *The human connectome*).

¹²³ Cf. first of all F. CRICK, *The astonishing hypothesis*; J.E. LEDOUX, *Emotional brain*; G.M. EDELMAN, *The remembered present*; G.M. EDELMAN, *Wider than the sky*; G.M. EDELMAN, G. TONONI, *A universe of consciousness*; A.R. DAMASIO, *Descartes' error: Emotion, reason, and the human brain*; A.R. DAMASIO, *The feeling of what happens*; A.R. DAMASIO, *Self comes to mind*; A. DAMASIO, *Feeling and knowing*; S. DEHAENE, *Consciousness and the brain*; B. LIBET, *Mind time*.

¹²⁴ Cf. *supra*, § 1.

¹²⁵ Cf. first of all R. Dawkins, *The selfish gene*; S.J. GOULD, *The mismeasure of man*.

¹²⁶ B. DAHLBOM (ed.), *Dennett and his critics: Demystifying mind*; D. ROSS, A. BROOK, D.L. THOMPSON (eds.), *Dennett's philosophy: A comprehensive assessment*.

¹²⁷ P.M. CHURCHLAND, *Eliminative materialism and the propositional attitudes*. Cf. also W. RAMSEY, *Eliminative materialism*.

¹²⁸ R.N. MCCAULEY (ed.), *The Churchlands and their critics*; W. HIRSTEIN, *On the Churchlands*.

¹²⁹ Cf. his doctoral dissertation: D.C. DENNETT, *Content and consciousness*.

¹³⁰ W.G. LYCAN, *Homuncular functionalism and other teleological theories*, p. 59.

¹³¹ Cf. first of all D.C. DENNETT, *Intentional systems*; D.C. DENNETT, *Brainstorms: Philosophical essays on mind and psychology*; D.C. DENNETT, *Intentional stance*.

¹³² D.C. DENNETT, *Consciousness explained*, chapter 8, § 2.

¹³³ *Ibid.*, chapter 13, § 3.

¹³⁴ Cf. first of all D.C. DENNETT, *Darwin's dangerous idea*; D.C. DENNETT, *Kinds of minds*; D.C. DENNETT, *Brainchildren*; D.C. DENNETT, *Freedom evolves*; D.C. DENNETT, *Breaking the spell: Religion as a natural phenomenon*; D.C. DENNETT, *From bacteria to Bach and back*.

¹³⁵ Cf. D.C. DENNETT, *Kinds of minds*, chapter 4.

¹³⁶ Cf., e.g., P.M. CHURCHLAND, *The engine of reason, the seat of the soul: A philosophical journey into the brain*, pp. 64-69.

¹³⁷ P.S. CHURCHLAND, V.S. RAMACHANDRAN, *Filling in: Why Dennett is wrong*.

¹³⁸ Cf. P.M. CHURCHLAND, *Eliminative materialism*; P.M. CHURCHLAND, *Matter and consciousness*, pp. 43-49.

¹³⁹ P. M. CHURCHLAND, *Matter and consciousness*, p. 44.

¹⁴⁰ P.M. CHURCHLAND, *The engine of reason*, pp. 21-54.

¹⁴¹ *Ibid.*, pp. 97-121.

¹⁴² *Ibid.*, pp. 137-150.

¹⁴³ *Ibid.*, cit., pp. 215.

¹⁴⁴ Cf. first of all R.N. MCCAULEY (ed.), *The Churchlands and their critics*.

¹⁴⁵ Cf. P.S. CHURCHLAND, *Brain-wise: Studies in neurophilosophy*. Cf. also P.S. CHURCHLAND, *Neurophilosophy*; P.S. CHURCHLAND, T.J. SEJNOWSKI, *The computational brain*; P.M. CHURCHLAND, P.S. CHURCHLAND, *On the contrary: Critical essays, 1987-1997*; P.S. CHURCHLAND, *Braintrust*; P.S. CHURCHLAND, *Touching a nerve*; P.S. CHURCHLAND, *Conscience*.

¹⁴⁶ P.M. CHURCHLAND, P.S. CHURCHLAND, *Do we pro-*

pose to eliminate consciousness?, p. 298.

¹⁴⁷ Cf. P.S. CHURCHLAND, *Consciousness: The transmutation of a concept*, pp. 80-95; P.S. CHURCHLAND, *Neurophilosophy*, p. 321.

¹⁴⁸ P.M. CHURCHLAND, P.S. CHURCHLAND, *Do we propose to eliminate consciousness?*, p. 298.

¹⁴⁹ Cf. S. NANNINI, *The mind-body problem in the philosophy of mind and cognitive neuroscience*.

¹⁵⁰ Cf. *supra*, § 3.

¹⁵¹ Cf. D.M. WEGNER, *The illusion of conscious will*.

¹⁵² Cf. S. NANNINI, *Naturalismo cognitivo*, pp. 135-158.

¹⁵³ Cf. E.L. SCHWARTZ, *Computational neuroscience*, MIT Press, Cambridge (MA).

¹⁵⁴ Cf. *supra*, § 1.

¹⁵⁵ Cf. *supra*, § 4.

¹⁵⁶ Cf. N. DANIELS, *Reflective equilibrium*.

Literature

ARMSTRONG, D.M. (1993). *A materialist theory of mind* (1968), Routledge, London/New York, revised edition.

BARRETT, J.A. (2006). *A quantum mechanical argument for mind-body dualism*. In: «Erkenntnis», vol. LXV, n. 1, pp. 97-115.

BARSALOU, L.W. (1992). *Cognitive psychology: An overview for cognitive scientists*, Erlbaum, Hillsdale (NJ).

BEANEY, M. (2021). *Analysis*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Summer edition – URL: <https://plato.stanford.edu/archives/sum2021/entries/analysis/>.

BECHTEL, W. (1988). *Philosophy of science. An overview for cognitive sciences*, Erlbaum, Hillsdale (NJ).

BECKERMANN, A., FLOHR, H., KIM, J. (eds.) (1992). *Emergence or reduction? Essays on the prospects of non-reductive physicalism*, De Gruyter, Berlin/New York.

BENINI, A. (2009). *Che cosa sono io? Il cervello alla ricerca di se stesso*, Garzanti, Milano.

BERKELEY, G. (2017). *A treatise concerning the principles of human knowledge* (1710). In: T. GOLDSCHMIDT, S. STAPLEFORD (eds.), *Berkeley's principles expanded and explained*, Routledge, London/New York.

BICKLE, J. (2020). *Multiple realizability*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Summer edition – URL: <https://plato.stanford.edu/archives/sum2020/entries/multiple-realizability/>.

BLOCK, N. (1978). *Troubles with functionalism*. In: C. WADE SAVAGE (ed.), *Perception and cognition. Issues in the foundations of psychology*, University of Minnesota Press, Minneapolis, pp. 261-325.

BODEN, M. (ed.) (1990), *The philosophy of artificial intelligence*, Oxford University Press, Oxford.

BUCKNER, C., GARSON, J.W. (2019). *Connectionism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Fall edition – URL: <https://plato.stanford.edu/archives/fall2019/entries/connectionism/>.

CARNAP, R. (1931-1932). *Die physikalische Sprache als Universalsprache der Wissenschaft*. In: «Erkenntnis», vol. II, pp. 433-465.

CARNAP, R. (1932-1933). *Psychologie in physikalischer Sprache*. In: «Erkenntnis», vol. III, pp. 107-142.

CHAKRAVARTTY, A. (2017). *Scientific realism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Summer edition – URL: <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>.

CHALMERS, D. (1996). *The conscious mind. In search of a*

- fundamental theory*, Oxford University Press, Oxford/New York.
- CHALMERS, D. (2010). *The character of consciousness*, Oxford University Press, Oxford/New York.
- CHOMSKY, N. (1959). *Review of Skinner's "Verbal Behavior"*. In: «Language», vol. XXXV, n. 1, pp. 26-58.
- CHURCHLAND, P.M., CHURCHLAND, P.S. (1998). *On the contrary: Critical essays 1987-1997*, MIT Press, Cambridge (MA).
- CHURCHLAND, P.M. (1988). *Matter and consciousness*, MIT Press, Cambridge (MA), revised edition.
- CHURCHLAND, P.M. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*, MIT Press, Cambridge (MA).
- CHURCHLAND, P.M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain*, MIT Press, Cambridge (MA).
- CHURCHLAND, P.M., CHURCHLAND, P.S. (1996). *Do we propose to eliminate consciousness?*. In: R.N. MC CAULEY (ed.), *The Churchlands and their critics*, Blackwell, Cambridge, pp. 297-300.
- CHURCHLAND, P.S. (1983). *Consciousness: The transmutation of a concept*. In: «Pacific Philosophical Quarterly», vol. LXIV, pp. 80-95.
- CHURCHLAND, P.S. (1986). *Neurophilosophy: Toward a unified science of the mind-brain*, MIT Press, Cambridge (MA).
- CHURCHLAND, P.S. (2002). *Brain-wise: Studies in neurophilosophy*, MIT Press, Cambridge (MA).
- CHURCHLAND, P.S. (2011). *Braintrust. What neuroscience tells us about morality*, Princeton University Press, Princeton (NJ).
- CHURCHLAND, P.S. (2013). *Touching a nerve: The self as brain*, Norton & Company, New York/London.
- CHURCHLAND, P.S. (2019). *Conscience: The origin of moral intuition*, Norton & Company, New York/London.
- CHURCHLAND, P.S., RAMACHANDRAN, V.S. (1993). *Filling in: Why Dennett is wrong*. In: B. DAHLBOM (ed.), *Dennett and his critics: Demystifying mind*, Blackwell, Oxford/Cambridge, pp. 177-203.
- CHURCHLAND, P.S., SEJNOWSKI, T.J. (1992). *The computational brain*, MIT Press, Cambridge (MA).
- CORBELLINI, G., SIRGIOVANNI, E. (2013). *Tutta colpa del cervello. Un'introduzione alla neuroetica*, Mondadori, Milano.
- CRICK, F. (1994). *The astonishing hypothesis: The scientific search for the soul*, Scribner, New York.
- DAHLBOM, D. (ed.) (1993). *Dennett and his critics: Demystifying mind*, Blackwell, Oxford/Cambridge (MA).
- DAMASIO, A.R. (1967). *Looking for Spinoza: Joy, sorrow and the feeling brain*, Harcourt, Orlando (FL).
- DAMASIO, A.R. (1999). *The feeling of what happens: Body, emotion and the making of consciousness*, Harcourt Brace, New York.
- DAMASIO, A.R. (2005). *Descartes' error: Emotion, reason, and the human brain* (1994), Penguin Books, London.
- DAMASIO, A.R. (2010). *Self comes to mind: Constructing the conscious brain*, Pantheon Books, New York.
- DAMASIO, A.R. (2021). *Feeling and knowing: Making minds conscious*, Pantheon Books, New York.
- DANIELS, N. (2020). *Reflective equilibrium*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Summer edition - URL: <https://plato.stanford.edu/archives/sum2020/entries/reflective-equilibrium/>.
- DARWIN, C. (1981) *The descent of man, and selection in relation to sex* (1871), Princeton University Press, Princeton (NJ).
- DAVIDSON, D. (2001). *Essays on actions and events* (1980), Clarendon Press, Oxford University Press, Oxford/New York.
- DAWKINS, R. (2016). *The selfish gene* (1976). In: R. DAWKINS, *The selfish gene: 40th anniversary edition*, Oxford University Press, Oxford/New York.
- DE CARO, M., MACARTHUR, D. (2005). *Naturalismo scientifico e naturalismo liberalizzato. Introduzione all'edizione italiana*. In: M. DE CARO, D. MACARTHUR (a cura di), *La mente e la natura. Per un naturalismo liberalizzato*, traduzione di L. GRECO, G. PELLEGRINO, Fazi, Roma, pp. xv-xxi.
- DE CARO, M., MACARTHUR, D. (eds.) (2004). *Naturalism in question*, Harvard University Press, Cambridge (MA).
- DEHAENE, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*, Viking, New York.
- DELLANTONIO, S. (2007). *La dimensione interna del significato. Esternismo, internismo e competenza semantica*, ETS, Pisa.
- DELLANTONIO, S., PASTORE, L. (2017). *Internal perception: The role of bodily information in concepts and word mastery*, Springer, Berlin/Heidelberg.
- DENNETT, D.C. (1971). *Intentional systems*. In: «The Journal of Philosophy», vol. LXVIII, n. 4, pp. 87-106.
- DENNETT, D.C. (1987). *Intentional stance*, MIT Press, Cambridge (MA).
- DENNETT, D.C. (1991). *Consciousness explained*, Little, Brown and Co., Boston.
- DENNETT, D.C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life*, Simon & Schuster, New York.
- DENNETT, D.C. (1996). *Kinds of minds: Toward an understanding of consciousness*, Basic Books, New York.
- DENNETT, D.C. (1998). *Brainchildren: Essays on designing minds*, MIT Press, Cambridge (MA).
- DENNETT, D.C. (2003). *Freedom evolves*, Viking, New York.
- DENNETT, D.C. (2006). *Breaking the spell: Religion as a natural phenomenon*, Viking, New York.
- DENNETT, D.C. (2010). *Content and consciousness* (1969), Routledge, London/New York.
- DENNETT, D.C. (2017). *Brainstorms: Philosophical essays on mind and psychology* (1978), MIT Press, Cambridge (MA).
- DENNETT, D.C. (2017). *From bacteria to Bach and back: The evolution of minds*, Norton & Company, New York/London.
- DESCARTES, R. (2008). *Meditations on first philosophy* (1641), with Selections from the *Objections and Replies*, translated by M. MORIARTY, Oxford University Press, Oxford.
- DILTHEY, W. (1988). *Introduction to the human sciences: An attempt to lay a foundation for the study of society and history* (1881), translated by R.J. BETANZOS, Wayne State University Press, Detroit.
- DREYFUS, H. (1972). *What computers can't do. A critique of artificial reason*, Harper and Row, New York.
- EDELMAN, G. (2004). *Wider than the sky: The phenomenal gift of consciousness*, Yale University Press, New Haven.
- EDELMAN, G.M. (1989). *The remembered present: A biological theory of consciousness*, Basic Books, New York.
- EDELMAN, G.M. (2004). *Wider than the sky: The phenomenal gift of consciousness*, Yale University Press,

- New Haven.
- EDELMAN, G.M., TONONI, G. (2000). *A universe of consciousness: How matter becomes imagination*, Basic Books, New York.
- EINSTEIN, A. (1920). *Relativity: The special and general theory*, translated by R.W. LAWSON, Henry Holt and Company, New York.
- FEIGL, H. (1967). *Postscript after ten years*. In: H. FEIGL, *The "Mental" and the "Physical": The essay with a postscript*, University of Minnesota Press, Minneapolis, pp. 135-160.
- FEIGL, H. (1934). *Logical analysis of the psychophysical problem: A contribution of the new positivism*. In: «Philosophy of Science», vol. I, pp. 420-445.
- FEIGL, H. (1963). *Physicalism: Unity of science and the foundations of psychology*. In: P.A. SCHILPP (ed.), *The philosophy of Rudolf Carnap*, Open Court, La Salle (ILL), pp. 227-267.
- FEIGL, H. (1958). *The "Mental" and the "Physical"*. In: H. FEIGL, M. SCRIVEN (eds.), *Concepts, theories, and the mind-body problem*, University of Minnesota Press, Minneapolis, pp. 370-497.
- FODOR, J.A. (2000). *Mind doesn't work that way: The scope and limits of computational psychology*, MIT Press, Cambridge (MA).
- FODOR, J.A. (1975). *The language of thought*, Crowell, New York.
- GALLAGHER, S., ZAHAVI, D. (2021). *The phenomenological mind*, Routledge, Taylor and Francis, London/New York, third edition.
- GARDNER, H. (1985). *The mind's new science: A history of the cognitive revolution*, Basic Books, New York.
- GEORGE, G. (2019). *Behaviorism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Spring edition – URL: <https://plato.stanford.edu/archives/spr2019/entries/behaviorism/>.
- GOULD, S.J. (1996). *The mismeasure of man*, W.W. Norton, New York.
- GUYER, P., HORSTMANN, R.P. (2022). *Idealism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Spring edition – URL: <https://plato.stanford.edu/archives/spr2022/entries/idealism/>.
- HAMILTON, P.D. (2003). *Historicism*, Routledge, London/New York.
- HEMPEL, C.G. (1980). *The logical analysis of psychology (1935)*. In: N. BLOCK (ed.), *Readings in the philosophy of psychology*, vol. I, Harvard University Press, Cambridge (MA), pp. 14-23.
- HIRSTEIN, W. (2004). *On the Churchlands*, Thomson Wadsworth, Belmont (CA).
- HUSSERL, E. (2001). *Logical investigations*, 2 voll. (1900-1901), translated by J.N. FINDLAY, Routledge, London/New York.
- HUSSERL, E. (1982). *Ideas pertaining to a pure phenomenology and a phenomenological philosophy. First book: General introduction to a pure phenomenology*. In: E. HUSSERL, *Collected work*, vol. II, translated by F. KERSTEN, Martinus Nijhoff, The Hague.
- HUSSERL, E. (1989). *Ideas pertaining to a pure phenomenology and a phenomenological philosophy. Second book: Studies in phenomenology of constitution*. In: E. HUSSERL, *Collected work*, vol. III, translated by R. ROJCEWICZ, A. SCHUWER, Kluwer, Dordrecht.
- HUSSERL, E. (1980). *Ideas pertaining to a pure phenomenology and a phenomenological philosophy. Third book: Phenomenology and the foundation of sciences*. In: E. HUSSERL, *Collected work*, vol. I, translated by T. KLEIN, W. POHL, Martinus Nijhoff, The Hague.
- JIBU, M., YASUE, K. (1995). *Quantum brain dynamics and consciousness: An introduction*, John Benjamins, Amsterdam/Philadelphia.
- KANT, I. (1998). *Critique of pure reason (1787)*, translated by P. GUYER, A.W. WOOD, Cambridge University Press, Cambridge.
- KÜNG, H. (2005). *Der Anfang aller Dinge: Naturwissenschaft und Religion*, Piper, München.
- LAVAZZA, A. (a cura di) (2008). *L'uomo a due dimensioni. Il dualismo mente-corpo oggi*, Bruno Mondadori, Milano.
- LEDoux, J.E. (1996). *Emotional brain: The mysterious underpinnings of emotional life*, Simon & Schuster, New York.
- LEVIN, J. (2021). *Functionalism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Winter edition – URL: <https://plato.stanford.edu/archives/win2021/entries/functionalism/>.
- LEWIS, D.K. (1966). *An argument for the identity theory*. In: «Journal of Philosophy», LXIII, n. 1, pp. 17-25.
- LIBET, B. (2004). *Mind time: The temporal factor in consciousness*, Harvard University Press, Cambridge (MA).
- LOWE, E.J. (2000). *An introduction to the philosophy of mind*, Cambridge University Press, Cambridge.
- LYCAN, W.G. (1990). *Homuncular functionalism and other teleological theories*. In: W.G. LYCAN (ed.), *Mind and cognition: A reader*, Blackwell, Oxford, pp. 59-62.
- MANCIA, M. (ed.) (2006). *Psychoanalysis and neuroscience*, Springer, Berlin.
- MCCAULEY, R.N. (ed.) (1996). *The Churchlands and their critics*, Blackwell, Cambridge (MA).
- MCDOWELL, J. (1996). *Mind and world (1994)*, Harvard University Press, Cambridge (MA).
- MCDOWELL, J. (2004). *Naturalism in the philosophy of mind*. In: M. DE CARO, D. MACARTHUR (eds.), *Naturalism in question*, Harvard University Press, Cambridge (MA), pp. 91-105.
- MCGINN, C. (1982). *The character of mind*, Oxford University Press, Oxford.
- MCGINN, C. (1990). *The problem of consciousness*, Blackwell, Oxford.
- MCGINN, C. (2000). *The mysterious flame. Conscious minds in a material world*, Basic Books, New York.
- MILLER, A. (2021). *Realism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Winter edition – URL: <https://plato.stanford.edu/archives/win2021/entries/realism/>.
- NAGEL, T. (1974). *What is it like to be a bat?*. In: «The Philosophical Review», vol. LXXXIII, n. 4, pp. 435-450.
- NAGEL, T. (1987). *What does it all mean? A very short introduction to philosophy*, Oxford University Press, Oxford/New York.
- NANNINI, S. (1992). *Cause e ragioni. Modelli di spiegazione delle azioni umane nella filosofia analitica*, Editori Riuniti, Roma, pp. 124-128.
- NANNINI, S. (2007). *Naturalismo cognitivo. Per una teoria materialistica della mente*, Quodlibet, Roma/Macerata.
- NANNINI, S. (2018). *The mind-body problem in the philosophy of mind and cognitive neuroscience: A physicalist naturalist solution*. In: «Neurological Sciences», vol. XXXIX, n. 9, pp. 1509-1517.

- NANNINI, S. (2019). *Alla riscoperta di Spinoza: in compagnia di Antonio Damasio e Emanuela Scribano*. In: M. FAVARETTI CAMPOSAMPIERO, M. PRIAROLO (a cura di), *La logica delle idee: studi di filosofia moderna in onore di Emanuela Scribano*, Mimesis, Milano/Udine, pp. 205-216.
- NANNINI, S. (2020). *Mente e cervello. La seconda svolta cognitiva in filosofia della mente e altri saggi*, Corisco, Roma/Messina.
- NANNINI, S. (2021). *L'anima e il corpo. Un'introduzione storica alla filosofia della mente*, Laterza, Roma/Bari, II edition.
- NANNINI, S. (2021). *The second cognitive turn in the philosophy of mind*, preprint, August 2021 – doi: 10.13140/RG.2.2.35071.94881.
- NANNINI, S., SANDKÜHLER, H.-J. (eds.) (2000). *Cognitive naturalism in the philosophy of mind*, Peter Lang, Frankfurt a.M.
- NEISSER, U. (1967). *Cognitive psychology*, Appleton-Century-Crofts, New York.
- NEURATH, O. (1957). *Physikalismus* (1931). In: «Scienza», vol. L, pp. 297-303.
- O'CONNOR, T. (2021). *Emergent properties*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Winter edition – URL: <https://plato.stanford.edu/archives/win2021/entries/properties-emergent/>.
- PAPINEAU, D. (2021). *Naturalism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Summer edition – URL: <https://plato.stanford.edu/archives/sum2021/entries/naturalism/>.
- PETITOT, J., VARELA, F.J., PACHOUD, B., ROY, J.-M. (eds.) (1999). *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*, Stanford University Press, Stanford.
- PINKER, S. (1997). *How the mind works*, Norton, New York.
- PLACE, U.T. (1954). *The concept of heed*. In: «The British Journal of Psychology», vol. XLV, n. 4, pp. 243-255.
- PLACE, U.T. (1956). *Is consciousness a brain process?*. In: «The British Journal of Psychology», vol. XLVII, n. 1, pp. 44-50.
- PLACE, U.T. (1988). *Thirty years on: Is consciousness still a brain process?*. In: «Australasian Journal of Philosophy», vol. LVI, n. 2, pp. 208-219.
- PLACE, U.T. (1998). *From mystical experience to biological consciousness. A pilgrim's progress?*. In: M. CHEUNG CHUNG (ed.), *Current trends in history and philosophy of psychology*, vol. I, British Psychological Society, Leicester, pp. 43-48.
- POPPER, K.R. (1994). *Knowledge and the body-mind problem*, Routledge, London/New York.
- POPPER, K.R., ECCLES, J.C. (1977). *The self and its brain*, Springer, New York.
- PUTNAM, H. (1960). *Minds and machines*. In: S. HOOK (ed.), *Dimensions of mind*, New York University Press, New York, pp. 148-180.
- PUTNAM, H. (1975). *Mind, language, and reality*, Cambridge University Press, Cambridge/New York.
- PUTNAM, H. (1981). *Reason, truth and history*, Cambridge University Press, Cambridge.
- PUTNAM, H. (2004). *The content and appeal of "naturalism"*. In: M. DE CARO, D. MACARTHUR (eds.), *Naturalism in question*, Harvard University Press, Cambridge (MA), pp. 59-70.
- PYLYSHIN, Z.W. (1986). *Meaning and cognitive structure: issues in the computational theory of mind*, Ablex Publishing, New York.
- QUINE, W.v.O. (1960). *Word and object*, Wiley, New York/London.
- QUINE, W.v.O. (1969). *Ontological relativity and other essays*, Columbia University Press, New York.
- QUINE, W.v.O. (1981). *Theories and things*, Harvard University Press, Cambridge (MA).
- RACINE, E. (2010). *Pragmatic neuroethics*, MIT Press, Cambridge (MA).
- RAMSEY, W. (2022). *Eliminative materialism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Spring edition – URL: <https://plato.stanford.edu/archives/spr2022/entries/materialism-eliminative/>.
- ROBINSON, H. (2020). *Dualism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Fall edition – URL: <https://plato.stanford.edu/archives/fall2020/entries/dualism/>.
- ROBINSON, W. (2019). *Epiphenomenalism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Summer edition – URL: <https://plato.stanford.edu/archives/sum2019/entries/epiphenomenalism/>.
- ROBSON, J., CURRIE, G. (2022). *Aesthetics and cognitive science*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Summer edition – URL: <https://plato.stanford.edu/archives/sum2022/entries/aesthetics-cogsci/>.
- RORTY, R. (ed.) (1992). *Linguistic turn: Essays in philosophical method* (1967), University of Chicago Press, Chicago.
- ROSKIE, A. (2021). *Neuroethics*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Spring edition – URL: <https://plato.stanford.edu/archives/spr2021/entries/neuroethics/>.
- ROSS, D., BROOK, A., THOMPSON, D.L. (eds.) (2000). *Dennett's philosophy: A comprehensive assessment*, MIT Press, Cambridge (MA).
- ROSSI, P. (1971). *Lo storicismo tedesco contemporaneo* (1956), Einaudi, Torino.
- ROTH, G., STRÜBER, N. (2014). *Wie das Gehirn die Seele macht*, Klett-Cotta, Stuttgart.
- ROWLANDS, M. (2009). *The mind-body problem*. In: W.P. BANKS (ed.), *Encyclopedia of consciousness*, Academic Press, Cambridge (MA), pp. 43-55.
- RUMELHART, D.E., MCCLELLAND, J.L. (eds.) (1986). *Parallel distributed processing: Exploration in the microstructure of cognition*, MIT Press, Cambridge (MA).
- RUSSELL, B. (1995). *The analysis of mind* (1921), Routledge, London.
- RYLE, G. (2002). *The concept of mind* (1949), University of Chicago Press, Chicago.
- SCHWARTZ, E.L. (1990). *Computational neuroscience*, MIT Press, Cambridge (MA).
- SEARLE, J.R. (1980). *Minds, brains and programs*. In: «Behavioral and Brain Sciences», vol. III, n. 3, pp. 417-457.
- SEARLE, J.R. (1992). *The rediscovery of the mind*, MIT Press, Cambridge (MA).
- SEARLE, J.R. (2004). *Mind. A brief introduction*, Oxford University Press, Oxford.
- SELLARS, R.W. (1921). *Epistemological dualism vs. metaphysical dualism*. In: «The Philosophical Review», vol. XXX, n. 5, pp. 482-493.
- SELLARS, W.S. (1956). *Empiricism and the philosophy of mind*, Harvard University Press, Cambridge (MA).
- SETH, A. (2021). *Being you. A new science of consciousness*, Faber, London.

- SKINNER, B.F. (2014). *Science and human behavior* (1953), The B.F. Skinner Foundation, Cambridge (MA).
- SKINNER, B.F. (2014). *Verbal behavior* (1957), The B.F. Skinner Foundation, Cambridge (MA).
- SMART, J.J.C. (1959). *Sensations and brain processes*. In: «Philosophical Review», vol. LXVIII, n. 2, pp. 141-156.
- SMART, J.J.C. (2017). *The mind/brain identity theory*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Spring edition – URL: <https://plato.stanford.edu/archives/spr2017/entries/mind-identity/>.
- SMART, J.J.C. (2023). *Materialism*. In: «Encyclopaedia Britannica», 2023, April 20th – URL: <https://www.britannica.com/topic/materialism-philosophy>.
- SMITH, B. (1988). *Gestalt theory. An essay in philosophy*. In: B. SMITH (ed.), *Foundations of gestalt theory*, Philosophia Verlag, Munich/Vienna, pp. 11-81.
- SOBEL, C.P. (2001). *Cognitive sciences: An interdisciplinary approach*, Mayfield Publication Company, Mountain View (CA).
- SPORNS, O., TONONI, G., KÖTTER, R. (2005). *The human connectome: A structural description of the human brain*. In: «PLOS – Computational Biology», vol. I, n. 4, Art. Nr. e42 – doi: 10.1371/journal.pcbi.0010042.
- SPURRETT, D., PAPINEAU, D. (1999). *A note on the completeness of "physics"*. In: «Analysis», vol. LIX, n. 1, pp. 25-29.
- STAPP, H.P. (ed.) (2009). *Mind, matter and quantum mechanics*, Springer, Berlin/Heidelberg.
- STEPHAN, A. (1999). *Emergenz: von der Unvorhersagbarkeit zur Selbstorganisation*, Dresden University Press, Dresden/München.
- STEPHAN, A., WALTER, S. (Hrsg.) (2013). *Handbuch Kognitionswissenschaft*, Metzler, Stuttgart/Weimar.
- STOLJAR, C. (2022). *Physicalism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Summer edition - URL: <https://plato.stanford.edu/archives/sum2022/entries/physicalism/>.
- STUBENBERG, L. (2018). *Neutral monism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Fall edition – URL: <https://plato.stanford.edu/archives/fall2018/entries/neutral-monism/>.
- SWINBURNE, R. (2007). *From mental/physical identity to substance dualism*. In: P.W. INWAGEN, D.W. ZIMMERMAN (eds.), *Persons: Human and divine*, Oxford University Press, Oxford, pp. 147-179.
- VON NEUMANN, J. (1966). *Mathematical foundations of quantum mechanics* (1932), Princeton University Press, Princeton (NJ).
- WATSON, J.B. (1998). *Behaviorism* (1924), Transaction Publishers, New Brunswick (NJ).
- WEGNER, D.M. (2002). *The illusion of conscious will*, MIT Press, Cambridge (MA).
- WIGNER, E.P. (1961). *Remarks on the mind-body question*. In: I.J. GOOD (ed.), *The scientist speculates*, Heinemann, London, pp. 284-302.
- YALOWITZ, S. (2021). *Anomalous Monism*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*, Spring edition – URL: <https://plato.stanford.edu/archives/spr2021/entries/anomalous-monism/>.
- YU, S., NIKOLIĆ, D. (2011). *Quantum mechanics needs no consciousness*. In: «Annalen der Physik», vol. DXXIII, pp. 931-938.

RICERCHE

“Cognition” – Let’s forget it?

Alan Costall^(a)

Ricevuto: 3 novembre 2022; accettato: 27 febbraio 2023

Abstract For many psychologists, “cognition” is an obvious object for study. A natural kind. What I want to do in this article is problematise “cognition”. Psychologists lived happily without “cognition” until the 1960’s and even then, its entry into psychological discourse was hardly smooth. Furthermore, the new cognitive psychology retained much of the behaviourism it wrongly claimed to have displaced. There are now some radical developments going on in “cognitive science” but those involved still retain the term “cognition”. But isn’t it like modern physicists claiming that they are coming up with new theories of phlogiston? “Cognition” – forget it?

KEYWORDS: Psychology; Cognition; Behaviourism; Cognitive Behaviourism; S-R Theory; Unconscious Mind

Riassunto “Cognizione”: dobbiamo lasciarla perdere? – Per molti psicologi la “cognizione” è un oggetto di studio che rasenta l’ovvietà. Un genere naturale. Ciò che mi propongo di fare in questo articolo è problematizzare la “cognizione”. Gli psicologi hanno vissuto felicemente senza la “cognizione” fino agli Anni ’60 e anche allora la comparsa di questa nozione all’interno del lessico psicologico non è stata cosa semplice. Inoltre, la nuova psicologia cognitiva ha conservato molto di quel comportamentismo che ha affermato, sbagliando, di aver scalzato. Ci sono oggi alcuni sviluppi, anche radicali, che si affacciano nella “scienza cognitiva”, ma tutti quelli che sono coinvolti usano ancora il termine “cognizione”. Ma non è come se i fisici di oggi sostenessero di avere nuove teorie del flogisto? La “cognizione”: dobbiamo lasciarla perdere?

PAROLE CHIAVE: Cognizione; Comportamentismo; Comportamentismo cognitivo; Teoria stimolo-risposta; Mente inconscia

^(a)Department of Psychology, University of Portsmouth, King Henry Street, PO1 2DY, Portsmouth (UK)

E-mail: alan.costall@port.ac.uk (✉)



“Cognitive” belongs to the vocabulary of
examination papers¹
Gilbert Ryle

[There are] no good grounds for supposing
this additional wheel work in the mind²
William James

Cognition, see *Knowing*³
William James

“Cognitive” is a classical term that implies a natural
cleavage between psychological processes, a cleavage that
confuses everything and clarifies nothing;
let’s forget it⁴
George Kelly

I BEGAN UNIVERSITY IN 1966 at the University College of North Wales (now Bangor University) to study physics, but the course was heavy on detail and light on big ideas, in contrast to the physics I had been taught at school. In 1968, I escaped to the Psychology Department. At the time, it was tiny compared to the Physics Department: four lecturers and just twelve students in my year. The main focus was on operant psychology and the philosophy of radical behaviourism, aided and abetted by Wittgenstein and Gilbert Ryle. (Both Skinner and Ryle turned up to give seminars – I guess Wittgenstein might have also showed up if he hadn’t been dead). However, a new lecturer had arrived from the University of London, and the cognitive revolution finally reached the outer reaches of North Wales. In his course, the required reading was the newly published *Cognitive Psychology*, by Ulric Neisser (1967). I was yet again confronted with more interminable detail – this time based around “cognitive boxology”.

I have become increasingly puzzled how Neisser could have written such a boring book. He had already written a penetrating critique of the computer analogy⁵ and later became an “apostate” of the cognitive revolution.⁶ When I gave a talk at Oxford in 1988, he made a point of coming up to

me before my talk to say he liked the edited book I had put together with Arthur Still against cognitivism.⁷ He, like some of the audience, must have already known that a ridiculously nasty review of the book had appeared in the *Quarterly Journal of Experimental Psychology*.⁸ I did not know about the review until I arrived to give my talk.⁹

As I have already explained, even before his 1987 book, Neisser had already written incisive criticism about the direction the new cognitive psychology was taking both in terms of theory and research practice and was turning away from boxology. As Neisser has explained in *Cognition and reality*, he deliberately avoided any discussion of consciousness precisely because he could see that it would not be sufficient merely to treat consciousness as yet another «stage of processing in a mechanical flow of information».¹⁰ He even included a spoof of his early boxology (cf. *Figure 1*).

■ 1 Cognition” prior to “the cognitive revolution”

The textbooks have been keen to enlist “pioneers” in the history of the cognitive approach, such as Jean Piaget, Frederic Bartlett, and Edward Tolman. On the whole, I think this was a case of retrospective reconstruction rather than a case of historical lineage. One important exception from the outset was Neisser who identified closely with Bartlett.¹¹ What I want to argue in this section, is that prior to the 1960’s, the use of “cognition” within the language of psychology was both sporadic and *selective*. Ernest Hilgard pointed out that “cognition” entered into psychological language initially as part of a *package*:

For two hundred years many psychologists took for granted that the study of mind could be divided into three parts: cognition, affection, and conation. They disagreed on whether these should be considered faculties of the mind or merely a classification of aspects of mental activity, but the threefold division was

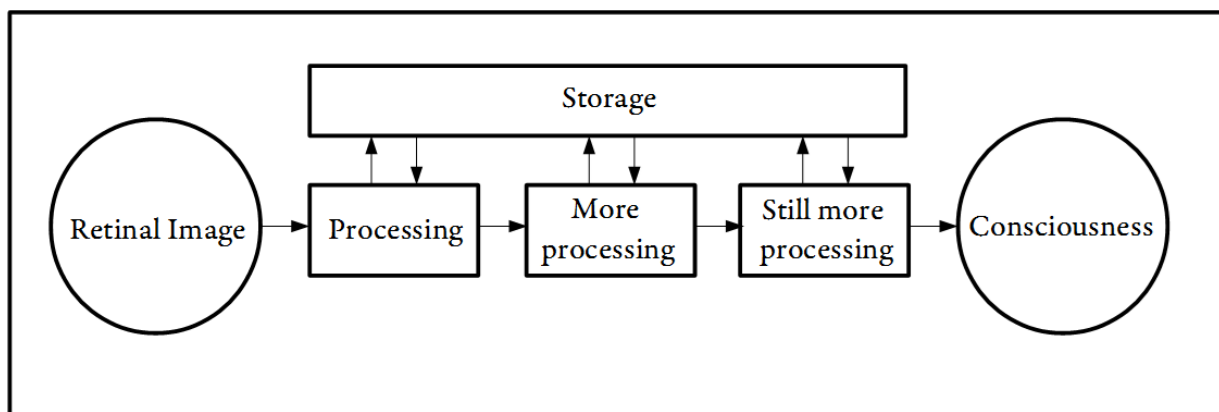


Figure 1. The internal information-processing model of perception

repeatedly revived. In the last twenty-five years, if we judge from the titles of books and journal articles, scientific psychology – whether its focus is on perception, learning and memory, development, or personality and social psychology – has become engaged with one of these aspects, now called cognitive psychology. An examination of the tripartite classification in historical perspective may show the extent to which affection and conation are now suffering neglect by contrast with cognition as their coequal. [...] the old trilogy helps to call attention to aspects that are neglected; it remains useful after all these years.¹²

As far as I can tell, at least in relation to the Anglo-American literature, “cognition” broke loose from this trilogy.

According to Gaines and Shaw (n.d.), «[Sir William] Hamilton (1859) introduced the term “cognition” into psychology».¹³ They also cite two early books with “cognition” or “cognitive” in their titles: *Elements of the psychology of cognition* (by R. Jardine) and *Psychology: The cognitive powers* (by J. McCosh).¹⁴ Jardine’s *Elements of the psychology of cognition* commences with the definition:

Cognition is a general name which we may apply to all those mental states in which there is made known in consciousness either some affection or activity of the mind itself, or some external quality or object. The Psychology of Cognition analyses knowledge into its primary elements, and seeks to ascertain the nature and laws of the processes through which all our knowledge passes in progressing from its simplest to its most elaborate condition.¹⁵

But then there was a “long” lull. There are two books that, from their titles, seem to be precursors to the “revolution”. One is Charles Spearman’s *Nature of intelligence and principles of cognition*.¹⁶ But this was misplaced in terms of an historical progression. It was about individual differences and based on psychometric testing rather than experimental psychology. The first English-language twentieth century textbook with the title, *Cognitive psychology*, however, was written by Dom Thomas Verner Moore,¹⁷ a Benedictine monk, but the approach adopted links back to the tradition of scholastic philosophy rather than, again, anticipating the forthcoming “cognitive revolution”.¹⁸

For me, a big surprise was to find Robert Leeper’s substantial chapter on “cognitive processes” in Stevens’ prestigious *Handbook of experimental psychology* published as early as 1951.¹⁹ Leeper begins his chapter by raising the issue of whether cognition should be defined in terms of particular psychological processes, or in relation to what he called an “approach”. In terms of processes, he

considered whether the scope of the definition should be restricted to thinking or even specific kinds of thinking, or else be more inclusive to cover perception and remembering. Leeper wondered whether “consciousness” should figure in the definition of cognition, but, drawing upon the early introspective research on imageless and sensationless thought, he concluded that cognition could indeed be unconscious. The definition Leeper finally settled upon sounds surprisingly modern:

[...] cognitive processes include all the means whereby the individual represents anything to himself or uses these representations as a means of guiding his behavior. It is in this broader sense that the term “cognitive processes”, after virtually disappearing from the vocabulary of psychology, has been reappearing in the writings of [some] psychologists.²⁰

Leeper’s definition of cognition in terms of representation embodies the source of our later perplexities. For his definition is ambiguous. It can be taken as either defining a field of inquiry or else setting out a theoretical, indeed metatheoretical, *approach*. Regarded as a field of inquiry, the study of cognition would be reasonably restricted to those limited but important areas of human life where people manifestly engage in representation of various kinds as a general way of guiding their activities. However, as an “approach” there is, in principle, no limit to the application of the term “cognition” well beyond the restricted realm of thinking, planning, classifying, and so on.

Already by the 1980s, psychology had almost entirely redefined itself as the science of “cognition”. As William Kessen later complained:

Friendship has become social cognition, affect is seen as a form of problem-solving, new-born perception is subsumed under a set of transforming rules, and psychoanalysis is reread as a variant of information processing. Cognition, the feeble infant of the late Fifties and early Sixties, has become an apparently insatiable giant.²¹

2 The “cognitive revolution”

In preparation for this paper, I checked on Google’s Ngram, and was struck by how abruptly references to “Cognitive Psychology” and, just a bit later, “Cognitive Science” took off from the 1960’s onwards.

I was also amazed by the terminal decline in references to both terms by 2012. But it turns out this statistical profile a fault in Ngram rather than evidence of the collapse of the “cognitive.” (Enter any other term, e.g., “chair,” and you get the same terminal profile!).

In this section, I want to make two points. The



Figure 2. References to “Cognitive Psychology” and “Cognitive Science” from 1960s onwards

reception of the term “cognitive” was not straightforward, and the “revolution” has hardly been complete.

According to Moroz,²² the first mention of “cognition” in the introductory textbooks of psychology was as recent as 1966, in McKeachie and Doyle’s *Psychology*.²³ Moroz admitted that he had no clear idea what the term was supposed to mean, and actually referred back to the Leeper chapter (referred to above) for help. He was not alone:

Querying colleagues about its meaning [cognition], I received a response direct but confusing (as that a child might receive when asking where he came from) or operationally precise but circular [...].²⁴

In his presidential address to the American Psychological Association, the influential psychologist, Donald Hebb, was also puzzled, but, in the end, decided there was an intimate connection between cognitive theorizing and stimulus-response *behaviorism*:

[...] the whole meaning of the term “cognitive” depends on [the stimulus-response idea], though cognitive psychologists seem unaware of the fact. The term is not a good one, but it does have meaning as a reference to features of behavior that do not fit the S-R formula; and no other meaning at all as far as one can discover.²⁵

In fact, even after many years, the textbooks continue to define “cognition” as what goes on between the stimulus and response. (More of this later.)

3 Traditional cognitivism

Traditional cognitivism usually comes as a package. The obvious component has been the commitment to *representational theory*: we do not experience the world but an internal mental surrogate of it. But there is also a commitment to un-

worldliness. Despite important caveats, Neisser claimed that «psychology, like economics, is a science concerned with the interdependence between certain events rather than with their physical nature».²⁶ However, the most extreme statement of this commitment I have encountered goes as follows: «a science of structure and function *divorced* from material substance».²⁷

Then there is *intellectualism*. To repeat my initial quote from Ryle, a commitment to «the language of the examination room». George Mandler gives an amusing comment on this in relation to a session he attended with Noam Chomsky:

[...] being the ultimate theorist of pure mentalism, [Chomsky] finally broke out to discuss the behavior of a baseball outfielder catching a flying ball. For Chomsky, this was a complex computational problem because the task obviously involved the “solution of several simultaneous equations”. Tell that to Joe DiMaggio!²⁸

There has also been the fundamental commitment to “*the unconscious mind*” stemming from Helmholtz’s notion of “unconscious inference”.²⁹ Once you take this route you can get away with theoretical murder.³⁰ For example, if Di Maggio didn’t realize that he was both as an athletic and mathematical genius, it was simply because he was unaware of it in the latter case. His cognitive unconscious was doing all the work.

4 The “cognitive revolution” as the overthrow of behaviourism?

As the social psychologist Solomon Asch suggested:

current cognitive psychology, despite the striking change of language it has introduced, [is] perhaps too often a guise for a newly attired behaviorism, a species of [...] cognitive behaviorism.³¹

As I see it, there are (at least) three “hang-overs”

within traditional cognitivism from behaviourism.

First, there is the continuing commitment to Stimulus-Response, or Input-Output, psychology. “Cognition” is supposed to be what goes on between. The mind is active, the body is passive and receptive. Ironically, a rare psychologist, James Gibson, who eventually rejected this scheme is (misrepresented) as its purist exponent.³²

Second, there is the retention of the hypothetical-deductive method:

The activity that dominates cognitive psychology today is not empirical exploration but something quite different: namely, the making and testing of hypothetical models. Ironically, the “hypothetico-deductive method” that was so strongly advocated by Hullian behaviorists half a century ago has become the stock-in-trade of their cognitivist successors. They argue that research should always begin with a theory; not just any theory, but a specific model of the internal processes that underlie the behavior of interest. That mental model is then tested as thoroughly as possible in carefully designed experimental paradigms. When it has been proven false (as it invariably is), a revised model is constructed so that the cycle can begin anew. The aim of the research is not to discover any secret of nature; it is to devise models that fit a certain range of laboratory data better than their competitors do.³³

In the same year, Neisser gave a specific example:

Images, models, and human nature. Why does the theory suggested here strike the reader as clever rather than insightful, as cute model making rather than serious psychology? I think it is because the thinking of Kosslyn and his collaborators is completely detached from everything we know about human nature or about perception, thinking, and the nervous system. Like much contemporary work in “information processing”, it attempts to “account for” a sharply restricted body of experimental results (usually reaction latencies) by relating it to an equally restricted class of models (usually computer programs or something similar).³⁴

The most insidious carry-over is “methodological behaviourism” – that all we can experience of another person or other animal directly is just, to use Clark Hull’s term, mere colourless movements. In short, there is a dualism of mind and behaviour. Here are three examples:

Your private experience is a theoretical construct to me. I have no direct access to your private experience. I do have direct access to your behavior. In that sense I’m a behaviorist.

In that sense, *everybody* is a behaviorist today.³⁵

Because psychologists were growing impatient with introspection, the new behaviorism caught on rapidly [...]. The modern cognitive perspective is in part a return to the cognitive roots of psychology and in part a reaction to the narrowness of behaviorism and the S-R view [...]. Like the 19th century version, the modern study of cognition is concerned with mental processes such as perceiving, remembering, reasoning, deciding, and problem solving. Unlike the 19th-century version, however, modern cognitivism is not based on introspection. Instead, it assumes (1) only by studying mental processes can we fully understand what organisms do, and (2) we can study mental processes in an objective fashion by focusing on specific behaviors (just as behaviorists do) but interpreting them in terms of underlying mental processes.³⁶

Behaviorism was an exciting adventure for experimental psychology but by the mid-1950s it had become apparent that it could not succeed. As Chomsky remarked, defining psychology as the science of behavior was like defining physics as the science of meter reading. If scientific psychology were to succeed, mentalistic concepts would have to integrate and explain the behavioral data. We were still reluctant to use such terms as “mentalism” to describe what was needed, so we talked about cognition instead.³⁷

“Behaviour” has for a very long time been a highly problematic term and that is a serious problem given it is such a key psychological term. It can refer to mindless, colourless movements, or to *meaningful* action, or even to how we “behave ourselves” in relation to social norms. As Tolman pointed out, Watson “dallied” with the first two senses.³⁸

All this stuff about behaviour as colourless movement (Hull’s term) when it comes to psychological research is, I insist, just plain rhetorical nonsense. What the behaviourists were *actually* studying were meaningful activities: rats trying to find a goal in a maze, or cats, when they could be bothered, getting out of puzzle boxes. Although, Skinner also dallied with the official term “behavior”, he finally came clean: «operant behavior is the very field of purpose and intention. By its nature it is directed towards the future».³⁹ In short, the behaviorists were hardly “behaviorists” in the reductionist and physicalist sense they are still cracked up to be.

All this should be ancient history, if not for the fact that many current psychologists have invoked this idea of “colourless movement” and made highly successful careers by coming up with incoherent theories of how we all get beyond “behaviour” (as colourless movement) to our fellow be-

ings' minds in relation to *Theory of Mind*.⁴⁰

5 Conclusion: Protecting the cognitivist core?

In 1974 I went to a seminar Imre Lakatos gave when I was a Ph.D. student at Birmingham. (He died just a few days later). His work was directed against Karl Popper's principle of falsification. His argument was that theories are not easily falsified at all. The core of a theory becomes surrounded by theoretical elaborations when problems are identified in order to protect the theoretical core – in this case what I will call “cognitivism”.

I am well aware that a lot of clever people have been elaborating the concept of “cognition” in various ways, most notably in relation to 4E cognition: embodied, embedded, enactive and extended. Curiously, other alternative approaches have been forgotten or side-lined, such as situated action, cognition in practice, ethnomethodology, and the ecological approach. They are, it seems, not part of the club.⁴¹

What I have tried to show in this chapter is that “cognition” was peripheral in psychological language until the 1960's and its entry was not completely smooth. Furthermore, some central aspects of the behaviorism that the “cognitive revolution” was supposed to have “overthrown” have been retained to the present day.

The term “cognition” has not only gone a long way since the 1960's but has also headed off in several different directions in relation to psychological theory, even within 4E. So, are we still talking about the same “thing” as the traditional cognitivists and, indeed, are the current generation of theorists talking about the same thing even among themselves? After all, as Kitchener⁴² has explained, the behaviourists ended up using a crazily diverse range of meanings for “behaviour”: so many that they could hardly keep track of them. Is it just possible that we ourselves don't really know what we are talking about when we talk about “cognition.” If so, let's forget it and find some better words.

Notes

¹ G. RYLE, *The concept of mind*, p. 258.

² W. JAMES, *The principles of psychology*, vol. I, p. 112

³ Entry to the index of William James's *Principles of psychology*.

⁴ G. KELLY, *Man's construction of his alternatives*, p. 91

⁵ Cf. U. NEISSER, *The imitation of man by machine*.

⁶ Cf. R. SHAW, *Theoretical hubris and the willingness to be radical: An open letter to James J. Gibson*, p. 246.

⁷ Cf. A. COSTALL, A. STILL (eds.), *Cognitive psychology in question*.

⁸ Cf. I. STUART-HAMILTON, Review of A. Costall, A. Still (eds.), *Cognitive psychology in question*.

⁹ Jonckheere (“Jonck”) was also visiting Oxford at the time and told me about the review and seemed un-

kindly amused. (I had earlier worked with him at UCL.)

¹⁰ U. NEISSER, *Cognition and reality: Principles and implications of cognitive psychology*, p. xiii.

¹¹ For some reason, Neisser seemed never to understood the simple logic of schema theory: when the material is strange (as with Bartlett's “*War of the Ghosts*”) recall should be difficult; but when the material is familiar and therefore assimilable to the available schemas, recall, according to the theory, should be relatively easy (see J. OST, A. COSTALL, *Misremembering Bartlett: A study in serial reproduction*; J. OST, J. UDELL, S. DEAR, J. ZINKEN, H. BLANK, A. COSTALL, *The serial reproduction of an urban myth: revisiting Bartlett's schema theory*).

¹² E.T.R. HILGARD, *The trilogy of mind: Cognition, affection, and conation*, pp. 106 and 116. Neisser makes essentially this same point in U. NEISSER, *The imitation of man by machine*.

¹³ B.R. GAINES, M.L.G. SHAW, *Personal construct psychology and cognitive revolution*.

¹⁴ Cf. R. JARDINE, *Elements of the psychology of cognition*; J. MC COSH, *Psychology: The cognitive powers*.

¹⁵ R. JARDINE, *Elements of the psychology of cognition*, pp. 1-2.

¹⁶ C. SPEARMAN, *Nature of intelligence and principles of cognition*.

¹⁷ Cf. T.V. MOORE, *Cognitive psychology*.

¹⁸ Cf. T.J. KNAPP, *The emergence of cognitive psychology in the latter of the twentieth century*.

¹⁹ Cf. R.S. LEEPER, *Cognitive processes*.

²⁰ *Ibid.*, p. 737.

²¹ W. KESSEN, *Early settlements in new cognition*, p. 168.

²² M. MOROZ, *The concept of cognition in contemporary psychology*, p. 178.

²³ W.J. MC KEACHIE, C.L. DOYLE, *Psychology*.

²⁴ A.S. BROWN, *Review of Information Processing and Cognition: The Loyola Symposium*, edited by Robert L. Solso, p. 357.

²⁵ D. HEBB, *The American revolution*, p. 737.

²⁶ U. NEISSER, *Cognitive psychology*, p. 7.

²⁷ Z.W. PHYLYSHYN, *Computation and cognition*, p. 68 - emphasis added.

²⁸ G. MANDLER, *Interesting times: An encounter with the 20th Century 1924*, p. 208. Joe Di Maggio was a famous base-ball player.

²⁹ This was the point about William James's complaint about the appeal to «this additional wheel work in the mind» (W. JAMES, *The principles of psychology*, vol. II, p. 112). For an uncritical celebration of “the cognitive unconscious”, cf. T.C. MEYERING, *Historical roots of cognitive science: The rise of a cognitive theory of perception from antiquity to the nineteenth century*.

³⁰ Cf. J. COULTER, *Rethinking cognitive theory*.

³¹ S. ASCH, *Social psychology*, Oxford University Press.

³² A. COSTALL, P. MORRIS, *The “Textbook Gibson”: The assimilation of dissidence*.

³³ U. NEISSER, *The future of cognitive science: An ecological analysis*, p. 248.

³⁴ U. NEISSER, *Commentary on “The demystification of mental imagery*, p. 561.

³⁵ G. MANDLER, *Interview with George Mandler*, p. 256.

³⁶ R.L. ATKINSON, R.C. ATKINSON, E.E. SMITH, D.J. BEM, S. NOLEN-HOEKSEMA, *Hilgard's introduction to psychology*, pp. 12-13 – emphasis added.

³⁷ G.A. MILLER, *The cognitive revolution*, p. 142.

³⁸ E.C. TOLMAN, *Purposive behavior in animals and men*, p. 6.

³⁹ B.F. SKINNER, *About behaviorism*, p. 61.

⁴⁰ Cf. I. LEUDAR, A. COSTALL (eds.), *Against theory of mind*.

⁴¹ Admittedly, Chemero draws upon ecological psychology but it is the reductionist Connecticut version (cf. A. COSTALL, *The hope of a radically embodied science*).

⁴² Cf. R.F. KITCHENER, *Behavior and behaviorism*.

Literature

- ASCH, S. (1987). *Social psychology*, Oxford University Press, Oxford, new edition.
- ATKINSON, R.L., ATKINSON, R.C., SMITH, E.E., BEM, D.J., NOLEN-HOEKSEMA, S. (2000). *Hilgard’s introduction to psychology*, Harcourt College Publishers, Fort Worth (TX), 13th edition.
- BROWN, A.S. (1976). *Review: Information processing and cognition: The Loyola symposium, edited by Robert L. Solso*. In: «American Journal of Psychology», vol. LXXXIX, n. 2, pp. 357-361.
- COSTALL, A. (2011-2012). *The hope of a radically embodied science*. In: «Behavior and Philosophy», vol. XXXIX-XL, pp. 345-353.
- COSTALL, A., MORRIS, P. (2015). *The “Textbook Gibson”: The assimilation of dissidence*. In: «History of Psychology», vol. XVIII, n. 1, pp. 1-14.
- COSTALL, A., STILL, A. (eds.) (1987). *Cognitive psychology in question*, Harvester, Brighton.
- COULTER, J. (1983). *Rethinking cognitive theory*, Macmillan, London.
- GAINES, B.R., SHAW, M.L.G. (n.d.). *Personal construct psychology and the cognitive revolution*, available at URL: <https://pages.cpsc.ucalgary.ca/~gaines/reports/PSYCH/SIM/SIM.pdf>.
- HEBB, D. (1960). *The American revolution*. In: «American Psychologist», vol. XV, pp. 735-745.
- HILGARD, E.T.R. (1980). *The trilogy of mind: Cognition, affection, and conation*. In: «Journal of the History of the Behavioral Sciences», vol. XVI, n. 2, pp. 107-117.
- JAMES, W. (1890). *The principles of psychology*, vol. I, Holt, New York.
- JAMES, W. (1890). *The principles of psychology*, vol. II, Holt, New York.
- JARDINE, R. (1885). *Elements of the psychology of cognition*, MacMillan, London.
- JOHNSON-LAIRD, P.N. (1980). *Mental models in cognitive science*. In: «Cognitive Science», vol. IV, n. 1, pp. 71-115.
- KELLY, G. (1969). *Man’s construction of his alternatives*. In: B. MAHER (ed.), *Clinical psychology and personality: The selected papers of George Kelly*, Wiley, New York, pp. 66-93.
- KESSEN, W. (1981). *Early settlements in new cognition*. In: «Cognition», vol. X, n. 1-3, pp. 167-171.
- KITCHENER, R.F. (1977). *Behavior and behaviorism*. In: «Behaviorism», vol. V, n. 2, pp. 11-71.
- KNAPP, T.J. (1986). *The emergence of cognitive psychology in the latter of the twentieth century*. In: T.J. KNAPP, L.C. ROBERTSON (eds.), *Approaches to cognition: Contrasts and controversies*, Lawrence Erlbaum, Hillsdale (NJ), pp. 13-35.
- LEEPER, R.S. (1951). *Cognitive processes*. In: S.S. STEVENS (ed.), *Handbook of experimental psychology*, Wiley, New York, pp. 730-757.
- LEUDAR, I., COSTALL, A. (eds.) (2011), *Against theory of mind*, Palgrave Macmillan, London.
- MANDLER, G. (1986). *Interview with George Mandler*. In: B.J. BAARS (ed.), *The cognitive revolution in psychology*, Guilford Press, New York, pp. 253-269.
- MANDLER, G. (2001). *Interesting times: An encounter with the 20th Century 1924-*, Erlbaum, Mahwah (NJ).
- MC COSH, J. (1886). *Psychology: The cognitive powers*, Scribner’s, New York.
- MC KEACHIE, W.J., DOYLE, C.L. (1966). *Psychology*, Addison-Wesely, Reading (MA).
- MEYERING, T.C. (1989). *Historical roots of cognitive science: The rise of a cognitive theory of perception from antiquity to the nineteenth century*, Kluwer, Dordrecht.
- MILLER, G.A. (2003). *The cognitive revolution: A historical perspective*. In: «Trend in Cognitive Sciences», vol. VII, n. 3, pp. 141-144.
- MOORE, T.V. (1939). *Cognitive psychology*, Lippincott, Philadelphia.
- MOROZ, M. (1972). *The concept of cognition in contemporary psychology*. In: J.R. ROYCE, W.W. ROZEBOOM (eds.), *The psychology of knowing*, Gordon and Breach, New York, pp. 177-205.
- NEISSER, U. (1963). *The imitation of man by machine*. In: «Science», vol. CXXXIX, n. 3551, pp. 193-197.
- NEISSER, U. (1967). *Cognitive psychology*, Century Crofts, Appleton/New York.
- NEISSER, U. (1979). *Commentary on “The demystification of mental imagery”*. In: «Behavioral and Brain Sciences», vol. II, n. 4, pp. 561.
- NEISSER, U. (1986). *Cognition and reality: Principles and implications of cognitive psychology*, W.H. Freeman & Co., New York.
- NEISSER, U. (1997). *The future of cognitive science: An ecological analysis*. In: D.M. JOHNSON, C.E. ERNELING (eds.), *The future of the cognitive revolution*, Oxford University Press, Oxford/New York, pp. 247-260.
- OST, J., COSTALL, A. (2002). *Misremembering Bartlett: A study in serial reproduction*. In: «British Journal of Psychology», vol. XCIII, n. 2, pp. 243-255.
- OST, J., UDELL, J., DEAR, S., ZINKEN, J., BLANK, H., COSTALL, A. (2002). *The serial reproduction of an urban myth: revisiting Bartlett’s schema theory*. In: «Memory», vol. XXX, n. 6, pp. 775-783.
- PHYLYSHYN, Z.W. (1986). *Computation and cognition*, MIT Press, Cambridge (MA).
- RYLE, G. (1949). *The concept of mind*, Hutchinson, London.
- SHAW, R. (2002). *Theoretical hubris and the willingness to be radical: An open letter to James J. Gibson*. In: «Ecological Psychology», vol. XIV, n. 4, pp. 235-247.
- SKINNER, B.F. (1974). *About behaviorism*, Random House, New York.
- SPEARMAN, C. (1923). *Nature of intelligence and principles of cognition*, Arno Press, New York.
- STUART-HAMILTON, I. (1988). *Review: A. Costall, A. Still (eds.), Cognitive psychology in question*. In: «Quarterly Journal of Experimental Psychology», vol. XL, 1988, pp. 415-416.
- TOLMAN, E.C. (1932). *Purposive behavior in animals and men*, Appleton Century-Crofts, New York.

Recensioni

Federico Faggin

Irriducibile. La coscienza, la vita, i computer e la nostra natura

Mondadori, Milano 2022

Collana: Gaia

Pagine: 296; € 20,90

Nel 2019 vedeva la luce, per i tipi di Mondadori, il volume dal titolo *Silicio. Dall'invenzione del microprocessore alla nuova scienza della consapevolezza* di Federico Faggin, fisico e imprenditore italiano, del quale tutti siamo in certo qual modo debitori, dal momento che egli è l'inventore del microprocessore, ovvero di quella tecnologia che sta alla base di qualsiasi *smartphone*, di cui ciascuno di noi è ormai dotato. Strettamente legato a quel testo è un secondo volume, pubblicato sempre per Mondadori e dato alle stampe nell'agosto del 2022 con il titolo *Irriducibile. La coscienza, la vita, i computer e la nostra natura*. Si tratta di un testo ambizioso che ha lo scopo di illustrare la posizione dell'autore sulla natura della consapevolezza o – come chiarisce sin da subito lo stesso Faggin – della *coscienza* (p. 11).

Il volume è di piacevole lettura, anche se non è sempre facile cogliere l'andamento delle argomentazioni, che spesso lascia fin troppo spazio a molteplici *gap*. Per meglio comprendere la finalità dell'opera sarebbe anzitutto necessario, a mio avviso, leggere prima l'*Introduzione* e poi la *Prefazione*, al fine di cogliere un elemento che costituisce un *file rouge* dell'intero discorso: uno spiccato sincretismo che si colloca a differenti livelli. Con il termine *sincretismo* non intendo essere eccessivamente critico; vorrei solo porre in risalto il tipo di approccio che l'autore adotta nei confronti dei temi trattati e che egli etichetterebbe piuttosto come *olismo*. Sincretismo, tuttavia, mi sembra l'espressione più appropriata.

Si tenga presente che non si tratta di un testo scientifico nel senso tecnico della parola e che, pertanto, il volume risponde ai canoni formali della saggistica generale e divulgativa (sebbene non manchino alcuni riferimenti ad articoli di periodici specializzati).

Il primo elemento sincretistico emerge dalla struttura del libro. Nelle due parti di cui si compone – una prima prettamente storica, una seconda più speculativa (il lettore edotto su certi argomenti potrà passare direttamente alla seconda parte) – si combinano, infatti, ricerca scientifica e vita personale, al punto tale da raggiungere, di tanto in tanto, i toni delle *Confessiones* di Agostino (p. 18). Credo proprio che quest'ultimo autore, che viene citato dallo stesso Faggin un paio di volte, possa aiutare a comprendere lo scopo dell'opera: suggerire un cammino di conversione speculare a

quello che l'autore dice di aver compiuto in giovane età. Quella volta egli passò dal dogmatismo religioso al dogmatismo scientifico (p. 19). Si può pensare che *illo tempore* quella conversione avvenne per via razionale. Ora il cammino di conversione suggerito al lettore è stato intrapreso, come racconta l'autore, per via "mistica", a seguito di un'esperienza significativa che lui chiama *risveglio* (pp. 20-21 e 160), in cui lo stesso autore può dire di sé: «ero contemporaneamente l'osservatore del mondo e il mondo. Ero il mondo che osservava se stesso!» (p. 20).

Da questa breve premessa potrebbe già emergere un secondo elemento di sincretismo, soltanto se non vengono assunti i presupposti dell'autore, il quale invita a non vedere più *mens* e *cor* come due organi disgiunti, ma come due parti dell'unico ente che noi siamo. Ed è questo *ente* (mi sia lecito il termine per il momento) che sta al centro dell'opera di Faggin.

La personalissima intuizione "mistica" di Faggin, infatti, converge nel testo con la teoria dell'informazione quantistica (teoria OPT) del fisico Giacomo Mauro D'Ariano dell'Università di Pavia. Dall'unione delle loro prospettive nasce il QIP, *Quantum Information-based Panpsychism*. All'interno del QIP è quindi possibile introdurre la nozione centrale per tutto il volume di Faggin, la *seity*, che significa letteralmente *possedere individualità* (p. 159). Nel caso specifico della teoria formulata da Faggin e D'Ariano, *seity* è «un ente cosciente che sa di esserlo, che può agire con libero arbitrio e che ha un'identità permanente come l'abbiamo noi» (p. 159).

Nell'universo, quindi, tutto ha coscienza, ma vi è una differenza sostanziale tra gli esseri umani, che «distinguiamo da coloro che sono coscienti ma non sanno di esserlo, e che pertanto mancano di identità propria», e questi enti che sono detti *thoughtform*, i quali sono enti quantistici che «non avendo un senso di sé, non possono guidare la loro esperienza con il libero arbitrio che non sanno di avere» (p. 159). Un organismo vivente, invece, è «sia quantistico sia classico, e può "ospitare" la coscienza e il libero arbitrio di una *seity* perché la sua parte quantistica può comunicare direttamente con la *seity* immediata informazione viva» (p. 174).

Le conclusioni alle quali si giunge, tuttavia, sembrano più supposte che provate: «Visto che lo stato puro non è conoscibile da fuori e non sarebbe conoscibile nemmeno da dentro, se non accettassimo la teoria QIP, esso non corrisponderebbe ad alcuna realtà» (p. 200). Sono tesi, infatti, che, come spesso ribadisce lo stesso autore, richiedono soltanto di essere accettate (p. 221).

A tal fine, però, la *seity* deve essere compresa in un contesto molto più ampio, che abbraccia tutto il

reale senza distinzione. In effetti per assumere questi postulati si rende necessario un nuovo modo di comprensione del reale che viene definito *olistico*.

Viene postulato, infatti, un «tutto olistico che contiene non solo le proprietà fondamentali che permettono l'evoluzione dell'universo inanimato, ma anche i semi del libero arbitrio, della coscienza e della vita» (p. 161). Impossibile non avvertire una eco plotiniana, soprattutto quando questo *Tutto* è definito *Uno* al fine di distinguerlo dal *Campo della fisica*, «perché da *Uno* emergono i campi coscienti con libero arbitrio (le *seity* elementari che chiamo UC, unità di coscienza) e non i campi inanimati delle particelle elementari della fisica, che interagiscono secondo leggi prestabilite» (p. 162). Non può sfuggire, inoltre, come i prodromi della riflessione siano profondamente leibniziani (come dichiara lo stesso autore a pagina 180); e ciò emerge quando si rende necessario esplicitare come le *seity* comunichino tra loro: «le *seity* elementari che emergono da *Uno* [...] comunicano tra di loro per approfondire la conoscenza di sé» (p. 171).

L'aspetto notevole dell'intera opera rimane comunque il voler indicare un cammino di conversione che faccia passare dal materialismo sfrenato, che domina per larga parte i contesti scientifici a partire dall'Ottocento – un orientamento che da «cocreatori dell'universo-vita» ci ha relegati a essere «effetti senza senso di qualche algoritmo classico presto dimenticato con la morte dei nostri corpi» (p. 162) –, a una teoria che non descriva più gli esseri umani come macchine biologiche analoghe ai computer (p. 22).

In questa ricomprensione generale del reale, sia dal punto di vista ontologico sia da quello epistemologico, la critica al monismo materialista di Faggin non deborda però nel dualismo sostanzialista, ma in una *visione altra*, che a tratti sembra assomigliare all'idealismo, inteso come monismo filosofico (p. 52), ma che dall'autore viene categorizzato come panpsichismo (p. 157), argomento al quale viene dedicato l'intero capitolo undicesimo, giacché essa è quella teoria filosofica che riconosce la coscienza come proprietà fondamentale, ovvero come proprietà irriducibile delle «particelle elementari». Il panpsichismo di Faggin, tuttavia, viene emendato grazie alla teoria QIP, così che questa versione perfezionata sembrerebbe risolvere il *problema della combinazione* che affligge le diverse forme di panpsichismo (p. 198).

Anche in questa scelta di categoria ontologica, tuttavia, emerge un certo sincretismo, dal momento che, se da un lato si difende il panpsichismo, dall'altro quello stesso dualismo che viene criticato (p. 106) riemerge sovente nelle modalità di espressione e di argomentazione (p. 150). Proprio quando si passa a trattare dell'essere umano, infatti, la visione panpsichista sembra dissolversi.

Prendendo l'esempio dell'essere-umano-Dante-Alighieri si dice che egli è una «*seity* incarnata nel suo corpo [di Dante Alighieri]» (p. 225). In questi e in altri passi sembra proprio che il dualismo giustamente ricusato – anche a parere dello scrivente – venga recuperato, e ciò venga fatto, ahimè, assumendone la sua variante peggiore. Per richiamare la famosa espressione di Gilbert Ryle, infatti, sembra che Faggin faccia ritornare l'incubo del «fantasma nella macchina» quando afferma, seppur *en passant*, che i robot possono «agire in maniera non-algoritmica [...] soltanto se sono controllati mediante informazione classica da *organismi viventi, a loro volta controllati da seity*» (174, corsivo mio). E ancora: «la *seity* è un ente cosciente che può anche «incarnarsi» in un corpo che è preposto al suo controllo. Il corpo è una struttura fisica quasi autonoma, quantistica-classica, che è distinta dalla *seity*» (p. 191). Una tale descrizione non potrebbe non attirare le molteplici critiche mosse al dualismo interazionista, e, sebbene l'autore se ne renda conto, non compie il procedimento argomentativo necessario per corroborare la sua tesi: «nessuno scienziato sa spiegarci come segnali elettrici o biochimici del cervello producano questi qualia. Il fatto che ciascuno di noi ne sia invece cosciente è incontestabile, e ciò falsifica l'idea che le teorie fisiche correnti siano complete, cioè che descrivano tutta la realtà» (pp. 23-24).

Andando ancora oltre si potrebbe vedere un recupero dell'impostazione popperiana (si rammenti che Popper fu uno dei sostenitori contemporanei di un certo dualismo delle sostanze), nel momento in cui ai tre mondi di Popper vengono sostituiti da Faggin altri tre mondi/spazi: lo spazio della coscienza, lo spazio dell'informazione (tra loro indivisibili) e lo spazio fisico. La *seity*, infatti, è «un «campo» in uno stato puro che esiste in una realtà più vasta del mondo fisico che contiene il nostro corpo» (p. 159). Tale *seity* esiste quindi anche senza il corpo fisico e, di conseguenza, la nostra esistenza non dipende dal corpo: «siamo *seity* che abitano temporaneamente i nostri corpi. Siamo esseri eterni, coscienti, non corpi deperibili» (p. 162).

Sebbene sia necessario mantenere la visione complessiva della teoria dell'autore e la sua impostazione olistica, sono innegabili dei veri e propri *salti* dall'epistemologia all'ontologia. Non viene descritto, infatti, come la dimensione-corpo interagisca con la dimensione-olistica della *seity*, tant'è che l'autore stesso sostiene che le *seity* non possono essere «né separabili da *Uno*, né essere fatte di parti» (p. 171), ma, nondimeno, che noi «siamo *seity* incarnate, siamo una combinazione misteriosa di *seity* e di corpo. A volte ci comportiamo come *seity*, altre volte come macchine, e quasi sempre come una combinazione delle due» (p. 237).

Altro elemento di sincretismo/olismo è la compresenza di riferimenti religiosi appartenenti a tradizioni differenti. Non mancano, infatti, vaghe

assonanze con la teologia cristiana. Sebbene si asserisca che «l'unica "immagine e somiglianza" di Uno che conosciamo siamo proprio noi» (p. 170), le conclusioni cui si perviene sono profondamente differenti, quasi bruniane (anche se l'autore a tal proposito si richiama a Schelling): «Uno è sia il creatore che la creazione e il beneficiario della propria traccia del proprio divenire [...] Noi siamo parti-intero di Uno, e ciò vuol dire che Uno è dentro di noi e noi siamo dentro Uno» (p. 170). In alcuni passaggi finali è possibile cogliere anche una certa vicinanza con la teoria novecentesca dell'*Intelligent Design*: «a mio modo di pensare, e correntemente con i principi della teoria QIP, la prima cellula vivente non si è autoassemblata per caso, ma è il risultato dell'enorme creatività e crescita di conoscenza delle seity. Esse, mosse dal bisogno di capire l'esistenza della riproduzione con cui loro stesse sono state create, hanno saputo manipolare cooperativamente i simboli vivi fino al punto di creare gradualmente un simbolo in grado di autoriprodursi» (p. 236).

Dal punto di vista strettamente filosofico l'autore si pone domande a suo dire innovative, ma che sono le medesime che si sono poste alcuni tra i più grandi autori della filosofia: Descartes (p. 52), Berkeley (p. 51), diversi filosofi del linguaggio (p. 54). È possibile anche avvertire l'eco, se pur indistinto, di Platone – «normalmente esiste una certa comunicazione tra l'ego [il corpo] e la seity, che si manifesta sottoforma di intuizioni, idee, slanci, emozioni, pensieri e immaginazione. [...] L'ego percepisce come oggetti ciò che nella realtà più vasta della seity sono combinazioni estremamente complesse di simboli vivi più o meno ordinati» (p. 192) – e un più vago riferimento a Hegel: «la seity che conosce diventa ciò che conosce» (p. 178). Nondimeno, se non sono nuove le idee, come quella per cui il tutto sarebbe maggiore della semplice somma delle sue parti (p. 123), è certamente nuova la prospettiva con cui queste sono offerte. Ciò che reputo cruciale del testo di Faggin, infatti, è la proposta di un nuovo paradigma (p. 154) che riesca a «unificare la fisica»; e per guadagnare questo paradigma, «dobbiamo abbandonare l'impostazione attuale e aprirci a una nuova visione» (p. 155).

Anche in questo caso ad essere notevole non è tanto la proposta in sé, ma la persona dalla quale proviene: un ingegnere, tra i padri del microprocessore, che invita ad abbandonare, quantomeno in via metodologica, l'assunto materialista da cui muove la maggior parte delle teorie scientifiche odierne: «non possiamo lasciare che il fiscalismo e il riduzionismo definiscano la natura umana e la natura più profonda dell'universo» (p. 24).

La proposta è ambiziosa, ma pienamente lecita; e l'autore ce lo spiega. È ambiziosa perché viene suggerito un ribaltamento di prospettiva e, in particolare, della prospettiva evolucionistica.

Quest'ultima, in particolare, non deve essere intesa *strictu sensu* – ovvero come la teoria dell'evoluzione della specie di Charles Darwin (teoria che l'autore sottoscrive) – ma come procedimento secondo cui una proprietà generale si evolve da una proprietà particolare (p. 155). Davanti a questo paradigma imperante giunge il suggerimento di Faggin: «far marcia indietro e considerare che possa esistere una realtà più profonda e finora insospettata da cui emerge la realtà fisica che conosciamo» (p. 155). L'universo inanimato, infatti, emergerebbe «dalle interazioni coscienti e libere delle seity» e questo «è esattamente l'opposto dell'interpretazione corretta, in cui la vita e gli enti coscienti sono diversi dalle interazioni della materia inanimata. Ciò che pensavamo fosse l'effetto ne è invece è la causa» (p. 174).

L'ambizioso programma di questa *inversione di marcia* coinvolge anche altre idee cardine del bagaglio concettuale occidentale: causalità (pp. 135-137), probabilità (p. 214), informazione (pp. 224-225) e, ancora oltre, invita a un complessivo ripensamento dell'intera impostazione della fisica classica (a cui è dedicata l'intera opera), oltre alla reintroduzione nel mondo della natura del principio teleologico (estromesso dalla fisica ormai da secoli). La seguente citazione esemplifica in maniera significativa questo invito a un *cambio di passo*: «i materialisti considerano ragionevoli le leggi della fisica classica, ma queste non contemplano che la Natura abbia la coscienza e il libero arbitrio che ci distinguono. Se la natura è cosciente e ha libero arbitrio, come suppongo, le leggi fondamentali della fisica devono essere deterministiche e probabilistiche, proprio come lo sono le leggi della fisica quantistica. L'assurdità non proviene quindi dalla Natura, ma dal pregiudizio umano che ritiene che le leggi debbano essere deterministiche» (p. 166).

Dalla formulazione (intellettualmente feconda) di queste tesi si passa, tuttavia, alla tacita ammissione di due postulati, la coscienza e il libero arbitrio (p. 156), i quali non starebbero a conclusione del cammino evolutivo – così come autori, anche molto distanti tra loro (si pensi a John Searle e Daniel Dennett) hanno asserito – ma al suo inizio (p. 162). Questo dato è, a detta di Faggin, «un miracolo che dobbiamo accettare» (pp. 161-162).

La tesi mantiene, quantomeno in via ipotetica, tutta la sua validità; e concordo col giudizio dell'autore quando sostiene che la «QIP rappresenta un'innovazione significativa. Essa cambia radicalmente l'interpretazione della realtà, perché ci dice che la coscienza e il libero arbitrio esistono da sempre e, nel contempo, mantiene la credibilità della fisica radicata sui risultati sperimentali ottenuti, che ora trovano una spiegazione molto più soddisfacente» (p. 232).

Faggin avrebbe potuto benissimo fermarsi all'enunciazione della tesi e alle prove della sua

fattibile percorribilità, piuttosto che azzardare una storia dell'evoluzione della *seity* – scadendo nel capitolo undicesimo, per quanto mi è dato di intendere, in una sorta di antropomorfizzazione delle *seity* (pp. 200-202). Dall'enunciazione di una tesi, infatti, non è lecito compiere inferenze affrettate come le seguenti: la «QIP dà ora pieno supporto alla mia conclusione che è la materia che può essere spiegata come una proprietà emergente della coscienza» (p. 231) o «è assurdo pensare che semplici molecole organiche, che rappresentano una frazione infinitesima della complessità del più semplice batterio conosciuto, abbiano potuto trasformarsi e autoassemblarsi fino a formare un

organismo in grado di autoriprodursi senza nessuna guida» (p. 236).

Una cosa rimane fuor di dubbio: chiunque legga il testo – storici della filosofia o filosofi della mente, filosofi del linguaggio o filosofi della scienza – troverà interessanti spunti (oltre a una miriade di preziosissimi aforismi); e, qualora questi non lo soddisfacessero, sicuramente troverà un punto di vista differente che proviene da un ambito sinora ermetico, com'è sempre stato quello dell'ingegneria più *hard*.

Cristiano Cali

*Dipartimento di Filosofia e Scienze dell'Educazione
Università degli Studi di Torino*

Recensioni

Riccardo Manzotti

La mente allargata. Perché la coscienza e il mondo sono la stessa cosa

Traduzione di Allegra Panini

Il Saggiatore, Milano 2019

Pagine 340; € 25,00

Che cos'è la coscienza e in che relazione è con la materia? Dove si colloca l'esperienza che ciascuno di noi vive in prima persona nello spazio ontologico della natura? Come si passa dall'attività metabolica del cervello alla particolare esperienza di un colore, di un suono, di un'emozione?

Queste domande, in forme diverse, costituiscono gli interrogativi che occupano i filosofi da tempo immemore e trovano precisa collocazione nel progetto di naturalizzazione del mentale. Obiettivo di questo programma è risolvere – o dissolvere – *il problema difficile della coscienza* formulato da Chalmers: come è possibile che dalla materia, in particolare dal cervello, si passi all'esperienza cosciente. In questa cornice naturalistica si inserisce la proposta radicale di Riccardo Manzotti.

Rifiutata l'identità mente-cervello, l'autore abbraccia una posizione esternalista: l'esperienza di una mela è identica a nient'altro che alla mela stessa. Coscienza e mondo sono la stessa cosa. Sebbene respinga la localizzazione della coscienza nel cervello, la teoria dell'identità mente-oggetto si iscrive a pieno titolo nella famiglia delle teorie physicaliste. L'esperienza fenomenica è parte del regno della natura e le leggi a cui è soggetta sono quelle della fisica.

La mente allargata è un viaggio in una nuova analisi ontologica della coscienza, in cui il lettore viene accompagnato passo dopo passo. Traspare una scelta linguistica volutamente lontana dai concetti filosofici tradizionali, l'*explanandum* è il rosso della mela, non il *quale-rosso* o la *rossezza-del-rosso* di cui spesso parlano i filosofi. A giudizio di Manzotti ogni termine tecnico è espressione di un particolare punto di vista, che lo stesso Manzotti, in questa sede, si sforza di abbandonare prima di intraprendere il lavoro della scrittura. Al lettore viene chiesto di mettere da parte le categorie classiche con cui si è abituati a pensare la coscienza; in cambio viene restituita una radicale e nuova ipotesi sul mentale.

L'analisi parte dal fossato galileiano, la differenza che Galilei ha descritto fra le qualità dell'esperienza (soggettive e non misurabili) e quelle del mondo fisico (oggettive e misurabili). Da una parte c'è l'esperienza e dall'altra il mondo, una porzione di materia che si riferisce ad altra materia. Fra questi due poli la storia della filosofia ha cercato di costruire dei ponti per spiegare come sia possibile che la materia cosciente divenga altro da sé. Manzotti rifiuta ogni mediazione fra il cervello e l'oggetto. La tesi proposta è molto ecologica: non esiste una

relazione fra soggetto e oggetto, l'unica relazione che esiste in natura è l'identità, l'esperienza della mela è identica alle proprietà della mela. L'esperienza della mela è la mela relativa a un corpo; la chiusura causale del mondo fisico è rispettata. Il corpo è la condizione d'esistenza di una proprietà che è esterna ed è dell'oggetto, non del cervello. In contrasto con le teorie allucinatorie della percezione, Manzotti descrive tutta l'esperienza come percezione, anche le esperienze relative a *casi anomali* quali illusioni, sogni e allucinazioni.

La lettura di *La mente allargata* apre il campo ad alcune domande: cosa fanno cervello e corpo per essere la condizione d'esistenza dell'oggetto relativo? In cosa si differenziano rispetto ad altri oggetti naturali o macchine in rapporto causale con l'ambiente?

Se la seconda domanda richiama ancora una volta il *problema difficile della coscienza*, la risposta alla prima richiede alla teoria dell'identità mente-oggetto un salto dall'ontologia all'epistemologia. Se il corpo è ciò che permette all'oggetto di esistere in quanto effetto-esperienza, si può legittimamente nutrire il sospetto per cui, a un riesame dell'attività cerebrale e dell'azione corporea alla luce della teoria della *mente allargata*, la distanza, ora abissale, rispetto alle teorie neuroscientifiche potrebbe ridursi.

Per esempio, la descrizione percettiva delle allucinazioni potrebbe non essere così incompatibile con la teoria opposta: il modello allucinatorio della percezione. Infatti, secondo Manzotti, le allucinazioni sono percezioni di oggetti ricomposti di cose incontrate in precedenza. E, tuttavia, la ricombinazione di percezioni esistenti (con il modello a doppia inibizione) è coerente anche con l'idea per cui le allucinazioni siano un film creato dal cervello con percezioni avute in precedenza.

Molto probabilmente questo nodo potrà essere sciolto e verificato mediante una indagine empirica vertente sulla causalità e sull'informazione, in cui anche il concetto di rappresentazione (che Manzotti rifiuta nettamente) dovrà essere preso nuovamente in considerazione e riesaminato.

Un'analisi empirica potrà conferire maggiore peso a una delle due teorie così come portare alla luce un legame tra di esse, mostrando come non ci si trovi di fronte a una profonda e inconciliabile dicotomia. Se si assume che le relazioni non facciano parte dell'ontologia del mondo, non possiamo far altro che localizzare la coscienza nel cervello (e, quindi, nel corpo) oppure nell'oggetto. Se rinunciamo a questo assunto, o almeno facciamo *epoché*, potremmo ripensare la coscienza come un processo che unisce cervello, corpo e mondo.

Delle ulteriori linee di ricerca per la teoria dell'identità mente-oggetto potrebbero riguardare la

questione del sé a partire dall'essere un polo unificante delle proprietà dell'oggetto esterno, gli stati d'animo e le emozioni.

Il variegato quadro teorico della filosofia della mente è ora arricchito da un'altra originale proposta. *La mente allargata* è un tentativo di fornire una risposta a una domanda esistenziale relativa a *chi siamo*. Il fatto che non ci sia una differenza fra la velocità di un corpo e il rosso della mela come proprietà relative fornisce una immagine di un universo in cui la coscienza non ha una natura particolare rispetto agli altri fenomeni fisici.

Le immagini descritte suggeriscono un universo fatto da una unica sostanza, dove forma e materia non sono separate. Fra la dimensione spaziotemporale e gli oggetti relativi questo libro ci ricorda continuamente che noi siamo il nostro mondo. L'auspicio è che possa tracciare nuove prospettive per una più antica concezione della persona come soggetto esteso. Se la mente è allargata, allora, quanto è importante scegliere con cura la nicchia ecologica nella quale viviamo?

Daniela Troiano
Ricercatrice indipendente