



Topics in Cognitive Science 12 (2020) 1224–1240

© 2020 The Authors. *Topics in Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society

ISSN:1756-8765 online

DOI: 10.1111/tops.12503

This article is part of the topic “Levels of Explanation in Cognitive Science: From Molecules to Culture,” Matteo Colombo and Markus Knauff (Topic Editors). For a full listing of topic papers, see [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1756-8765/earlyview](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview)

## Editors’ Review and Introduction: Levels of Explanation in Cognitive Science: From Molecules to Culture

Matteo Colombo,<sup>a</sup> Markus Knauff<sup>b</sup>

<sup>a</sup>*Tilburg Center for Logic, Ethics and Philosophy of Science, Tilburg School of Humanities, Tilburg University*

<sup>b</sup>*Department of Psychology, Experimental Psychology and Cognitive Science, Justus Liebig University, Giessen, Germany*

Received 26 February 2020; accepted 11 April 2020

---

### Abstract

Cognitive science began as a multidisciplinary endeavor to understand how the mind works. Since the beginning, cognitive scientists have been asking questions about the right methodologies and levels of explanation to pursue this goal, and make cognitive science a coherent science of the mind. Key questions include: Is there a privileged level of explanation in cognitive science? How do different levels of explanation fit together, or relate to one another? How should explanations at one level inform or constrain explanations at some other level? Can the different approaches to the mind, brain, and culture be unified? The aim of this issue of *topiCS* is to provide a platform for discussing different answers to such questions and to facilitate a better understanding between the different strands of thinking about the right levels of explanation in cognitive science.

*Keywords:* Levels in cognitive science; Explanation; Unification; Reductionism; Scale; Representations

---

---

Correspondence should be sent to Matteo Colombo, Tilburg Center for Logic, Ethics and Philosophy of Science, Tilburg School of Humanities, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, North Brabant, The Netherlands. E-mail: [m.colombo@uvt.nl](mailto:m.colombo@uvt.nl)

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## 1. Introduction

How should we study cognition? How can we understand how the mind works? Such questions are inextricably linked to the notion of *levels of explanation* for cognitive phenomena. Some researchers argue that cognitive phenomena are best studied and understood at some particular level, for example at the functional level, at the level of neural networks in the brain, or at the level of biological or cultural evolution. Such positions are sometimes associated with the conviction that the ultimate aim of cognitive science is to establish a single, comprehensive account of cognitive phenomena based on their causal mechanisms, or on a single set of principles, at some privileged level of explanation.

Other researchers disagree and maintain that for any cognitive phenomenon there is no privileged level of explanation. From this view, a full understanding of cognitive phenomena requires explanations that integrate multiple levels of explanation. For these researchers, cognitive science should pursue integration and become a genuine interdisciplinary endeavour.

A third position is that the explanatory targets of cognitive science have a much more disunified and unstable character than the targets of disciplines such as physics. Hence, the understanding of cognitive phenomena necessarily resembles a patchwork of relatively autonomous levels and approaches. These are just three positions, and there are many other opinions in between.

Intuitively, it is plausible that any cognitive phenomenon can be studied and explained at different levels. For example, if you want to explain how you can learn something new by reading this text, you might seek an explanation at some “high” level, considering psychological, social, or cultural factors that are relevant to acquiring new knowledge by reading. But you might also pursue explanations at some “lower” levels, for instance, the functional level of syntactic parsing, lexical processing, and memory update. Or you might go at “lower” levels still, and study eye movement or brain activity during reading.

But does it really make sense to talk about different levels of explanations in cognitive science? Does the notion of a *level of explanation* play any important role in studying and understanding cognitive phenomena? A positive answer would beget several other questions. What exactly is a level of explanation? Is there any privileged level of explanation for a given cognitive phenomenon? How are different levels of explanation in cognitive science related, or how should they be related? How can unified multi-level accounts of cognitive phenomena be effectively pursued? A negative answer would instead suggest that the notion of *level of explanation* is confused and does not do much epistemic work in cognitive science. *Level of explanation* would be a generic notion related to a number of distinct research strategies and questions, which are more precisely defined in terms of concepts such as scale, composition, complexity, and hierarchy.

Most cognitive scientists have reflected on these questions at some point in their work. Perhaps you also discussed these questions with other colleagues. The authors in this topic titled “Levels of Explanation in Cognitive Science: From Molecules to Culture” are cognitive scientists with different disciplinary backgrounds, including cognitive

psychology, philosophy of science, robotics, neuroscience, evolutionary biology, anthropology, and computer science. When we invited the authors, we asked them to write their contribution, keeping in mind the following questions:

- What is a level of explanation for cognitive phenomena? Is there a privileged level or kind of explanation in cognitive science? How could we tell?
- How do different levels of explanation fit together, or relate to one another? How should explanations at one level inform or constrain explanations at some other level?
- Can the different approaches to the mind, brain, and culture be unified? Or is a plurality of approaches and levels of explanation a genuine feature of cognitive science? What would it take to unify or integrate different levels of explanation?
- What is reductionism in the sciences of mind, brain, and culture? How does reductionism promote or hinder our understanding of the mind?
- Which kind of explanations should be more represented in the future of cognitive science?

Such questions have always been the subject of controversy. So it is unsurprising that this is not the first *topiCS* on the theme. For example, there is another issue devoted to David Marr and “Levels of analysis in Cognitive Science” (Peebles & Cooper, 2015). However, the authors of the present volume express opinions that go far beyond the debates about Marr’s trichotomy between what is computed, how it is computed, and in which hardware it is computed.

In the remainder of this introduction, we first provide an overview of the different contributions in this special issue. Then we are going to do an experiment: We present a dialogue between a student of cognitive science and two professors representing different views on the theme of this volume. The aim of this dialogue is to lay out a complex pattern of arguments and counterarguments concerning the topic of levels of explanation in the cognitive science to a wider readership. We close with some general remarks and a list of recommended readings.

## 2. Outline of the contributions in this topic

The aim of the issue is to acquaint readers with a large range of different positions on levels of explanation in cognitive science. The contributions should not only be variations of similar positions, or positions which are already mainstream in cognitive science. The goals are rather to mirror the complexity of the issue, to stimulate further discussion, and to serve as a helpful resource for students of cognitive science. The fact that we have achieved these goals—we think—is thanks to our outstanding authors and their clear way of laying out their positions for a broad cognitive science audience. It is also thanks to the reviewers for this journal, who offered constructive feedback on all the contributions.

The first article in this volume is by John Bickle, who discusses the distinction between mechanistic approaches to levels of explanation and his own “ruthlessly

reductive” approach. His argument is based on a detailed examination of experimental practices in cognitive neuroscience that rely on optogenetics and DREADDs (Designer Receptors Exclusively Activated by Designer Drugs). His main claim is that some work in current neuroscience is ruthlessly reductive, since it aims to directly link mind and molecular activities in the brain.

Using the case of microscopic and macroscopic dynamics displayed by recurrent neural networks, the second article, by Gregor Schöner, argues that the dynamics of neural populations form a privileged level. This level would be privileged, because it would display the general laws and principles (not the mechanisms) of behavior and cognition.

Maxwell Bertolero and Danielle S. Bassett examine network neuroscience and argue that this increasingly popular framework can bridge explanations at multiple scales from a micro to a macro level in the brain, while it can also highlight aspects of the neural mechanisms producing cognitive phenomena.

Bernhard Hommel considers the concept of mechanistic explanation, which stands in contrast with the concept of a fundamental laws or fundamental causal relation. His main claim is that the great majority of explanations in psychology and cognitive neuroscience is “pseudo-mechanistic,” and that cognitive neuroscience should move beyond this “Aristotelian” phase to become a more mature “Galilean” science that uncovers actual mechanisms.

Angela Potochnik and Gui Sanches De Oliveira also discuss the role of cognitive mechanisms. They provide a survey of reductionist, mechanist, and pluralist approaches for explanation in the cognitive science. They argue that these approaches are best understood as different styles of explanation that capture different, cross-cutting patterns in cognitive phenomena. Discovering mechanisms should not be the only explanatory goal of cognitive science.

Sara Aronowitz and Tania Lombrozo distinguish between what they call “experiential explanations” and “abstractive explanations,” and provide empirical evidence from experiments in cognitive psychology that people use both modes of explanation. In particular, they argue that experiential explanations play a special, distinctive epistemic role within the psychological and social sciences.

Andrew Shtulman and Cristine Legare focus on “explanatory co-existence,” which is the cognitive process by which people appeal to seemingly incompatible explanations of the same phenomenon at once. Based on new experimental data, they argue that this process “presents a unique challenge to understanding how distinct mental representations may give rise to the same behavioral phenomena.”

Gerd Gigerenzer distinguishes three frameworks for finding explanations of cognitive phenomena, namely: a tools-to-theories framework, an “as-if” framework, and an “adaptive toolbox of heuristics” framework, which he considers to be a process-oriented approach to human rationality. Discussing their virtues and limitations, he clarifies how these three approaches have “considerable potential to inform each other and to generate points of integration.”

Patricia Rich, Mark Blokpoel, Ronald de Haan, and Iris van Rooij also examine the adaptive toolbox of heuristics in the context of the question of how evolved systems can

tractably implement cognitive functions. They consider three possible answers to this question, namely: Resource Rationality, the Adaptive Toolbox, and Massive Modularity. Their claim is that none of these answers resolve the tractability challenge. They just relocate it to the level of evolution.

This volume of *topiCS* ends with a contribution on the cultural evolution of cognition by Andrea Bender. Her paper uses the examples of writing, number representation, and causal cognition to argue that cultural evolution has specific causal influence on cognition. For example, culture effects gene selection in a population, the neural substrate of cognitive capabilities of individuals, and the evolution of culture itself. This position and many others are represented in the following fictitious dialogue.

### 3. A dialogue on the right level of explanation to understand the mind

This dialogue allegedly took place at one of the Annual Meetings of the Cognitive Science Society, and it has three protagonists: *Alex*, a student of cognitive science, and two well-known professors, Professor B from the University of Brain City and Professor M from the University of Mind City. Alex just presented a poster for the first time at an international conference, received very good feedback on the poster, and started to think...

“I’m now seriously considering to apply for a PhD program in cognitive science. There are two professors over there right now, who work at different universities, which both have an excellent reputation for their cognitive science programs. But I’ve also heard that the two programs are different in many ways. Most of the professors at the University of Brain City use brain imaging and techniques from genetics to try to understand the biological foundations of cognition. The program at the University of Mind City relies mainly on methods from cognitive psychology, artificial intelligence, linguistics, and anthropology to understand cognition beyond the level of biology. So I may well introduce myself to Professors B and M and ask them about their two programs and their ‘philosophy.’”

“Hi, I am Alex!”—I started the conversation. “I’m interested in your opinion, if you have a moment. I’d love to apply for a PhD position in cognitive science, and I’m considering applying to the programs at your universities.”

The two professors were happy to tell me about their programs. They initially asked me if I liked any particular talk at the conference. After thinking about it for a moment, I replied: “Did you hear this talk on understanding human cognition from the ground up at the level of genes? I very much enjoyed it. The speaker argued that genetic factors will become a more prominent level of explanation in the cognitive sciences. What do you two think about that?”

Professor B, who is working at the University of Brain City and well known as an advocate of biological approaches to understand the mind said: “No, I didn’t hear that presentation.”

Professor M, from the University of Mind City, did not visit the talk either, but asked me what the main claim of the talk was, and why I liked it.

I stammered a bit, but then said: “I think, the main claim was that genetic variation can help to predict cognitive performance in a wide swathe of tasks. Certain genes would predict, for example, which people are the best at solving different spatial problems, or who is at risk to develop specific spatial disorders.”

Professor M, a well-known critic of cognitive scientists trying to understand how the mind works only by looking in the brain or genes, asked me right away: “But did the speaker say anything substantial about the cognitive mechanisms of spatial problem-solving, or of any other cognitive ability?”

Before I could answer, Professor B intervened: “Well, you’re asking too much. At this stage of research, it’s unrealistic to fully explain how something like spatial cognition works from the ‘bottom-up’—all the way from individual genes to knowledge structures for spatial navigation.”

“But that should be the goal, right?”—I asked. Professor M did not hesitate and argued that human reasoning and problem-solving rely on cognitive representations: “I don’t think that we can, or should, reduce phenomena that appear on the cognitive level all the way down to the level of genes and molecules.”

I was a bit disappointed, and wondered: “But why not? I think it would be exciting if we could understand the genetic components of cognitive capacities.”

Professor M responded: “I’m sceptical about how much that helps, when we want to understand how spatial thinking works and why it works the way it does. Results from cognitive research also allow us to predict human performance, for instance, when people get lost or draw false inferences in reasoning. Predicting behavior is also an important goal, but I’ve never seen that this is possible based on brain imaging or DNA sequencing.”

I asked: “But wouldn’t evidence from biological levels also be relevant to confirm or disconfirm competing cognitive explanations?”

Professor B nodded: “I agree with you that the levels of genes, molecules, and neural pathways are relevant sources of evidence about the correct explanatory model of spatial thinking.”

Professor M looked at Professor B and said: “But most explanations of how genes affect human cognition are currently too unconstrained to give us testable predictions. And the existing empirical evidence is too weak.”

“I think there’s a risk of conflating explanatory issues with issues of evidential relevance here”—said Professor B —“One issue is: ‘What is the relevant evidence to evaluate an explanation?’ A different issue is: ‘To what level does the explanation belong?’ Consider, for instance, alternative models that explain spatial reasoning in terms of representations, processes, and resources. These models belong to the cognitive level. But this doesn’t mean that evidence from neuroscience and genetics cannot be relevant to evaluate them.

I interrupted: “You know, I have just read an article where the researchers empirically identified 18 candidate genes for depression, which have often been studied for their relevance to depression phenotypes. No evidence was found for any of these candidate genes to be associated with depression. None of these candidate genes for depression were more

strongly associated with depression phenotypes than noncandidate genes. I was disappointed.”

Professor B replied: “Don’t worry, there’s still a lot more work to do before we can answer questions about the genetic pathways and mechanisms underlying cognitive capacities and mental illnesses; but it’s an exciting time. We have many new experimental tools we can use to examine and test cognitive models at a micro-level. Our lab at the University of Brain City just bought state-of-the-art technologies for neurostimulation and imaging. I once collaborated with a geneticist on a project on intelligence and conditional reasoning; and the geneticist was much less reductionist than what many of our colleagues in cognitive science, including my colleague here, seem to believe.”

“And how did it work?” I asked.

Professor B answered: “The main challenge was to find a common vocabulary for talking about the same things. It took some time, but it eventually worked.”

Professor M replied: “I’ve heard several talks from genetic psychology, and I must say that they were always disappointing. The effect sizes were often extremely small, if there were any. Typically, there was no interpretable pattern of results in these studies. The only thing most of these studies indicate is that the people who do this research are strong believers of biology as key to understanding cognition.”

I was finding this discussion entertaining. Professor B said sharply: “Not sure that’s fair. Those studies certainly show, at least, that we should clearly define *what* phenomenon we want to explain, before we can say anything sensible about the most adequate level in its causal structure, at which testing should be conducted, and understanding pursued.”

A question was already revolving in my mind: “Professor B, you mentioned your collaboration with a geneticist on conditional reasoning. But aren’t the norms of what we consider rational defined by social agreement? It seems strange to use the vocabulary of biology to explore accurate or fallacious conditional reasoning. Or am I wrong?”

Professor B said: “No, you’re not wrong, in a way. But I think many colleagues would agree that cognitive theories should at least be consistent with relevant neural evidence, otherwise they are wrong.”

“But why shouldn’t it be the other way around?” I asked.

Professor M smiled: “Good point, I actually think that a theory on the neural level must be consistent with what we know on the cognitive level; otherwise it is wrong. We know much more on the cognitive level than on the neural level, and most of the terms that cognitive neuroscientists use are cognitive terms. Imagine if neuroscientists just talked about synapses, dendrites, neurotransmitters, action potentials, and so forth. It would be rather boring, don’t you think?”

“Well, sorry, but this is a bad argument,” Professor B protested. “You know that brain research doesn’t boil down to cognitive neuroscience. It includes important—and, if you ask me, exciting—fields like neuroanatomy, neurochemistry, and biophysics. These fields can do just fine without cognitive terms. And their results are relevant to understanding cognitive capacities anyway.”

“Okay, brain research might be exciting just as physics and many other areas of the sciences and humanities”—said Professor M. “But I disagree”—M continued—“regarding the importance for cognitive science. Reductionist approaches can tell us little, if anything, about the nature of cognitive phenomena.”

“I am sorry”—I said—“but what do you exactly mean with *reductionism*? You mentioned this term already, but I don’t know what it exactly means.” Professor B explained that reductionism is a term that broadly refers to the view that we can fully understand cognitive phenomena by studying their underlying physical or biological component structures like genes and neurons.”

But then I asked: “Do cognitive scientists really believe that any cognitive phenomenon, spatial navigation or conditional reasoning for example, can be adequately explained at a single level? My sense is that most think molecular biology and its techniques are just one method, and should not serve as the only, or the best, model for gaining understanding in cognitive science.”

“You’re right,” Professor B replied. “But, just to tell you about another ‘ism,’ the view you have just described may be better called *monism*. That’s the opinion that there is only *one* correct level of explanation or *one* correct method targeting the correct level of explanation for a phenomenon.”

“I am sorry again. But what do you mean with *level*?”

Professor B explained: “When I use the term ‘level’ in my work, I simply mean ‘some scientific domain of interest.’ Nothing deep.”

“Does that mean that you think there’s no ‘right’ level for explaining any cognitive phenomenon?”, Professor M asked the colleague.

“You know what”—B admitted—“I really don’t have a rigorous justification for what makes a level ‘right’. I don’t think we need one. What the ‘right level’ is depends on the specific context of research.”

I was surprised because Professor M agreed: “It makes little sense to talk about a ‘right level’ of explanation in a vacuum, without taking into account our background knowledge about the phenomenon we want to understand, relevant theoretical and practical interests, and so on. When I use the term ‘level,’ I have in mind certain properties of a mechanism, like scale, granularity, or hierarchical composition.”

Once again I had to ask: “What do you mean by that?”

“I mean,” Professor M answered, “that higher levels are composed of things at lower levels. Explanations of cognitive phenomena at lower levels would refer to smaller entities, or to faster interactions between entities. Explanations that refer to axons are at a lower level than explanations that refer to, say, cultural processes, which rely on spatially distributed social networks.”

Professor B clarified that the kind of hierarchical composition Professor M was talking about is not a hierarchical model of scientific disciplines or theories, where each level corresponds to a theory, and fundamental physics is at the bottom level. I nodded, and Professor B continued to explain: “And it is important to see that scale is also related to the notion of *complexity*. Think of consciousness. Many scholars claim that conscious experience emerges from interactions among brain components at a micro-scale. Thus,



even if we perfectly knew how brains are organized and wired, we may not be able to predict the more complex phenomena emerging at a macro-scale.”

Professor M suddenly interjected: “Right, assume that Professor B here finds a 1-to-1 mapping between any two particular mental and physical states at least in one individual. This discovery would not be of much interest, because scientists are interested in general principles, not just explanation for single observations. So we want to find identities between kinds of mental states and kinds of neural states in the brain. Yet finding such identities is impossible, because any given type of mental state can be realized by many distinct types of brain states, both across and within individuals. And any given brain state can implement many different mental states.”

“That’s too quick” objected Professor B. “The idea that mental states are *multiply realized* by different types of physical states is less prevalent in scientific practice. The physical differences between humans come with functional differences and this means that it is really hard to identify a genuine case of multiple realization, where two different physical structures realize the same psychological capacity, but in very different ways.”

Professor M said: “This is really a strong statement. When you and I think about the concept of “love” that relies on the same neural processes in our brains? Okay, we might not have the same concept of love, but even if we had, that would not be represented in the same way in our brain because the concept is acquired by learning and what has been learned is mapped onto different neural structures.”

Professor B cringed, and turned to me: “How do you understand the term ‘level’?”

The first thing that came to mind was David Marr’s three-level framework for analysing cognitive systems, which I had studied in different courses in my Master’s.

Both professors smiled.

B said: “If anything, Marr’s three-level framework stands for one of the fundamental convictions of the cognitive sciences, namely that the mind consists in some sort of information processing.”

“Why?”—I asked—“Is there any other way to characterize what cognitive science is about?”

Professor M answered: “Well, an increasing number of people say there is. Many who subscribe to dynamical systems theory criticize the information processing paradigm. They say that cognitive systems should be studied in terms of their dynamics and interactions with the environment, instead of computations over representations. The idea the mind is an information processing system has become something like an unquestioned axiom of our discipline. Over the last decades, the term ‘information’ has developed into a vague and overly used term.”

I was a little confused at this stage, as I assumed that cognitive scientists do not have to decide between either dynamical systems theory, or computation and information. That seemed a false dichotomy to me.

Professor B clarified: “It’s important to remember, though, that not all information consists in symbolic, language-like representations. For example, at the representational level, we assume the states of a calculator represent specific numbers. We ascribe meaningful representational content to different states of the calculator. At the syntactic level,

we individuate operations on meaningless numerals; at this syntactic level, strictly speaking, the calculator doesn't represent numbers and doesn't perform arithmetical operations over numbers."

Growing impatient, Professor M asked: "What's your point, B?"

"My point is that to be a good calculator, it's not necessary the calculator employs symbolic representations in a way that involves comprehension of arithmetic. I think the debate between representationalists and anti-representationalists often seems to neglect this point, confusing representational ascriptions to a system with ascription of comprehension."

I interrupted and said: "The pocket calculator example reminds me of what I have learned in my classes about the question of whether a thermostat is a cognitive system. I think it isn't. Whenever the sensor detects a certain temperature, then it turns on or off the heating. It performs this function without any flexibility; a certain input always leads to a certain output. But that is not how cognitive systems behave. They respond flexibly to the input from the environment and make predictions. Don't you think?"

Professor M agreed and said that this is a topic of one of the popular courses in their program at Mind University.

Still a little confused, I asked for clarification: "Many cognitive scientists, also at this conference, seem to offer explanations in terms of representations. Does that mean the level of representations is particularly important to explain cognitive phenomena?"

Professor M replied confidently: "Genuinely cognitive phenomena can be adequately explained only by considering this level. For instance, when we want to deal with what the philosopher Daniel Dennett calls taking an intentional stance toward a cognitive system's behavior, then representations are essential in studying cognitive systems. We know from the history of psychological research that explanations that do not account for internal representations and processes don't have much power. The cognitive turn showed the assumption of intermediate internal processes can explain how the human mind works in a much more powerful way than representation-free approaches. Chomsky showed that for language, Bandura and Bruner for learning, Atkinson and Shiffrin for memory, the list is endless."

"But"—I insisted—"do you think that all adequate explanations of cognitive phenomena should be pitched at the level of algorithms and representations? It seems to me that capacities like perception and motor control may be better explained in terms of a representation-free, dynamic, and distributed interaction between organisms and their environment."

Professor B said: "Our goal, at least the 'research philosophy' at our program at the University of Brain City, is not to explain the less fundamental, whatever that is, in terms of the more fundamental. Our goal is to integrate different explanations, both at a level and across levels of organization and functionality. From this perspective, different researchers in different sub-disciplines in our department contribute different causal, constitutive, and contextual constraints to mechanistic or biophysical explanations of cognitive phenomena."

Professor M pointed out: “Mechanistic integration and unification in cognitive science are more easily asserted than achieved. Have you seen the program of this conference? It includes sessions on ‘memory encoding,’ ‘explanation,’ ‘perception,’ ‘creativity,’ ‘word learning,’ ‘neural dynamics,’ ‘rationality,’ and so on. Within each session, different researchers use different theoretical frameworks, different methodological protocols, and different datasets, to study what they assume to be the same kinds of phenomena or capacities. In fact, our field is often explicitly named in the plural, as the cognitive sciences. What makes it, or should be making it, an integrated, cohesive, science, is the goal to explain cognition on the algorithmic level. At least, that’s the ‘research philosophy’ at our program at the University of Mind City. More generally, we aim to understand the mind at the *personal level*.”

I was starting to get a better idea of the programs at Universities of Brain City and Mind City; but, once again, I had a question: “What’s the personal level?”

Professor B explained: “The distinction between the levels of *sub-personal* and *personal explanations* is a distinction between the explanation of cognitive phenomena in terms of concepts that refer to components of a system—for instance, in terms of brain states or circuits—and in terms of concepts that refer to mental states of a whole system, for instance in terms of a person’s thoughts, motivations, and emotions.”

Professor M added: “Another way to cash out the distinction is in terms of causal-mechanic explanation and reason-based explanation. Sometimes we need explanations that are grounded in people’s beliefs, goals, motives, and values. But we also sometimes need mechanistic explanations that view the human mind as driven, not by reasons, but by sub-personal causal factors.”

“So, if I understand the distinction correctly, we may explain why a person is reading a book on cognition, by saying this person is a student of psychology, has an exam next week, and wants to pass the exam. This would be a type of causal explanation. However, if she has forgotten some parts of the content and thus cannot remember that during the exam, this requires a mechanistic explanation of how human memory works.”

Professor M said: “Well stated. Learning is essential. By the way, we should not forget that behaviorism primarily flourished in the US. European psychology was largely unaffected by behaviorism and always more open to mental concepts.”

I was surprised and said, “Oh, I didn’t know that. Very interesting.”

Professor M added: “Yes, and the link between science and the values of a society are also important when we talk about the levels of explanation. Today, many people in our society think about biological explanations as something unchangeable and hardwired. But this is a fatal, although quite common, misunderstanding.”

Professor B asked: “Are you talking about epigenetics?”

Professor M responded: “Not only. Indeed, epigenetics shows that genes can be switched on or off in particular environmental conditions. Nevertheless, biological approaches are often used to justify injustice and the unfair distribution of goods and chances in society. We should emphasize much more convincingly to the wider public that learning fundamentally shapes our brains, and that environmental and cultural conditions have an enormous impact on the functioning of our brains and genes.”

Professor B nodded and said: “Yes that’s right; and there are good reasons to say that, during the evolution of humankind, biology has been largely replaced by culture as the major driving force in human evolution. The famous example is the selection of lactose tolerance in groups with dairy traditions. And heritable diseases can persist and spread if they occur in families with social power. But this is not the point here. I agree we should think very carefully about the consequences that explanations pitched at a certain level can have for our society.”

Now it suddenly became loud, because one session ended and everyone rushed to the coffee. Professor B said, “Oh, there is my husband, I should leave now.” Professor M said, “Oh, and my husband, too.” They both thanked me for approaching them and wished me all the best with my applications. I had a nice conversation with these two professors, which helped me to get a better sense of the programs at their Universities, and their approaches to explaining the mind. I am sure I will encounter several of the ideas and arguments we discussed in my future studies. I just have to be admitted into a PhD program in cognitive science now. Maybe one that combines the two lines of thinking that I have learned about.

#### **4. Conclusion and final remarks**

Alex now has a better sense of some of the current debates about levels of explanation in cognitive science. And we hope our readers do, too. Of course, this is essential for the methodology and self-understanding of the discipline. But it is also important for another reason. Some researchers (e.g., Núñez et al., 2019) have pointed out that cognitive science is dominated by psychology and marginalizes fields like anthropology, education, and philosophy. This might be right. Yet an even more important question—for both theoretical and practical reasons—is how cognitive science should relate to artificial intelligence (AI) and neuroscience.

Since the early days of cognitive science in the 1950s, AI and neuroscience have been making central contributions to the discipline. Today, however, there is a boom of deep learning and big data approaches in AI. And, in neuroscience, flagship projects such as the “human brain project” aim at mapping the entire human brain in computer models. In the last years, such projects have become relatively independent from cognitive science and moved toward engineering applications, which have received considerable attention in the popular media, but have been criticized for their inability to advance our understanding of the mind (Lake, Ullman, Tenenbaum, & Gershman, 2017; Mainen, Häusser, & Pouget, 2016; Marcus, 2018). Is the departure of these parts of AI and neuroscience from cognitive science something to be concerned about?

Such questions are both practically and theoretically important. It lies at the heart of cognitive science to have a clear standpoint regarding the relationship between biological and artificial intelligence. This is important for scientific progress, of course. But such a standpoint is also important for social, cultural, and political reasons. Today, many people want to understand the relation between biological and artificial intelligence. And the

different answers to this question will certainly have enormous effects on our daily life, our self-understanding as human beings, and the future constitution of society. How we answer such questions also depends on one's views about the right level of explanation for cognitive phenomena and the role of different approaches in fostering our understanding of the mind. Related questions are: Where does cognitive science stand today? Where is it heading? What should its agenda be? Of course, we could not address all these questions in this issue of *topiCS*. But we believe that finding consensus on the right level of explanations for cognitive phenomena is an important step toward satisfactory answers.

## Acknowledgments

We thank James Grayot, Günther Knoblich, and Naftali Weinberger for their feedback on previous versions of this introduction. This work has been partially supported by the German Research Foundation (DFG) within the Priority Program “New Frameworks of Rationality.” MK's work was supported by the DFG under contract number KN 465/9-2. MC gratefully acknowledges financial support from the Alexander von Humboldt Foundation.

## References and recommended readings

- Anderson, P. W. (1972). More is different. *Science*, 177(4047), 393–396.  
A classic on how change in scale can cause fundamental change in the phenomenon under study.
- Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry*, 176, 376–387.  
Alex in the dialogue above refers to this paper on genes and depression.
- Chalmers, D. (2011). A computational foundation for the study of cognition. *The Journal of Cognitive Science*, 1B, 323–357.  
A prominent proposal on how to relate the level of computation with the level of cognitive mechanisms.
- Churchland, P. S., & Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science*, 242(4879), 741–745.  
A discussion of how the relationship between experimental techniques, modelling tools, and cognitive processes taking place at multiple levels of organization in the nervous system.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, 36(3), 181–204.  
A manifesto of a Predictive Processing approach to the mind and how it promises to unify computational and embodied approaches in cognitive science.

Colombo, M. (2017). Experimental philosophy of explanation rising: The case for a plurality of concepts of explanation. *Cognitive Science*, 41(2), 503–517.

An argument in support of the existence of a plurality of concepts of explanation in human psychology, which reflect different philosophical models of explanation.

Craver, C. F. (2015). Levels. In M. Windt (Ed.), *Open MIND*. (26): 8, (pp. 1–26). <https://doi.org/10.15502/9783958570498>

A comprehensive overview of different ways of understanding level in the cognitive and brain sciences.

Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. Cambridge, MA: MIT Press.

A sustained argument that a cognitive architecture grounded in probabilistic graphical models promotes unification in cognitive science.

Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy*, 60(23), 685–700.

An influential essay in the philosophy of action, arguing that giving the reason for which an action was done amounts to giving a causal explanation.

Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy*, 68(4), 87–106.

A statement and defence of the concept of *intentional system*, and of the distinction between the *physical stance*, the *design stance*, and the *intentional stance*.

Dennett, D. C. (1969). *Content and consciousness*. London: Routledge.

Chapter 4 introduces the distinction between personal and subpersonal levels of explanation.

Drayson, Z. (2014). The personal/subpersonal distinction. *Philosophy Compass*, 9(5), 338–346.

A survey of how philosophers and cognitive scientists picked up Dennett's personal/subpersonal distinction.

Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines*, 18(3), 303–329.

An analysis and defence of “levels of abstraction” in artificial intelligence and computer science.

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28(2), 97–115.

One of the most significant expressions of anti-reductionism.

Guttman, B. S. (1976). Is 'levels of organization' a useful biological concept? *BioScience*, 26(2), 112–113.

An argument that the concept of levels of organization in biotic systems “is vague, useless, and misleading.”

Kaplan, D. M. (Ed.). (2017). *Explanation and integration in mind and brain science*. Oxford, UK: Oxford University Press.

A collection of papers from leading philosophers of science and cognitive science on the nature of explanation in mind and brain science, and on the prospects for genuine integration in the cognitive science.

Kellert, S., Longino, H., & Waters, C. K. (2006). Introduction: The pluralist stance. In S. H. Kellert, H. E. Longino, & C. K. Waters (Eds.), *Scientific pluralism, Minnesota studies in philosophy of science*, 19 (pp. vii–xxix). Minneapolis: University of Minnesota Press.

Distinguishes and discusses different brands of scientific pluralism.

Knauff, M., & Spohn, W. (Eds.). (in press). Psychological and Philosophical Frameworks of Rationality – A systematic survey. In: *The Handbook of rationality*. Cambridge, MA: The MIT Press.

Discusses the relation between cognitive science, psychology, and philosophy in our attempt to understand mind and rationality.

Kosslyn, S. M. (1999). If neuroimaging is the answer, what is the question? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 354(1387), 1283–1294.

A paper that discusses foundational and methodological questions about the epistemic status of neuroimaging in cognitive science.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3), 480–490.

A critical comparison between a reductionist and a non-reductionist program for understanding the relationship between brain, mind and behaviour.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Science*, 40, e253.

A discussion of challenges of recent neural network advances, and a proposal for how to combine them with structured cognitive models.

Mainen, Z. F., Häusser, M., & Pouget, A. (2016). A better way to crack the brain. *Nature News*, 539(7628), 159–161.

A critical look at the promise and limitations of “big” neuroscience projects like the Europe an Commission’s Human Brain Project.

Marcus, G. (2018). Deep learning: A critical appraisal. *Arxiv*, 1801, 00631.

A critical appraisal of deep learning

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7(3), 141–144.

A historical account of the cognitive revolution from one the pioneers of cognitive psychology.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

An argument for the need of a set of general principles, which could account for all cognitive phenomena. Uses Soar and ACT-R as examples of unified theories of cognition.

Núñez, R., Allen, M., Gao, R., Rigoli, C. M., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, 3, 782–791.

An argument that so far cognitive science has failed as a multi-disciplinary, integrated science.

Oppenheim, P., & Putnam, H. (1958). The unity of science as a working hypothesis. In G. Maxwell, H. Feigl, & M. Scriven (Eds.), *Concepts, theories, and the mind-body problem* (pp. 3–36). Minneapolis: Minnesota University Press.

An account of science as a hierarchy of disciplines whose ontologies and explanatory power are derivative from the most fundamental physical one.

D. Peebles, & R. P. Cooper (Eds.). (2015). Thirty years after Marr's vision: Levels of analysis in cognitive science. *Topics in Cognitive Science*, 7(2), 185–335.

A volume of *topiCS* dedicated to David Marr's three levels of analysis.

Piccinini, G., & Craver, C. F. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3), 283–311.

A framework for building a unified science of cognition grounded in the notion of a *mechanism*.

Polger, T. W., & Shapiro, L. A. (2016). *The multiple realization book*. New York, NY: Oxford University Press.

A book aimed at clarifying and criticizing, both on empirical and theoretical grounds, the idea that mental states are multiply realized or at least multiply realizable.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106, 467–482.

An influential paper focused on general systems theory and its prospect to explain complexity in terms of the order and hierarchical structures emerging from a dynamical system of interacting elements.

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: The MIT Press.

A polemic book against brain imaging in psychology.

Wimsatt, W. C. (1994). The ontology of complex systems: Levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy*, 24(sup1), 207–274.

An assessment of notions of *level* relevant to the analysis of complex systems.

## Papers in this Topic

Aronowitz, S., & Lombrozo, T. (2020). Experiential explanation. *Topics in Cognitive Science*, 12(4), 1321–1336.

Bender, A. (2020). The role of culture and evolution for human cognition. *Topics in Cognitive Science*, 12(4), 1403–1420.

Bertolero, M. A., & Bassett, D. S. (2020). On the nature of explanations offered by network science: A perspective from and for practicing neuroscientists. *Topics in Cognitive Science*, 12(4), 1272–1293.

Bickle, J. (2020). Laser lights and designer drugs: New techniques for descending levels of mechanisms “in a single bound”? *Topics in Cognitive Science*, 12(4), 1241–1256.

Colombo, M., & Knauff, M. (2020). Editors' introduction: Levels of explanation in cognitive science: From molecules to culture. *Topics in Cognitive Science*, 12(4), 1224–1240.

Gigerenzer, G. (2020). How to explain behavior? *Topics in Cognitive Science*, 12(4), 1363–1381.

Hommel, B. (2020). Pseudo-mechanistic explanations in psychology and cognitive neuroscience. *Topics in Cognitive Science*, 12(4), 1294–1305.



- Potochnik, A., & Sanches de Oliveira, G. (2020). Patterns in cognitive phenomena and pluralism of explanatory styles. *Topics in Cognitive Science*, *12*(4), 1306–1320.
- Rich, P., Blokpoel, M., de Haan, R., & van Rooij, I. (2020). How intractability spans the cognitive and evolutionary levels of explanation. *Topics in Cognitive Science*, *12*(4), 1382–1402.
- Schöner, G. (2020). The dynamics of neural populations capture the laws of the mind. *Topics in Cognitive Science*, *12*(4), 1257–1271.
- Shtulman, A., & Legare, C. H. (2020). Competing explanations of competing explanations: Accounting for conflict between scientific and folk explanations. *Topics in Cognitive Science*, *12*(4), 1337–1362.