

Why go for a computation-based approach to cognitive representation

Dimitri Coelho Mollo

Humboldt-Universität zu Berlin

Exzellenzcluster Science of Intelligence & Berlin School of Mind and Brain
& Institut für Philosophie

This is a pre-print (pre-peer review) of an article forthcoming in *Synthese*

Abstract

The traditional view in (philosophy of) cognitive science is that computation in cognitive systems conceptually depends on representation: to compute is to manipulate representations. I argue that accepting the non-semantic teleomechanistic view of computation lays the ground for a promising alternative strategy, in which computation helps to explain and naturalise representation, rather than the other way around. I show that this computation-based approach to representation presents six decisive advantages over the traditional view. I claim that it can improve the two most influential current theories of representation, teleosemantics and structural representation, by providing them with precious tools to tackle some of their main shortcomings. In addition, the computation-based approach opens up interesting new theoretical paths for the project of naturalising representation, in which teleology plays a role in individuating computations, but not representations.

1 Introduction

A long-lived and fruitful approach to understanding cognition and intelligence is to see cognitive systems as performing computations over representational states. The basic idea is that, at least for many complex tasks,

organisms behave appropriately and intelligently partly because they have internal states — representations — that stand-in for things in the body and world. What representations stand-in for, what they are about, are their contents. Representational states are operated on by cognitive systems by means of computational rules in a way consistent with their contents. If the system is working correctly, the resulting representations —for instance motor commands — are the appropriate ones for the task at hand, as they follow from adequate computational processing of accurate (enough) representations. This allows surrogative reasoning: using a surrogate to think, reason, or process information about something else (Swoyer 1991). Instead of directly manipulating the world, with all the risks that this entails, organisms can first exploit internal surrogates to draw conclusions, integrate information, simulate and evaluate behavioural strategies, etc. — all that before trying something out in a world often intolerant of mistakes.

Its influence and fruitfulness notwithstanding, alternatives to the computational-representational framework have been proposed, ranging from varieties of eliminativism about representation (Churchland 1981, Stich 1983, Hutto & Myin 2013), to varieties of subjectivism and perspectivalism about computation (Searle 1992, Schweizer 2016, Dewhurst 2018*a*), to alternative frameworks for understanding cognition in general (Gibson 1979, Varela et al. 1991, Thompson 2007). This paper will have little to say about the ongoing debate between those in favour of representationalism and computationalism about cognition, and those that oppose these ideas partially or fully. My focus is on issues internal to the representational-computational camp, and especially on the project of accounting for the nature and role of representation, which lies at its philosophical centre. More specifically, I will be concerned exclusively with cognitive representation — the notion of representation most relevant to the cognitive sciences — which is typically ascribed to subpersonal states and processes in cognitive systems¹.

If representation and computation are to be notions that lie at the foundations of the study of cognition and intelligence, it is widely held that we

¹I will often talk of representation *simpliciter* in what follows, even though by this I refer only to cognitive representation.

must offer naturalistic, scientifically-acceptable accounts of both notions. Otherwise the risk remains that the representational-computational research programme in the cognitive sciences is grounded on scientifically problematic notions, causing the whole theoretical edifice eventually to crumble². Consequently, the twin projects of naturalising representation and naturalising computation have occupied philosophers for the past forty years or so. Interestingly, naturalistic theories of representation, with few exceptions (e.g. Cummins 1989, Milkowski 2013, Egan 2014), have given little attention to the ways in which computation may help shed light on representation. Quite on the contrary, the tendency has been to see representation as helping account for computation, a ‘received view’ that the Fodorian slogan ‘no computation without representation’ nicely captures (Sprevak 2010). Part of the reason for the neglect of computation may originate from the fact that alternatives to the received view tend to be very liberal in bestowing computational nature on physical systems — making them of little explanatory use.

I suggest that abandoning the received view of computation in favour of a robust teleology-based view is a promising strategy for theories of representation. The best such theory currently on offer is the teleomechanistic view, which reserves an important role for teleology in accounting for computation (Piccinini 2015, Milkowski 2013, Coelho Mollo 2018). Appealing to teleology in making sense of computation within a theory of representation, I argue, has much to recommend it, insofar as it makes theories of representation more solid, whilst doing justice to the currently widespread suspicion that teleology is, in some way or another, important for representation. I suggest that instead of explaining computation by means of representation, as per the traditional view, we should do exactly the opposite: explaining representation by means of computation — a theoretical path adumbrated by computational mechanists such as Piccinini and Milkowski. Instead of

²Some are less exercised by this worry, holding that the scientific fruitfulness of representation and computation are sufficient to assure us of their scientific respectability (Burge 2010, Rescorla 2013). Although these are indeed good evidential grounds for their respectability, the history of science is rich in examples of similarly explanatorily useful notions that turned out to be misguided. Trying to get additional guarantees is thus a commendable pursuit.

Fodor’s, we should endorse Milkowski’s (2013, p. 166) dictum: ‘there is no representation without computation’.

In this paper, I will lay out in detail some of crucial advantages that a computation-based account offers for shedding light on cognitive representation, especially once we subscribe to the teleomechanistic view of computation. My aim here is not to offer a full-fledged computation-based theory of representation, but rather to outline the shape that such accounts may take. Some accounts are best seen as improvements over extant theories of representation, but the computation-based approach also opens the way to novel strategies for naturalising representation. In this paper, I will remain neutral on the best way to go between these two options: my aim is to show why we should go for a computation-based approach to representation at all, upending the traditional picture.

Here is how I will proceed in what follows. In section §2 I briefly review the motivations and main shortcomings of the currently most influential approaches to representation: teleosemantics, and structural representation. In section §3, I propose that theories of representation should abandon the received view of computation and rather endorse a teleo-based view, such as the teleomechanistic account. In section §4, I argue that placing a robust, teleo-based theory of computation at the basis of theories of representation presents six distinctive advantages, yielding accounts of cognitive representation more capable of confronting some of the weighty hurdles that lie in the path of the naturalisation project.

2 Representation: between teleology and resemblance

Perhaps the main challenge to theories of representation is to identify a set of natural relations between internal states and processes and external entities such that those internal states and processes come to be endowed with fairly determinate and explanatorily useful contents. The quest for determinacy of content is closely related to another central issue for theories of

representation: making space for misrepresentation. A significant notion of misrepresentation seems to be possible only when representations have fairly determinate contents. If a representation represents all sorts of different entities, it becomes unclear when it is correct or adequate, and when it is an instance of representational error, or misrepresentation. A further aim is to show how physical vehicles can play distinctively representational roles, which Ramsey (2007) calls this the ‘job description challenge’. Determinacy of content and misrepresentation have commanded the attention of most of the philosophical literature on cognitive representation.

In this section, I will argue that the three most promising naturalistic theories of representation currently on offer — producer- and consumer-based teleosemantics, and structural representation — illustrate a metatheoretical problem with the traditional approach to naturalising representation. Either traditional naturalistic theories appeal to teleology in fixing content so as to try and secure determinate contents and the possibility of misrepresentation, and in consequence incur in problems with teleology itself and its coarse-grainedness; or they eschew appeal to teleology, but then end up being overly liberal in their bestowal of representational status and contents, jeopardising the explanatory role of the notion and the possibility of misrepresentation.

2.1 Teleological theories of representation

Teleological, or teleo-based theories of representation, appeal to teleology in fixing representational status and/or content. Teleology refers, roughly, to what something is for, to its function, purpose or aim, and it is typically cashed out naturalistically in terms of the grounds for the past selection of a trait or entity by some selection process — natural selection being an important, though non-exclusive such selective process (Garson 2019). What something is for, in this sense, is its teleofunction. Traits of biological organisms are clear examples: hearts have the teleofunction, the purpose, of pumping blood since it is because past instances of the type pumped blood that hearts contributed to the survival and inclusive fitness of organisms having them. This led in turn to hearts persisting across time: they have

been selected for because they pump blood³. Teleology embodies a weak form of normativity: an entity that fails to fulfil its teleofunction — e.g. a heart that fails to pump blood adequately, or at all — malfunctions.

There are two main families of teleo-based theories of representation, both of which remain popular to this day. Producer-based teleosemantics, or teleoinformational semantics, focuses on the information carried by representations, typically determined by the worldly entities that they carry information about, or correlate with (Dretske 1988, Neander 2017). The guiding idea is that the content of a representation are those conditions that the representation has the teleofunction to carry information about (or be correlated with). For instance, a cognitive (sub)system may have the teleofunction to produce representations that carry information about the presence of food, whereby the representations it produces have the content *food present* or some such.

Consumer-based teleosemantics, often referred to simply as teleosemantics, focuses instead on what the teleofunctions of the (sub)systems that use representations are (Millikan 1984, 2017). Roughly, on this picture the content of a representation consists in the conditions that led the system using the representation to fulfil its teleofunction. If what caused the selection of a certain system is that it responded to certain representations in such a way as to bring the containing organism to ingest food, then the condition that led the system to selective success was the presence of food — whereby the representations it consumes have the content *food present* or some such.

Although in some cases the content ascriptions that follow from the two families of theories cohere, they often do not⁴. At any rate, both have *prima*

³In the case of possible future artificial cognitive systems — should real, ‘strong’ artificial intelligence ever be achieved — the kinds of factors relevant for bestowing teleological functions on them will almost certainly be mind-dependent and distinct from the factors underpinning natural teleology, being dependent either directly on intentions, ideas, and purposes of designers and users; or indirectly, in the likely case that such systems will be the result of artificial processes that may or may not mimic the natural processes that led to the evolution of cognitive systems and intelligent organisms.

⁴Famous cases of divergent content ascriptions include (toy versions of) magnetotactic bacteria, so-called bug-detectors in toads, and imaginary cases such as Pietroski’s (1992) *kimus*.

facie the tools to yield determinate contents and make space for misrepresentation. Teleology helps fix the content of representations, discarding those candidate contents that are not grounded in natural teleology. Since it involves a weak form of normativity, the appeal to teleology makes misrepresentation possible. A token representation produced or consumed in ways that are at odds with the teleofunctions of its producers and/or consumers is a misrepresentation, insofar as it has been produced or consumed by systems that fail, in that instance at least, to fulfil their teleofunctions.

Appeal to teleology in theories of representation has been met with important challenges.

An influential objection is the functional indeterminacy challenge. In a nutshell, the objection purports to show that teleology fails to yield sufficiently determinate representational contents, being unable to cut contents finely enough (Fodor 1990). Dedicated neural circuitry in toads generates snapping of their tongues when they detect small moving dark spots in their visual fields. Given that in toads' usual habitat small moving dark spots happen to be edible insects, this circuitry has been plausibly selected by evolution given its adaptive value. At least three different contents, *insect present*, *food present*, and *moving dark spot present*, can lead to selective success, given the environments where frogs live and thrive. In consequence, teleology per se cannot decide between these different content ascriptions, leading to indeterminacy of content.

A more general objection to appeals to teleology in accounting for representation is due to Cummins (1996), Burge (2010). They point out that natural teleology concerns only behavioural success, since it is this kind of success that leads to selection. Behavioural success, they argue, is separate, and sometimes in conflict with, representational success — representational accuracy may often be too costly to be adaptive. Teleology cannot thereby play a role in theories of representation, for what generates appropriate behaviour need not represent the world correctly. Teleological theories run together two types of correctness, behavioural and representational, that are independent of one another, and although they may coincide in some cases, they need not so do.

2.2 Non-teleological theories of representation

I will not assess the merits of such objections. They are partly responsible, however, for motivating theorists to try and come up with theories of representation that do not appeal to teleology. Such theories, with few exceptions, have lost popularity in the past few decades, which have been marked by a dominance of teleo-based views — as well as of a less optimistic attitude towards theories of representation in general (Godfrey-Smith 2006). Important examples of non-teleological theories include causal-informational semantics (Dretske 1981, Fodor 1987, Usher 2001), structural representation (Cummins 1989, Ramsey 2007, Isaac 2012), and content pragmatism (Egan 2014, Coelho Mollo 2020*b*). Content pragmatism rejects the naturalisation project, and thus I will not discuss it further. Causal-informational semantics has been largely abandoned, since it has proved to have serious problems with indeterminacy of content, falling victim to the disjunction problem (Fodor 1984; but see Usher 2001 for an improved version of the view). An additional ground for its decline is the fact that a teleo-based alternative to it, which keeps much of its spirit, is far superior, namely producer-based teleosemantics. For these reasons, I will mostly be concerned with the non-teleological theory of representation that has enjoyed most attention in the past few years: structural representation.

Structural representation is a sophisticated version of resemblance theories of representation. It holds that a large factor in determining representational status and content is that representations and what they represent share relational structure. For instance, the spatial relations between points on a map are similar to the spatial relations between locations in the city mapped. Sharing relational structure, or structural resemblance, is seen as crucial for something to stand in for something else, i.e. to play a representational role. It is because two structures share all or part of their relational structures that one can be used in the place of the other in reasoning or cognitive processing — as I can use a map to plan my future movements when visiting a new city. Content is partly, if not fully, determined by the relevant structural resemblance relation. Several proposals exist as to how to under-

stand structural resemblance precisely, typically having recourse to different kinds of mapping relation, or morphisms, between structures, such as isomorphism and homomorphism. Often the relevant relational structure over the representational vehicles is taken to be computational in nature, yielding thus versions of computational role semantics (Cummins 1989, Ramsey 2007).

The central difficulty for structural representation theories is their liberality. Structural resemblance relations are notoriously easy to come by, and difficult to constrain in principled ways (McLendon 1955, Shea 2013). A problematic consequence of liberality is that representational status is bestowed on an enormous amount of vehicles, watering down the explanatory purchase of appeal to representation. Furthermore, resemblance relations are reflexive and symmetrical, while representation relations are not, suggesting that the former are the wrong tools to understand the latter (Goodman 1976). Finally, liberality also infects content determination: since any representational vehicle stands in structural resemblance relations to a large amount of things in the world, its contents, barring further constraining factors, will be wildly non-unique⁵. Explanation of appropriate behaviour in terms of content, and of inappropriate behaviour in terms of misrepresentation, thus risks losing much of its distinctive value, since representations will have disjunctive contents that include both task-appropriate and task-inappropriate contents. This makes appeal to content largely uninteresting, and moreover picking and choosing the appropriate contents out of the disjunction begs the questions that accounts of content are meant to answer.

Some authors add further requirements to constrain the preferred morphism and avoid liberality, including being generated by a causal process originating from what is represented (Isaac 2012), and being the one relevant for the use that the representation is put to by the system (Ramsey 2007, Shea 2014). Many teleo-based theories also make space for structural representation, adding a teleological constraint to the content-determining structural

⁵Cummins (1996) is a proponent of structural representation who bites this bullet: he claims that content is non-unique, but tries to save the explanatory power of representation by appeal to a teleological theory of representational targets.

resemblance relation (Millikan 2004, Neander 2017, Shea 2018).

It is not my aim to assess whether and to what extent these additions to the basic idea underlying structural representation are successful. I wish only to bring to the fore a metatheoretical problem in existing attempts at naturalising representation. For it seems that we either have to appeal to natural teleology in accounting for representational status and content, thus having trouble with functional indeterminacy and with the conflation of biological and representational success; or, if we reject appeal to teleology, we end up with overly liberal theories of representation, making appeal to the notion explanatorily uninteresting. In what follows, I wish to suggest one way of moving forward, building on insights by Piccinini (2015), Milkowski (2013). Namely, that a promising way to go is to appeal to teleology in individuating computation, in addition to or in alternative to, appealing to teleology in individuating representation and content.

3 Teleology in Computation

According to the semantic view of computation, dominant especially within philosophy of mind and cognitive science, physical computation presupposes representation: computation essentially involves the transformation of states with representational content (Sprevak 2010). Including a semantic constraint in a theory of computation is widely seen as a necessary move in order to curb pancomputationalism, that is, the claim that everything computes one or more functions (limited pancomputationalism), or that everything computes any function (unlimited pancomputationalism). Pancomputationalism, especially of the unlimited sort, jeopardises the explanatory role of computation in the computer and cognitive sciences, insofar as it makes the notion apply overly liberally and in ways that are at odds with scientific practice (Piccinini 2015, chap. 4). Imposing a semantic constraint on computation is a promising strategy to avoid pancomputationalism. Since plausibly relatively few systems are representational, the set of computational systems will at most be identical with, but more likely a subset of, the set of representational systems. Few physical systems will thereby count as

computational under the semantic view.

Appealing to semantics in accounting for computation comes with a price. First, the success of the account depends on the success of an independent theory of representation, and one that cannot rely on computation, on pain of circularity (Piccinini 2015, pp. 33-4)⁶. Second, the semantic view only succeeds in avoiding pancomputationalism and doing justice to scientific practice, two of its central aims, if the theory of representation it relies on is not overly liberal in its turn⁷, and if it individuates as computational (and non-computational) those physical systems that the computer and cognitive sciences take to be such. The latter is particularly troubling for the semantic view, since arguably many systems considered to be computational by the computer sciences do not seem to have a semantics, at least not of the externalist kind most relevant to cognitive representation. Third, the semantic view precludes what can be a promising avenue to explore in making sense of cognitive representation, namely basing it on physical computation⁸.

Until recently, the main alternatives to the received view were non-semantic views based on mappings between abstract computational states and rules, and causal transitions between physical states and processes (Chalmers 2011). To compute, on this view, is to have a causal structure onto which an abstract computational description — i.e. in terms of functions from input types to output types — can be mapped. Causal mapping views admittedly suffer from problems relating to limited pancomputationalism. Since nearly every physical system has internal causal structure, it follows that nearly all physical systems compute at least one function, and typically more than one.

Proponents of the view argue that limited pancomputationalism does not endanger the explanatory power of the notion of computation. I take this strategy to be misguided. First, by connecting computation too closely to

⁶Perhaps more promising are views that see representation and computation as mutually co-defining, such as Ramsey (2007).

⁷Theories that tend toward panrepresentationalism, such as Cummins' (1996), are therefore of no help.

⁸See Piccinini (2015) for further shortcomings of semantic views, and Shagrir (2018) for a recent partial defence.

causal structure, causal mapping views fail to give the notion of computation a distinctive explanatory role over and above that already granted to the notion of causal structure (Milkowski 2013). Moreover, limited pancomputationalism is more pernicious than they take it to be, as it flies in the face of the explanatory practices of the computational and cognitive sciences, which regard computational systems as a rather small subset of existing systems. Finally, since causal structures, as such, do not involve any sort of normativity, causal accounts cannot make space for an objective notion of miscomputation.

In the past years, a more compelling non-semantic view has been proposed: the teleomechanistic view (Milkowski 2013, Piccinini 2015, Coelho Mollo 2018). The theory makes use of the influential neo-mechanistic framework of scientific explanation to try and shed light on the nature of physical computation. Importantly, it reserves a central role for teleology in the individuation of computational systems. In the rest of this paper, I will argue that endorsing the teleomechanistic view of computation leads to many valuable advantages for theories of representation. A brief presentation of the teleomechanistic view will suffice for my purposes.

According to the neo-mechanistic approach to scientific explanation, a mechanism consists of the entities, their activities (i.e. what they do), and their organisation that are responsible for a certain phenomenon (Illari & Williamson 2012). Given a phenomenon to be explained, explanation proceeds by decomposing the system into the components, their organisation, and their causal roles that explain how the phenomenon comes about or is maintained across time. Mechanisms are thereby individuated by the phenomena they explain, and comprise only those components and causal roles that are relevant to the explanation of the phenomenon at hand. Some mechanisms may also have teleofunctions: there may be some function they are supposed to perform due to design or selection processes. They are teleomechanisms. Artefacts, and biological mechanisms such as hearts, are good examples of teleomechanisms.

The teleomechanistic view of physical computation applies the central aspects of mechanistic philosophy to try and shed light on the nature of

computational systems in the world (rather than in mathematical theory). According to the view, computational systems are mechanisms with the teleofunction to compute, where computing is understood as the manipulation of vehicles according to rules sensitive only to some dimensions of variation of the physical vehicles, i.e. their degrees of freedom. A rule, in its turn, is a mapping from input-types (and possibly types of internal states) to (types of internal states and) output-types — where the types are individuated in terms of the structural and/or functional organisation of the mechanism.

The appeal to teleology plays (at least) a triple role in the account. First, it reveals a privileged phenomenon, i.e. a teleological function, that guides the individuation and decomposition of the mechanism into its components and their activities. Second, it helps avoid pancomputationalism of any sort, since plausibly relatively few systems possess teleofunctions of any kind — clear cases being biological traits and artefacts — and even fewer the teleofunction to perform computations — plausibly only designed computers and, if computational theories of cognition are correct, cognitive systems. Third, it makes space for miscomputation, since computational systems can compute functions other than the ones it is their teleofunction to compute, therefore miscomputing⁹. For instance, a system may have the teleofunction to compute a certain function f from inputs to outputs — given selection history or design— but due to a defect or overload, may in some circumstances compute function g instead. Given that the system, in such instances, fails to compute the function it is its teleofunction to compute, it miscomputes.

The mechanistic view has been shown to be descriptively and explanatorily adequate (Milkowski 2013, Piccinini 2015). Importantly, it individuates computation without making recourse to representation. Computations may take place over representational vehicles, but need not: computation does not depend essentially on representation. Being robust, non-trivial, and non-semantic, such a view of computation opens the path to exploring the

⁹They may also malfunction more catastrophically, failing to compute any function at all. Although this is a case of malfunction, it is not a case of miscomputation, since no computation takes place. See Tucker (2018) for a detailed and compelling account of miscomputation within the mechanistic view.

idea that cognitive representation can be partly accounted for in terms of computation, rather than the other way around.

4 Representation and Computational Teleology

The teleomechanistic view of physical computation offers a solid non-semantic basis that theories of cognitive representation — within the computational-representational framework — can make use of. More generally, having a theory of representation rely on a teleological theory of computation presents many advantages. I will here explore six of the most central ones, occasionally putting them in connection with current theories. My contention is that the computation-based approach to representation is promising in at least two ways: it can modify extant theories of representation such that they are better able to tackle their main shortcomings; and it can suggest compelling theories of representation importantly at odds with current approaches. The rest of this paper will be dedicated to showing why this is so.

4.1 Naturalistic credentials

A first motivation in favour of computation-based accounts is that the scientific status of computation seems to be less controversial than that of representation, and the prospects for its direct naturalisation seem rosier. Representation is essentially defined by a property that is not easy to make sense of by recourse to the natural sciences. The property of being about something else, of being directed toward something else — aboutness, intentionality — seems importantly different from the kinds of properties at work in the basic sciences, and particularly difficult to explain in naturalistic terms. While significant progress has been made on the naturalisation of representation, the lack of consensus that persists despite decades of focused philosophical work attests to the difficulty of the project.

Computation, on the other hand, is a notion that came to theoretical maturity as already importantly tied to mechanisation (albeit of a purely abstract sort), in the form of the Turing machine, and later on of concretely

implementable computational architectures. Its relations to properties and processes amenable to mechanical explanation are much more direct if compared to representation. Using computation as an important stepping stone to account for representation therefore suggests itself as a promising theoretical move. For if computation can be more easily naturalised, we can use it as a starting point for a naturalistic theory of representation.

In light of these considerations, it is curious that most philosophical work in the area tended to start with representation, rather than computation. There is much to be said about the reasons for this, but I would like to mention three tentative ones. First, representation was recognised as central to explaining cognition and intelligence very early in the history of philosophy, starting at least with Plato (e.g. in the *Theaetetus*); while recognition of the importance of computation came much later, Hobbes being a precursor of the explosion in interest that took place in the second half of the twentieth century following advances in the theoretical and practical understanding of computational systems. Second, computation is appealed to in helping explain how representations are transformed in task-appropriate ways, with representation's aboutness taken to play the lion's share of explaining successful behaviour. Third, as hinted above, the predominant views of computation until recently were extremely liberal and of little explanatory use if not properly constrained, and representation, being already a central part of theories of cognition and intelligence, seemed the natural choice to curb that pernicious liberality.

The teleomechanistic view of computation allows us to do away with this latter motivation for the received picture. The former two, on their hand, do not justify the received view, but merely shed light on the mostly historical grounds for its apparent intuitiveness. They should therefore be weighted against the considerations I brought to bear above in favour of starting with computation instead, which I take to shift the balance toward the computation-based approach. Be as it may, the foregoing line of reasoning is suggestive, but far from decisive. Stronger reasons, I believe, come from the considerations that follow.

4.2 Causal relevance and explanation

Giving pride of place to the notion of computation in a theory of representation helps to account for the causal powers of representational vehicles, courtesy of a theory of computational implementation. This has been often seen as the main (if not the only) contribution that the notion of computation makes to the computational-representational framework in the cognitive sciences (Fodor 1975, Haugeland 1981): appeal to computation helps explain how physical vehicles can behave in ways that respect semantic and rationality constraints, such as preserving truth in deductive arguments, being responsive to practical reasons, or playing good chess. The physical vehicles must be so regimented as to follow computational rules that mirror to a relevant extent semantic or rationality constraints, thus being able to work as representational vehicles, i.e. the physical states that carry representational content. Representational vehicles are causally efficacious in virtue of their physical properties. Theories of computational implementation build a bridge between the somewhat abstract level of the computational individuation of a system, and the details of the causally efficacious physical processes that realise, or implement, the computations it performs. Computation, in this way, connects semantic, causally inefficacious properties with physical, causally efficacious ones. Thereby — the accepted picture goes — cognition, thought, and intelligence can be physically explained.

This may seem to be an advantage that the teleomechanistic view of computation, in contrast to other non-semantic views, cannot have, since the factors relevant for determining teleological functions are causally inert, being historical rather than occurrent¹⁰. Since the teleomechanistic view builds causally inert factors into its conditions of individuation for computation, one may worry that it cannot provide the desired bridge between representational and causal explanation. In this central respect, therefore, the foregoing view may seem to be inferior to its competitors.

Such a worry, I believe, trades on a failure to distinguish individuation

¹⁰Appealing to non-etiological theories of teleological function, as Piccinini (2015) recommends, does not avoid the problem, given that, as Garson (2019) compellingly argues, such theories make covert appeal to historical factors.

from implementation, a distinction that has played an important role in debates about mental causation. While individuating properties need not be causally efficacious, implementational properties do. Take for instance a 1€ coin. What individuates it as a 1€ coin, beside its physical properties (shape, weight, materials, colour, etc.), is the fact that it has been minted by an authorised European mint. Other objects may share all the occurrent properties of 1€ coins, but fail to be 1€ coins because they lack the appropriate history. Historical properties are causally inefficacious: a 1€ coin and a perfect counterfeit have the same causal powers, since they share the same occurrent properties. But this does not make 1€ coins causally inefficacious, for their causal powers (e.g. interacting appropriately with ticket machines) depend on the occurrent properties of their ‘implementation’, that is, the entities in the world that the individuation conditions pick out.

Computation, under the teleomechanistic view, is similarly a kind that is partly individuated by historical properties. This makes it so that two physically identical systems — with identical causal powers — may differ in computational nature if one has the teleofunction to compute — a historical property — and the other does not. However, individuation conditions, if they successfully apply to entities in the world, pick out physical systems that possess causal powers. Computational individuation picks out those physical systems that implement computations, insofar as they have the appropriate occurrent and historical properties. The implementing systems have causal powers, and a subset of these causal powers must be such that they fulfil the appropriate parts of the individuation conditions of computation, namely the systematic transformation of inputs into outputs. Therefore, the teleomechanistic view, as other theories of computation, does not threaten the role of computation in connecting semantic properties with physical, causally efficacious ones via computational implementation. For the mechanistic view appeals to teleology in the individuation of computational systems and computational functions, and not in order to account for computational implementation, which is purely mechanistic (Coelho Mollo 2018, Tucker 2018)¹¹.

¹¹For a detailed analysis of the relationships between computation and implementation

The fact that the causal powers of computational systems hinge on their occurrent, implementational properties, does not make the appeal to teleology idle. For teleology helps individuate those physical systems that are computational to start with, and thereby that implement physical computations by means of some of their causal properties. Without such a demanding individuation condition, a theory of computation is forced to embrace pan-computationalism — given the ubiquity of causal relations — with all the accompanying problems (see section 4.3 below) ¹².

In some cases, such as in structural representation, the causal powers of representational states may be more directly explained in terms of representations being computational structures that stand in structural mapping relations to representational contents. Here, computational structure captures a central part of what it is for physical states and processes to represent, helping explain how they generate appropriate behaviour by means of working as surrogates to their contents (Swoyer 1991, O’Brien 2015, Gladziejewski & Milkowski 2017). Importantly, a robust mechanistic theory of computation allows us to individuate the causal powers of computational states and processes independently of a theory of representation, thus paving the way, without the threat of liberality or circularity, to basing the causal powers of representations on the causal powers of computations.

4.3 Liberality

A robust, non-liberal theory of physical computation, such as the teleomechanistic view, can help constrain the states and processes in the world that are candidates for representational status. In a theory of representation that takes seriously the computational part of the representational-computational framework, only those states and processes that are part of computational mechanisms can count as representational. This narrowing down of candid-

in the mechanistic view, see Elber-Dorozko & Shagrir (2019). I endorse what they call the ‘separate hierarchies’ view.

¹²Dewhurst (2018*a*) is a good illustration of this problem: by eschewing appeal to teleology, he has to accept pancomputationalism. This leads to problems with the explanatory value of computation, which Dewhurst tries to tackle by appeal to epistemic perspectives. See Coelho Mollo (2020*a*) for criticism.

ates for representational status is something that only views that ascribe computations to an adequately limited set of physical systems can offer. Moreover, the account of computation must be non-semantic, on pain of being circular, or at least uninformative — if we individuate computation by means of representation, appeal to computation is idle in narrowing down the domain of the representational.

The teleomechanistic view is the only non-semantic account of computation that is constrained in this way. Causal mapping accounts, the most sophisticated alternative non-semantic view, lead (at least) to limited pan-computationalism. They fail to narrow down candidates for representational status by means of appeal to computation, since they ascribe computational nature to (almost) all systems. The teleomechanistic view, on the other hand, poses demanding constraints on candidate representational states and processes. Representational states and processes are a subset of computational states and processes, and given that the latter are relatively rare, the claim has considerable bite. Indeed, it excludes all those states and processes of the system that do not contribute to its computational capacities from being candidates for representational status.

In the case of biological cognitive systems, for instance, states and processes that involve glia and blood vessels are not representations, since as far as we know they do not play computational roles — even though they certainly play roles sustaining the states and processes that do play such roles, in partial analogy to what fans and batteries do in electronic computers. In a way true to the spirit of the computational-representational approach to the cognitive sciences, computation helps capture those properties of cognitive systems that are directly relevant to cognition and intelligent behaviour¹³.

In brief, a robust, teleo-based theory of computation helps to delimit the set of physical vehicles that can play representational roles. This is a

¹³There may be cognitive abilities that exploit in part non-computational processes, such as random processes. The foregoing account does not rule this out, but rather focuses on those abilities, arguably the majority, that either in part or fully employ computational processes. The account of computation at play here is moreover general enough to encompass digital, analogical, and possibly other forms of computation, such as neural computation (Piccinini & Bahar 2013).

welcome addition to theories of representation insofar as it helps to avoid the danger of being overly liberal about representational status. This is particularly important for structural representation theories, since the representational relation they rely on, i.e. mapping relations of some sort, is overly unconstrained, risking to lead both to too many things being bestowed representational status, and to radical indeterminacy of content. Endorsement of the teleomechanistic view of computation contributes to avoiding these unwanted results in at least two ways.

First, it constrains considerably the kinds of physical vehicles that can stand in the content-constituting mapping relations that structural representation theories identify. The teleomechanistic view individuates a relatively small subset of physical systems that are computational, and only a subset thereof will acquire representational status: those computational states and processes that stand in the relevant mapping relation to worldly states of affairs. As should be clear, this is only part of the solution to liberality. For it to succeed, it also requires that the overall theory of representation be able to individuate to some precision the worldly states of affairs that are candidates for standing in the relevant mapping relations, that is, to be representational contents. Promising solutions to this issue have been put forward by Ramsey (2007) and Shea (2018), who appeal to organismic embeddedness and behavioural salience to help identify the candidate contents. Together, teleomechanistic computation and additional requirements such as these further constrain the relevant content-fixing mapping relation, attenuating the problems that structural representation theories have with liberality.

Second, narrowing down the candidate vehicles for cognitive representations indirectly contributes to meeting indeterminacy of content challenges. In light of the fact that the theory of computation individuates the internal computational states and processes of the cognitive system in ways independent from representational status, it helps constrain the kinds of content that can be ascribed to such states and processes. Given the computational profile of such states, how they interface with other computational states and sensory and motor surfaces, only some contents are compatible with

the computational workings of the system, many others being implausible insofar as they clash with — or fit rather unnaturally — the computational profile of the vehicles. Admittedly, these constraints by themselves are insufficient to solve content indeterminacy problems, but they help, and given the difficulties involved, any help is welcome.

4.4 Objections to teleo-based representation do not apply

It is possible to do justice to the insight that representation is in some way tied to natural teleology, as per teleo-based views of representation, whilst avoiding having to view the connection as directly tied to content determination. Instead, teleological considerations may come into play exclusively in the individuation of physical computation, generating theories that are teleo-based, insofar as they depend on a teleo-based notion of computation, but that do not use teleology to determine content. This can be helpful for preserving a role for teleology in accounting for representation, while avoiding the main lines of objection against teleo-based views. As I show below, appealing to teleology to individuate computation is not liable to the central kinds of objections that have been raised against a similar appeal in individuating representation.

First, since computations are not individuated in terms of representation, the claim that representational contents are finer-grained than what natural teleology can provide is irrelevant. Second, computations, in contrast to representations, are individuated in a relatively coarse-grained way, so that, even assuming that teleology is a rather blunt tool with which to carve nature, no indeterminacy challenge analogous to the one that confronts representation affects computation. Let me expand on this point.

If computations are considered to be a matter of performance of logical or mathematical functions, it may seem that similar indeterminacy problems appear. It is plausible that natural teleology cannot distinguish between performance of (some) equivalent logical or mathematical functions, since more than one such function maps equally well onto causal goings-on in physical systems. As Shagrir (2001), Sprevak (2010) have shown, this is true of the

basic logical functions performed by logic gates, the most basic computational components: AND-gates and OR-gates, in isolation, are functionally indistinguishable.

Even if this is true, it should not worry proponents of the teleomechanistic view of computation. For according to influential versions of the view, computations are input-output functions implemented by components of a mechanism. Input-output functions are understood extensionally, in terms of transformations of physical quantities (Dewhurst 2018*b*), or in terms of functional states and transitions revealed by the functional decomposition of the system (Coelho Mollo 2018, Fresco & Milkowski forthcoming). Teleofunctions to perform physical or functional input-output functions do not pose indeterminacy worries, since if two such functions are physically/functionally equivalent, they count as the same function, regardless of whether they can be mapped onto different logical or mathematical functions¹⁴. Computations can therefore be individuated partly by teleology without this leading to a multiplicity of individuated computations.

In brief, the indeterminacy problems that follow from appealing to teleology in individuating representations do not transpose to individuating computation by analogous means. In contrast to the fine-grainedness of representational individuation — which must (often) distinguish between co-extensional but distinct contents — computational individuation is coarse-grained enough for teleology to carve the computational domain adequately.

Furthermore, physical and functional states and processes are the kinds of states to which selection processes are sensitive; selection processes being at the core of the most promising family of theories of natural teleology, namely selected-effects theories. There is thereby no problematic gap between computational correctness and behavioural success, as there may be in the case

¹⁴This is an ontological, not an epistemological point. It may be challenging to discover which input-output function a computational mechanism has the teleofunction to perform, given the epistemic opacity of past selection processes, and the complexity of cognitive systems. However, this does not detract from the claim that, if the foregoing view is correct, there is a fact of the matter about what teleofunctions a computational system has. At any rate, recent developments in computational neuroscience provide reasons to be hopeful (Kriegeskorte & Douglas 2018). I thank Rosa Cao for pressing me on this point.

of representational correctness: computational systems have the computational teleofunctions they do because they have been selected for, and they have been selected for because they contribute, often enough, to behavioural success. Moreover, while it is arguable that there can be no biological functions to represent, as Burge (2010) insists — I remain neutral on whether his arguments succeed — it seems extremely implausible to claim that there can be no biological functions to perform input-output functions of specific types, since this is arguably what most biological functions, including non-computational ones, consist in (e.g. taking food as input and generating energy as output)¹⁵. In consequence, the second influential line of objection against teleo-based theories of representation does not apply to teleo-based theories of computation.

4.5 Internal complexity

The non-semantic individuation of the internal computational structure of cognitive systems that the teleomechanistic view offers also helps to address an additional, less discussed difficulty that confronts teleo-based views of representation. As Cao (2012) points out, teleosemantics seems ill-suited to capture the internal complexity involved in bringing about cognitive capacities. The intermediate states between sensory input and motor behaviour are, as cognitive psychology and neuroscience have shown, extremely complex both in terms of the number of contributing internal states, and of their interactions, which are marked by a plurality of dependence relations, feedback, excitatory and inhibitory connections, degeneracy, and redundancy. This makes it so that the causal nexus between initial representation producer — in most cases a sensory surface — and the final representation consumer — typically responsible for generating motor behaviour — is highly mediated.

Teleo-based theories of representation may pursue two main strategies in trying to accommodate this point, none of which is satisfying. On the

¹⁵The medium-independence of computation, however, does pose a challenge. See Coelho Mollo (2019) for an extended treatment of this problem and more generally for a defence of the idea that there are teleofunctions to compute.

one hand, focusing only on the initial representation producer and the final representation consumer is unlikely to vindicate the explanatory purchase of representational explanation in cognitive science, as it fails to make sense of the role played by the intermediate vehicles in bringing about complex and intelligent behaviour. Breaking down cognitive systems into their functional and mechanistic components, and shedding light on their contributions to cognition and intelligence, is the main aim of cognitive science. A theory of representation that is silent about intermediate vehicles is thereby inadequate to the cognitive sciences.

On the other hand, recognising the complex organisation of internal producers and consumers that cognitive science reveals makes appeal to selection processes in determining the content of intermediate states implausible. Most of these states, taken singly, make very limited and unspecific contributions to overall behaviour. In consequence, ascription of content to these states cannot be of the externalist kind relevant to the cognitive sciences. Their representational contributions to any instance of behaviour — the representational functions they have been selected to perform — plausibly do not involve entities external to the organism, but rather the proximal states and processes that capture their direct, specific contributions to cognitive processing. These states would thus have as their contents something along the lines of ‘upstream cognitive subsystem in state x ’ or ‘upstream neuron active’ (Cao 2012). However, such contents are unlikely to justify appeal to representational explanation, insofar as nothing seems to be lost in terms of explanatory power by shedding representational talk, and appealing rather to causal interactions between cognitive subsystems and/or between neural populations.

It may seem that a teleo-based notion of computation would share the same fate. Given that teleofunctions to compute are similarly dependent on selection processes, and since such processes care only about appropriate external behaviour, it seems that internal components cannot possess specific computational teleofunctions for reasons similar to the above¹⁶. Most

¹⁶This damaging line of reasoning is not limited to teleo-based theories of computation that appeal to selected-effects theories of teleology. Goal-based as well as design-based

of the intermediate computational states and processes in cognitive systems make very indirect and variable contributions to external behaviour. Since natural teleology relies on processes sensitive only to appropriate external behaviour, it becomes unclear how such intermediaries, and their computational functions, can be individuated by means of a teleo-based theory of computation.

Mechanistic and functional decomposition can come to the rescue. Although natural teleology, at first glance, can only bestow teleofunctions to compute functions that go from sensory input to motor output, functional and mechanistic decomposition help distribute computational responsibility across intermediate states. Selection processes settle on a complex algorithm or set of algorithms from sensory input to appropriate motor output. In light of the teleofunction to carry out such complex algorithms, there is a breakdown of the contributions that intermediate states and processes make to the overall algorithm. In this way, the theory bestows computational teleofunctions on mediating states and processes in light of the computational contributions they make to the performance of the complex algorithm it is the teleofunction of the whole system to compute. Do these intermediate teleofunctions help identify, in explanatorily powerful ways, the computational contributions that intermediate states and processes are supposed to make to cognition and intelligent behaviour?

In contrast to the case of representational individuation, the answer is positive. For unlike the case of representation, the teleofunctions that intermediate computational states and process have — according to the teleomechanistic view of computation — are to compute simpler input-output functions that together make up the complex algorithm from sensory input to motor output performed by the whole system. The teleomechanistic view bestows on intermediate subsystems the teleofunction to manipulate in specific ways the inputs they receive — individuated non-semantically, in terms of functional or physical states — and generate outputs to be fed to downstream computational subsystems. These bestowals of computational

theories also rely on external behaviour in order to bestow teleofunctions on systems, and thus would also be vulnerable to the problem.

function are explanatorily powerful. They help explain how the system comes to implement the complicated algorithm that leads from sensory input and internal states to behaviour by recognising the small, partial computational contributions that intermediate states and processes internal to the system make, and which together come to compose the overall function that the whole system computes. While in the case of teleo-based theories of representation, as Cao (2012) argues, we end up with content ascriptions that have little to no explanatory purchase, the teleomechanistic view provides computational ascriptions that play a distinctive explanatory role: they reveal the typically many-stepped richly-branched algorithm that leads from sensory stimulus to motor behaviour.

The difficulties of teleo-based views of representation in doing justice to internal cognitive complexity are therefore not shared by the teleomechanistic view of computation. A theory of representation can exploit this feature in order to avoid the problem of internal complexity, in at least two mutually-exclusive ways.

First, by relying on the teleomechanistic view of computation, a theory of representation may make use of the non-semantically individuated intermediate computational states and processes, and the contributions they have the teleofunction to make to overall behaviour, to help distribute representational responsibility across the components of the cognitive system. A theory of representation may be able to pick out a subset of computationally-individuated states and processes that play representational roles by standing in exploitable (and exploited) relations to salient features of the environment — such as cognitive maps in entorhinal cortex (Moser et al. 2008).

Alternatively, theories of representation may reject representational ascription to intermediate states and processes, while keeping allegiance to intermediate computations and computational teleofunctions. On this view, the units on which representations are bestowed are whole organisms, or at least whole brains, as Cao (2012) suggests, since it is at this level of description that talk of external behaviour is justified, and thereby appeal to selection processes takes a direct hold. This Dennettian view is strongly revisionary of mainstream cognitive science, since it rejects subpersonal representational

vehicles, and thereby subpersonal representational explanation. However, by means of the teleomechanistic view of computation, it is sensitive to subpersonal explanation cashed out in purely computational terms, thus following mainstream cognitive science at least in its focus on the complex internal goings-on that lead to cognition and intelligent behaviour.

Since my aim in this paper is to open up paths, rather than to tread them, I remain neutral on which route is the most promising.

4.6 Misrepresentation and miscomputation

Finally, appealing to natural teleology in accounting for computation allows for a sort of weak normativity already when it comes to computational workings. This can be useful for a theory of representation insofar as it makes it possible to distinguish different factors that may lead to inappropriate behaviour. In particular, it opens the way for distinguishing cases in which misrepresentation depends on computational error, from cases in which misrepresentation takes place despite the fact that the system correctly follows its computational norms (i.e. fulfils its computational teleofunctions).

Applying this point to teleosemantics does not entail that separate selection processes need be responsible for the representational and for the computational teleofunctions of the system. It is compatible with the foregoing that the same kinds of selective pressure led to both the representational and computational organisation of the system – the former hinging on the latter. However, it does not follow that there cannot be dissociations, i.e. cases in which there is misrepresentation but no miscomputation, and even, by a fluke, miscomputation without misrepresentation. For instance, if the environment is not the adequate one for the organism, the relations between internal states and world that are relevant for representational success will not obtain, leading to misrepresentation. But this is compatible with the internal computational processes performing the correct computations, since under the teleomechanistic view, the latter are non-semantic.

Moreover, attention to the role of (mis)computation in explaining behaviour can decrease the scope and weight of representational explanation,

thereby helping to avoid overly liberal ascription of representational nature. For explanation of behavioural inadequacy or error may be cashed out in terms of pure miscomputation in cases in which the relevant computational states and processes lack representational status. This is important to keep in check the temptation to overextend representational explanation to cases, capacities, or systems for which the notion makes little to no explanatory contribution — thus risking to make appeal to representation lose its distinctive explanatory power, and with it, its theoretical justification¹⁷.

5 Concluding Remarks

In this paper, I have argued that theories of cognitive representation have much to gain by paying more attention to the roles that computation can play in helping explain representation. More strongly, I have argued that the received view of the conceptual relation between representation and computation should be turned on its head: computation helps explain representation, rather than the other way around. I have shown that this strategy holds much promise if theories of representation rely on a robust, non-semantic, teleo-based theory of computation such as the teleomechanistic account.

I have examined six advantages that theories of representation can reap by adopting this approach: a) starting from computation seems more justified, since computation is arguably simpler or more primitive than representation, and has better naturalisation prospects; b) it helps to account for the causal powers of representations; while the teleomechanistic view more specifically c) helps to narrow down the states and processes that are candidates for representational status, making more tractable problems with liberality and, to a more limited extent, content indeterminacy; d) it individuates non-semantically the complex internal organisation of cognitive systems; and e) it makes space for useful distinctions between different sources of (behavioural) error; all the while f) avoiding the main lines of objection against appeal to teleology in theories of representation. While I have not provided here a de-

¹⁷See Artiga (2016) for a recent defence of radical liberality about representational explanation, and about representational status more generally.

tailed theory of representation along these lines, I hope to have shown that it is a promising way to go for theories of representation in general.

Acknowledgements

I am indebted to Rosa Cao, Manolo Martinez, Nicholas Shea, Marc Artiga, Fernando Martinez Manrique, Margherita Arcangeli, Michael Pauen, two anonymous referees, and audiences at the EuroCogSci conference, and the workshop ‘Mental Representations in a Mechanical World’, both in Bochum, Germany, in 2019, for valuable feedback on earlier versions of this material.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2002/1 “Science of Intelligence” – project number 390523135.

References

- Artiga, M. (2016), ‘Liberal representationalism: A deflationist defense’, *Dialectica* **70**(3), 407–430.
- Burge, T. (2010), *Origins of Objectivity*, Oxford University Press.
- Cao, R. (2012), ‘A teleosemantic approach to information in the brain’, *Biology and Philosophy* **27**, 49–71.
- Chalmers, D. J. (2011), ‘A computational foundation for the study of cognition’, *Journal of Cognitive Science* **12**(4), 323–357.
- Churchland, P. M. (1981), ‘Eliminative materialism and the propositional attitudes’, *Journal of Philosophy* **78**, 67–90.
- Coelho Mollo, D. (2018), ‘Functional individuation, mechanistic implementation: the proper way of seeing the mechanistic view of concrete computation’, *Synthese* **195**, 3477–3497.
- Coelho Mollo, D. (2019), ‘Are there teleological functions to compute?’, *Philosophy of Science* **86**(3), 431–452.
- Coelho Mollo, D. (2020a), ‘Against computational perspectivalism’, *The British Journal for the Philosophy of Science* .
URL: <https://doi.org/10.1093/bjps/axz036>

- Coelho Mollo, D. (2020*b*), ‘Content pragmatism defended’, *Topoi* **39**, 103–113.
- Cummins, R. C. (1989), *Meaning and Mental Representation*, MIT Press.
- Cummins, R. C. (1996), *Representations, Targets, and Attitudes*, MIT Press.
- Dewhurst, J. (2018*a*), ‘Computing mechanisms without proper functions’, *Minds & Machines* **28**(3), 569–588.
- Dewhurst, J. (2018*b*), ‘Individuation without representation’, *British Journal for the Philosophy of Science* **69**(1), 103–116.
- Dretske, F. (1981), *Knowledge and the Flow of Information*, Basil Blackwell.
- Dretske, F. (1988), *Explaining Behavior*, MIT Press.
- Egan, F. (2014), ‘How to think about mental content’, *Philosophical Studies* **170**, 115–135.
- Elber-Dorozko, L. & Shagrir, O. (2019), ‘Integrating computation into the mechanistic hierarchy in the cognitive and neural sciences’, *Synthese* .
- Fodor, J. A. (1975), *The Language of Thought*, Harvard University Press.
- Fodor, J. A. (1984), ‘Semantics, Wisconsin Style’, *Synthese* **59**(3), 231–250.
- Fodor, J. A. (1987), *Psychosemantics*, MIT Press.
- Fodor, J. A. (1990), *A Theory of Content and Other Essays*, MIT Press.
- Fresco, N. & Milkowski, M. (forthcoming), ‘Mechanistic computational individuation without biting the bullet’, *British Journal for the Philosophy of Science* .
- Garson, J. (2019), *What Biological Functions Are and Why They Matter*, Cambridge University Press.
- Gibson, J. J. (1979), *The Ecological Approach to Visual Perception*, Houghton Mifflin.
- Gladziejewski, P. & Milkowski, M. (2017), ‘Structural representations: causally relevant and different from detectors’, *Biology and Philosophy* **32**(3), 337–355.
- Godfrey-Smith, P. (2006), Mental representation, naturalism, and teleosemantics, in D. Papineau & G. Macdonald, eds, ‘Teleosemantics: New Philosophical Essays’, Clarendon Press.
- Goodman, N. (1976), *Languages of Art: An Approach to a Theory of Symbols*, Hackett Publishing.

- Haugeland, J. (1981), Semantic engines: An introduction to mind design, *in* ‘Mind Design’, MIT Press.
- Hutto, D. D. & Myin, E. (2013), *Radicalizing Enactivism: Basic minds without content*, MIT Press.
- Illari, P. M. & Williamson, J. (2012), ‘What is a mechanism? Thinking about mechanisms across the sciences’, *European Journal for the Philosophy of Science* **2**, 119–135.
- Isaac, A. M. C. (2012), ‘Objective similarity and mental representation’, *Australasian Journal of Philosophy* **91**(4), 683–704.
- Kriegeskorte, N. & Douglas, P. K. (2018), ‘Cognitive computational neuroscience’, *Nature Neuroscience* **21**, 1148–1160.
- McLendon, H. J. (1955), ‘Uses of similarity of structure in contemporary philosophy’, *Mind* **64**(253), 79–95.
- Milkowski, M. (2013), *Explaining the Computational Mind*, MIT Press.
- Millikan, R. G. (1984), *Language, Thought, and Other Biological Categories: New Foundations for Realism*, MIT Press.
- Millikan, R. G. (2004), *Varieties of Meaning: the 2002 Jean Nicod Lectures*, MIT Press.
- Millikan, R. G. (2017), *Beyond Concepts: Unicepts, Language, and Natural Information*, Oxford University Press.
- Moser, E. I., Kropff, E. & Moser, M.-B. (2008), ‘Place cells, grid cells, and the brain’s spatial representation system’, *Annual Review of Neuroscience* **31**, 69–89.
- Neander, K. (2017), *A Mark of the Mental: in defense of informational semantics*, MIT Press.
- O’Brien, G. (2015), How does mind matter? - solving the content causation problem, *in* T. Metzinger & J. M. Windt, eds, ‘Open MIND’, MIND Group.
- Piccinini, G. (2015), *Physical Computation: a Mechanistic Account*, Oxford University Press.
- Piccinini, G. & Bahar, S. (2013), ‘Neural computation and the computational theory of cognition’, *Cognitive Science* **34**, 453–488.
- Pietroski, P. M. (1992), ‘Intentionality and teleological error’, *Pacific Philo-*

sophical Quarterly **73**, 267–282.

Ramsey, W. M. (2007), *Representation Reconsidered*, Cambridge University Press.

Rescorla, M. (2013), ‘Against structuralist theories of computational implementation’, *British Journal for the Philosophy of Science* **64**, 681–707.

Schweizer, P. (2016), In what sense does the brain compute?, in V. C. Müller, ed., ‘Computing and Philosophy’, Vol. 375 of *Synthese Library*, Springer, pp. 63–79.

Searle, J. R. (1992), *The Rediscovery of the Mind*, MIT Press.

Shagrir, O. (2001), ‘Content, computation and externalism’, *Mind* **438**, 369–400.

Shagrir, O. (2018), ‘In defence of the semantic view of computation’, *Synthese* **197**(9), 4083–4108.

URL: <https://doi.org/10.1007/s11229-018-01921-z>

Shea, N. (2013), Millikan’s isomorphism requirement, in J. Kingsbury, D. Ryder & K. Williford., eds, ‘Millikan and her Critics’, Wiley-Blackwell.

Shea, N. (2014), ‘Exploited isomorphism and structural representation’, *Proceedings of the Aristotelian Society* **CXIV**, Part 2, 123–144.

Shea, N. (2018), *Representation in Cognitive Science*, Oxford University Press.

Sprevak, M. (2010), ‘Computation, individuation, and the received view on representation’, *Studies in History and Philosophy of Science* **41**, 260–270.

Stich, S. P. (1983), *From Folk Psychology to Cognitive Science: The Case Against Belief*, The MIT Press.

Swoyer, C. (1991), ‘Structural representation and surrogative reasoning’, *Synthese* **87**, 449–508.

Thompson, E. (2007), *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, Harvard University Press.

Tucker, C. (2018), ‘How to explain miscomputation’, *Philosophers’ Imprint* **18**(24), 1–17.

Usher, M. (2001), ‘A statistical referential theory of content: Using information theory to account for misrepresentations’, *Mind & Language* **16**(3), 311–34.

Varela, F., Thompson, E. & Rosch, E. (1991), *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press.