To appear in *What are Mental Representations?*, Joulia Smortchkova, Krzysztof Dolega, Tobias Schlicht (eds.), New York: Oxford University Press.

A Deflationary Account of Mental Representation

Frances Egan

Among the cognitive capacities of evolved creatures is the capacity to represent. Theories in cognitive neuroscience typically explain our manifest representational capacities by positing *internal* representations, but there is little agreement about how these representations function, especially with the relatively recent proliferation of connectionist, dynamical, embodied, enactive, and Bayesian approaches to cognition. In this paper I sketch an account of the nature and function of representation in cognitive neuroscience that couples a realist construal of representational vehicles with a pragmatic account of representational content. I call the resulting package a *deflationary* account of mental representation and I argue that it avoids the problems that afflict competing accounts.

1. Preliminaries

A commitment to representation presupposes a distinction between representational *vehicle* and representational *content*. The vehicle is a physically realized state or structure that carries or bears content. Insofar as a representation is causally involved in a cognitive process, it is in virtue of the representational vehicle. A state or structure has content just in case it represents things to be a certain way; it has a 'satisfaction condition' – the condition under which it represents accurately.

We can sharpen the distinction by reference to a simple example. See figure [1]. Most generally, a physical system computes the addition function just in case there exists a mapping from physical state types to numbers, such that physical state types related by

a causal state transition relation are mapped to numbers n, m, and n+m related as addends and sums. But a perspicuous rendering of a computational model of an adder depicts two

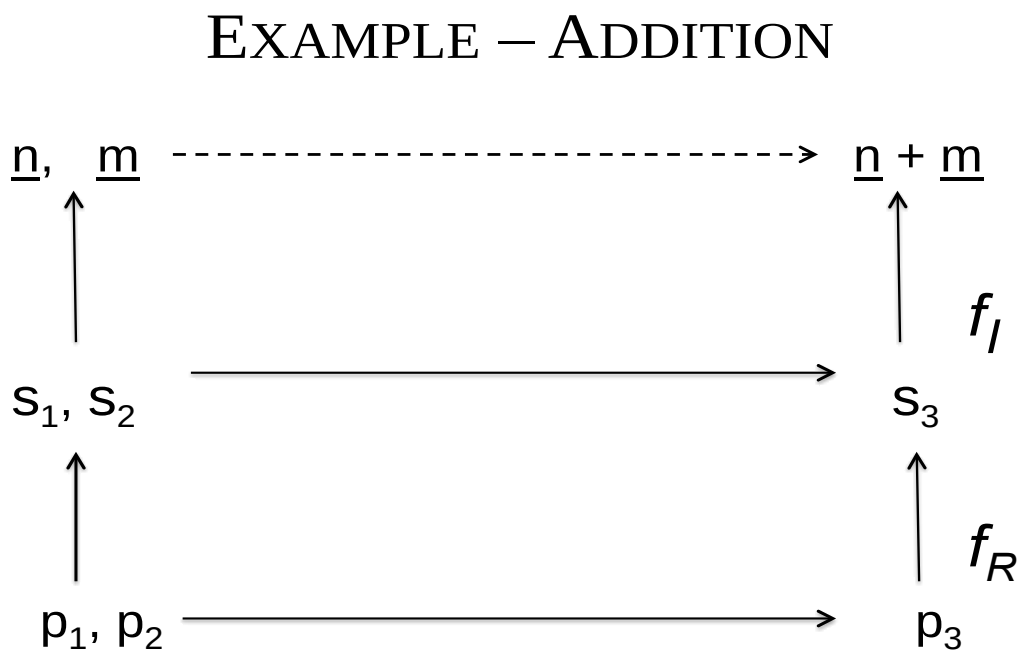

Figure 1

mappings: at the bottom, a *realization function* (f_R) that specifies the physically realized vehicles of representation – here, *numerals*, but more generally structures or states of some sort – and, at the top, an *interpretation function* (f_I) that specifies their content. The bottom two horizontal arrows depict causal relations (the middle at a higher level of abstraction); the top arrow depicts the arguments and values of the computed function. When the system is in the physical states that under the mapping represent the numbers *n*, *m* (for example, 2 and 3) it is caused to go into the physical state that under the mapping represents their sum (i.e. 5).

For any representational construal of a cognitive system we can ask two

questions: (1) How do the posited internal representations get their meanings? This, in effect, is the *problem of intentionality*; and (2) What is it for an internal state or structure to function as a representation, in particular, to serve as a representational vehicle? The appeal to representations should not be idle – positing representations should do some genuine explanatory work. In our terms, what is at stake with (1) and (2) is justifying the interpretation and the realization functions respectively. I shall discuss each in turn below.

First, however, it is useful to set out Ramsey's (2007) adequacy conditions on a theory of mental representation. This will provide a framework for evaluating the account to be defended here. Ramsey identifies at least five general constraints:

1. Mental representations should serve a function sufficiently like paradigm cases of representation. Public language is probably the clearest case; maps are another exemplar.
2. The content of mental representations should be causally relevant to their role in cognitive processes.
3. The account should not imply pan-representationalism: lots of clearly non-representational things should not count as representations.
4. The account should not *under-explain* representational capacities; it should not, for example, presuppose such intentional capacities as *understanding*.
5. Neither should it *over-explain* representational capacities, such that representation is *explained away*. For example, according to Ramsey, if representations function as 'mere causal relays' then, in effect, the phenomenon of interest has disappeared.

With Ramsey's adequacy conditions for a theory of mental representation on the table, let's turn to our first problem: the problem of *mental content*.

2. Representational content: the naturalistic proposals

We can identify several widely accepted constraints on an account of content for cognitive neuroscience:

(1) The account should provide the basis for the attribution of *determinate* contents to the posited states or structures.

(2) The account should allow for the possibility that the posited states can *mis*represent. The motivating idea is that genuinely representational states represent *robustly*, in the way that paradigmatic mental states such as beliefs represent; they allow for the possibility of *getting it wrong*.

There is a constitutive connection between constraints (1) and (2). If the theory cannot underwrite the attribution of determinate satisfaction conditions to a mental state (type), then it cannot support the claim that some possible tokenings of the state occur when the conditions are not satisfied, and hence would misrepresent.

(3) The account should be *naturalistic*. Typically, this constraint is construed as requiring a specification, in non-semantic and non-intentional terms, of (at least) a sufficient condition for a state or structure to have a particular content. Such a specification would guarantee that the theory makes no illicit appeal to the very phenomenon – meaning – that it is supposed to explain. This idea motivates so-called *tracking* theories, discussed below. More generally, the constraint is motivated by the conviction that intentionality is not *fundamental*:

> It's hard to see… how one can be a realist about intentionality without also being, to some extent or other, a reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe supervenience on) properties that are themselves *neither* intentional *nor* semantic. If aboutness is real, it must be something else. (Fodor 1987, 97)

> There are no "ultimately semantic" facts or properties, i.e. no semantic facts or properties over and above the facts and properties of physics, chemistry, biology, neurophysiology, and those parts of psychology, sociology, and anthropology that can be expressed independently of semantic concepts. (Field 1975, 386)

Finally, (4) the account should conform to the actual practice of content attribution in cognitive neuroscience. It should be empirically accurate.

Explicitly naturalistic theories explicate content in terms of a privileged relation between the tokening of an internal state and the object or property the state represents. Thus the state is said to 'track' (in some specified sense) the external condition that serves as its satisfaction condition. To satisfy the naturalistic constraint both the relation and the relata must be specified in non-intentional and non-semantic terms. Various theories offer different accounts of the content-determining relation. I will discuss very briefly the most popular proposals, focusing on their failure, so far, to underwrite the attribution of determinate contents to internal states. I will then illustrate how a pragmatic account of content of the sort I defend handles this thorny issue.

Information-theoretic accounts hold, very roughly, that an internal state S means *cat* if and only if S is caused by the presence of a cat, and certain further conditions

obtain.[1] Further conditions are required to allow for the possibility of misrepresentation, that is, for the possibility that some S-tokenings are *not* caused by cats but, say, by large rats on a dark night, and hence misrepresent a large rat as a cat. A notable problem for information-theoretic theories is the consequence that everything in the causal chain from the presence of a cat in the distal environment to the internal tokening of S, including cat-like patterns in the retinal image, appears to satisfy the condition, and so would fall into S's extension. Thus, information-theoretic theories typically founder on constraint (1), failing to underwrite determinate contents for mental states, and hence have trouble specifying conditions under which tokenings of the state would *mis*represent (condition (2)). The outstanding problem for such theories is to provide for determinacy without illicit appeal to intentional or semantic notions.

Teleological theories hold that internal state S means *cat* if and only if S has the natural function of indicating cats. The view was first developed and defended by Millikan (1984), and there are now many interesting variations on the central idea.[2] Teleosemanticists have notoriously been unable to agree on the natural function of states of even the simplest organisms.[3] Let's focus on a widely-discussed case. Does the inner state responsible for engaging a frog's tongue-lashing behavior have the function of indicating (and hence representing) *fly*, *frog food*, or *small dark moving thing*? Teleosemanticists, at various times, have proposed all three. We might settle on *fly*, but

[1] See Dretske 1981 and Fodor 1990 for the most developed information-theoretic accounts. Further conditions include the requirement that during a privileged learning period only cats cause S-tokenings (Dretske 1981) or that non-cat caused S-tokenings depend asymmetrically on cat-caused S-tokenings (Fodor 1990).
[2] See Matthen 1988, Papineau 1993, Dretske 1995, Ryder 2004, Neander 2006, 2017, and Shea 2007 for other versions of teleosemantics.
[3] See the discussion of the magnetosome in Dretske 1986 and Millikan 1989.

then Quinean indeterminacy[4] rears its head: a *fly stage* detector or an *undetached fly part* detector would serve the purpose of getting nutrients into the frog's stomach equally well. The problem is that indeterminate functions cannot ground determinate contents. Each of various function-candidates specifies a different satisfaction condition; unless a compelling case can be made for one function-candidate over the others, teleosemantics runs afoul of constraint (1). Moreover, the argument must not appeal to intentional or normative considerations (such as what makes for a good explanation), on pain of violating the naturalistic constraint.

A third type of tracking theory appeals to the type of relation that holds between a map and the domain it represents, that is, *structural similarity* or isomorphism.[5] Cummins 1989, Ramsey 2007, and Shagrir 2012 have proposed variations on this idea. Of course, since similarity is a symmetric relation but the representation relation is not, any account that attempts to ground representational content in similarity will need supplementation by appeal to something like *use*. Of more concern in the present context, a given set of internal states or structures is likely to be structurally similar to any number of external conditions. The question is whether structural similarity can be sufficiently constrained to underwrite determinate contents while still respecting the naturalistic constraint.

The upshot of this short discussion is that tracking theories of mental content face formidable problems in underwriting content determinacy, and hence the possibility of misrepresentation, in way that satisfies the naturalistic constraint. One might simply conclude that more work needs to be done, that naturalistic semantic theorists should

---

[4] See Quine 1960.

[5] Better, *homomorphism*, or what O'Brien & Opie 2004 call a 'second-order resemblance' (p.11).

continue to look for naturalistic conditions that would further constrain content. However, if the proposed meaning-determining relation becomes too baroque it will fail to be explanatory, leaving us wondering *why* that particular relation determines content.

Despite the fact that there is no widely accepted naturalistic foundation for representational content, computational theorists persist in employing representational language in articulating their models. It is unlikely that they have discovered a naturalistic meaning-determining relation that has so far eluded philosophers. Shea (2013, 499) claims that cognitive science takes semantic properties for granted, "offer[ing] no settled view about what makes it the case that the representations relied on have the contents they do. The content question has been largely left to philosophy…" There is something right about this idea, which I will say more about in the next section, but on its face it would be a bitter pill for the majority of philosophers of mind who look to the cognitive sciences, and in particular, to computational neuroscience, to provide a naturalistic explanation of our representational capacities. Their hopes would be dashed if cognitive science just kicks the project of naturalizing the mind back to philosophy.

The apparent mismatch between the theories of content developed by philosophers pursuing the naturalistic semantics project and the actual practice of computational theorists in attributing content in their models cries out for explanation; it motivates a different sort of account.

3. Representational content: a pragmatic alternative[6]

[6] See Egan (2014) for elaboration and defense of the view sketched here. See also Mollo (2017).

The view that I favor builds on the central insight of tracking theories – states of mind represent aspects of the world by tracking, in some sense, the distal objects and properties that they are about – but it doesn't suppose that a naturalistically specifiable relation is sufficient to determine a mental state's satisfaction condition. Additional, pragmatic, considerations play an essential role.

A content assignment requires empirical justification, and this requires a certain *fit* between the mechanism and the world. A content assignment that interprets states of a system as representing Dow-Jones stock index prices would be justified only if the states track the vagaries of the market, and to do that (barring a miracle) there must be a causal connection between the states of the system and market prices. The fit between biological systems and distal objects and properties is, of course, is a product of natural selection, but it doesn't follow, as teleosemanticists seem to assume, that evolutionary function – the historical relation that holds between a structure's tokening and its normal cause in the EEA – best serves the cognitive (scientific) theorist's explanatory goals. It may not, for example, if the goal is to explain how a cognitive mechanism works in the here and now. The various tracking relations privileged by naturalistic semantic theories characterize different ways that states of mind can fit the world, with the choice among tracking relations determined by explanatory, or broadly pragmatic, considerations.

Let me elaborate. In ascribing representational contents the cognitive theorist may look for a distal causal antecedent of an internal structure's tokening, or a homomorphism between distal and internal elements, but the search is constrained primarily by the cognitive capacity that the theory is developed to explain. For example, vision theorists will look to properties that can structure the light in appropriate ways; thus they construe

the states and structures they posit as representing light intensity values, changes in light intensity, and further downstream, changes in depth and surface orientation. Theorists of motor control construe the structures they posit as representing positions of objects in nearby space and changes in body joint angles. And the assignment of task-specific content – what I call *cognitive content* – is justified only if the theorist can explain how the posited structures are used by the system in ways that subserve the cognitive capacity in question.

We can see the extent to which pragmatic considerations figure in the ascription of content by revisiting some of the problems encountered by tracking theories in their attempt to specify a naturalistic content-determining relation. Far from adhering to the strict program imposed by the naturalistic constraint, as understood by tracking theorists, the computational theorist, in assigning content to posited internal structures, *selects* from all the information in the signal just what is relevant for the cognitive capacity to be explained and specifies it in a way that is salient for explanatory purposes. Typically, pragmatic considerations will privilege a distal cause (the cat) over a proximal cause (cat-like patterns in the retinal image), because a distal content ascription will facilitate an explanation of the interaction between the organism and its environment necessary for the organism's success. Recall the dispute among teleo-semanticists about whether the frog's internal state represents *fly* or *frog food* or *small dark moving thing*. The dispute is unlikely to be settled without reference to specific explanatory concerns. If the goal of the theoretical project is to explain the frog's role in its environmental niche, then the theorist is likely to assign the content *fly*. Alternatively, if the goal is to explain how the frog's visual mechanisms work, then *small dark moving thing* might be preferred. In other

words, explanatory focus resolves indeterminacy. Turning to Quinean indeterminacy, theories are articulated in public language and the ontology implicit in public language privileges *fly* over *fly stage*. These content choices are not motivated by naturalistic considerations – the naturalistic constraint prohibits appeal to specific explanatory interests or to public meaning. Attention to actual practice reveals that pragmatic considerations motivate the choice among naturalistic alternatives and secure content determinacy.

Cognitive content is not part of the essential characterization of a computational mechanism and is not fruitfully regarded as part of what I call the *computational theory proper*. The theory proper comprises a specification of the function (in the mathematical sense) computed by the mechanism,[7] specification of the algorithms, structures, and processes involved in the computation, as well as what I call the *ecological component* of the theory – typically facts about robust co-variations between tokenings of internal states and distal property instantiations under normal environmental conditions, which constrain, but do not fully determine, the attribution of cognitive content, as explained above. The computational theory proper is, strictly speaking, sufficient to explain the system's success (and occasional failure) at the cognitive task (seeing what is where in

[7] See Egan (2017) for elaboration and defense of what I call *function-theoretic* (FT) characterization, which is an environment-neutral, cognitive domain-general characterization of a mechanism. The inputs of a computationally characterized mechanism represent the arguments and the outputs the values of the mathematical function that canonically specifies the task executed by the mechanism: for example, smoothing functions for perceptual mechanisms (see Marr 1982, among many others), path integration for navigation mechanisms (see Gallistel 1990), vector subtraction for reaching and pointing (Shadmehr and Wise 2005). Hence, the FT characterization specifies a kind of content – *mathematical* content – that is distinct from the (cognitive) domain-specific content that philosophers typically have in mind when they talk about 'representational content' and which I call 'cognitive content'.

the scene, object manipulation, and so on) that is the explanatory target of the theory.

Cognitive content is not in the theory proper; rather it is best construed as a kind of *gloss* – an *intentional* gloss – on the computational theory. It is ascribed to facilitate the explanation of the relevant cognitive capacity. The primary function of an intentional gloss is to illustrate, in a perspicuous and concise way, how the computational theory addresses the intentionally-characterized phenomena with which the theorist began and which it is the job of the theory to explain. Cognitive content is 'connective tissue' linking the sub-personal (primarily mathematical[8]) capacities posited in the theory and the manifest personal-level capacity that is the theory's explanatory target (vision, grasping an object in view, and so on). But, as I noted above, the computational theory proper can fully explain the interaction between organism and environment, and hence the organism's success, without adverting to cognitive content. The intentional gloss characterizes the interaction between the organism and its environment that enables the cognitive capacity in terms of the former *representing* elements of the latter; the theory does not.

An important heuristic function served by the assignment of representational content is to help us keep track of the flow of information in the system, or, to be more explicit, help *us* – theorists and students of cognitive neuroscience – keep track of changes in the system caused by both environmental events and internal processes, with an eye on the cognitive capacity (e.g. seeing what is where) that is the explanatory target of the theory. The choice of content will be responsive to such considerations as *ease of explanation*, and so may involve considerable idealization.

[8] See footnote 7.

An additional function of content ascription is worth noting here; it will play a role in my argument later. A content ascription can serve as a temporary placeholder for an incompletely developed computational theory of a cognitive capacity and so guide the discovery of mechanisms underlying the capacity. For example, at the early stages of theory development, prior to the specification of the mathematical function computed and the structures and processes that enable the computation, a visual theorist may characterize a to-be-specified structure as representing edges or some other visible property of the distal scene. She may even call the structure an EDGE (as Marr does), foreshadowing the functional role that the structure will play in the processes to be described by the theory. Or a capacity may be characterized initially in intentional terms, as, say, *shape from shading*, prior to the development of the computational theory that explains the capacity. At this stage there may be little or no theory to gloss; nonetheless the intentional characterization plays an important role in the search for the mechanisms and processes underlying the intentionally described capacity.

Let me return to Shea's (2013) claim that cognitive science takes semantic properties for granted, leaving the project of specifying the conditions for content attribution to philosophy. On the account I have sketched, there is a clear sense in which computational theorists do take meanings for granted: they don't attempt to reduce mental content, nor do they assume that some naturalistically kosher relation grounds content attribution. Rather, they *use* unreduced, pragmatically motivated, content to explicate (gloss) their theories, and to serve the various explanatory functions described above. In doing so they help themselves to the ontology implicit in public language. But, pace Shea, I am sure they would be surprised to hear that the naturalistic *bona fides* of

their theories depend upon philosophers finding the holy grail of a naturalistic content-determining relation.

I shall conclude the discussion of representational content by returning to the constraints on an adequate account of content for cognitive neuroscience discussed above. In the first place, the account should provide the basis for *determinate* contents. The pragmatic account does this by explicitly recognizing the role of explanatory interests and other pragmatic considerations in determining content ascription. Secondly, the account should allow for the possibility of *misrepresentation*. Once determinacy is secured, we can see how misrepresentation can arise on the pragmatic account. Assume that the interpretation function (f_I), justified in part by reference to pragmatic considerations, assigns the determinate content *fly* to a posited internal state. If the system goes into that state in the absence of a fly, then it misrepresents some other condition as a fly.

The third constraint requires that the account be naturalistic. At first blush, it may seem that the appeal to explanatory and other pragmatic considerations in the determination of representational content would compromise the naturalistic credentials of cognitive neuroscience. That isn't so, because the pragmatic elements and the contents they determine are 'quarantined' in the intentional gloss, to use Mark Sprevak's (2013) apt description of my view. The theory proper does not traffic in ordinary (i.e. cognitive task-specific) representational contents, so its naturalistic credentials are not threatened.

I want to consider the empirical adequacy of the deflationary account of representation as a whole, so I shall postpone discussion of the final constraint until later.

4. Representational vehicles

Turning now to our second question: what is it for an internal state or structure to function as a representation, that is, to serve as a representational vehicle?

Many of our intuitions about representation are shaped by thinking about public language, which is the model for the most popular account of mental representation. According to the *language of thought hypothesis* (LOT) mental representations are literally symbols in an internal language (aka *mentalese*), and mental processes are to be understood as operations on internal sentences.[9] Like more familiar linguistic systems, LOT has a compositional syntax (specified by a realization function f_R) and semantics (specified by an interpretation function f_I). The content of LOT representations is said to be *explicitly represented*, as opposed to represented implicitly in the architecture of the system.[10] But the analogy with public language can be misleading. While the information encoded in printed text is (in some sense) explicit, it must be *usable*. Think, for example, of an encyclopedia without an index or a library without a catalogue. In addition to inert data structures there must be processes that read them. And the process that "reads" mental representations can't involve *understanding*, on pain of *under-explaining* our representational capacities, as Ramsey might put it. As Fodor (1980) noted with his *formality condition*, computational processes are sensitive only to *formal* (that is, *non-semantic*) properties of representations. The relevant properties of the symbols to which computational processes are sensitive will be specified by the realization function f_R.

A wide variety of cognitive models do not posit explicit representations, in the above sense. To mention just a few: (i) *connectionist* models typically explain cognitive

[9] Jerry Fodor is LOT's most ardent champion. See, especially, Fodor 1975 and 2008.
[10] See Kirsh 1990 for a useful discussion of the notion of explicit representation.

phenomena as the propagation of activation among units in highly connected networks; (ii*) dynamical* models characterize cognitive processes by a set of differential equations describing the behavior of the system over time; (iii) *enactive* models treat cognition as consisting, fundamentally, of a dynamic interaction between the subject and the environment, rather than a static representation of that environment. None of these models characterize cognitive processes as involving computational operations defined on symbol structures. A relatively recent development in Bayesian modeling, *predictive processing* models, treat the brain as a predictive machine that uses perception and action to minimize prediction error; it is not obvious that predictive processing models lend themselves naturally to a representational construal in the sense presumed by LOT. The proliferation of various types of cognitive modeling compels us to re-examine our intuitions about when and how information is encoded in a system. At very least, the linguistic model underlying LOT seems overly restrictive.

In general, intuitions differ on the representational status of the various types of models. Clark (1997), Bechtel (1998, 2001), and others argue for a representational construal of connectionist and dynamical models. Chemero (2009), Gallagher (2008), and Ramsey (2007) argue that they do not posit representations. According to Ramsey the structures posited in connectionist models are "mere causal relays." If they count as representations, he cautions, then *pan representationalism* threatens.

A locus of dispute has been the Watt governor [Figure 2], first introduced into the discussion by Van Gelder (1995).

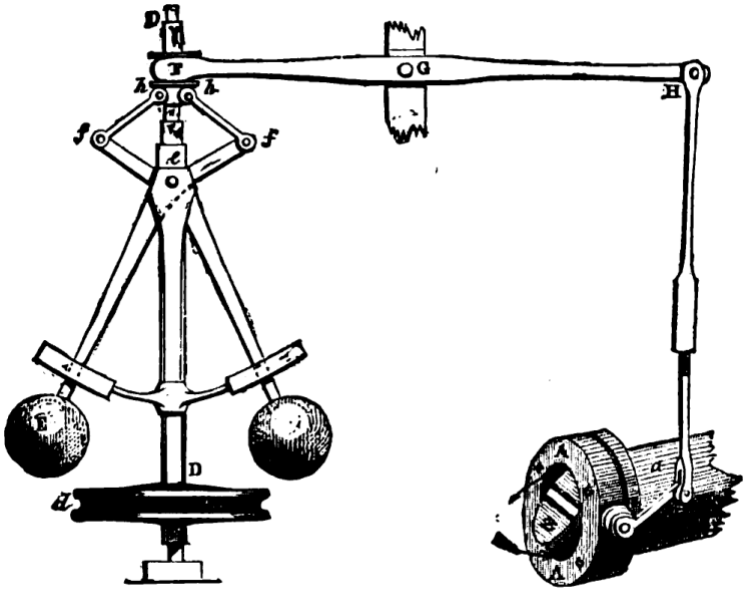

FIG. 4.—*Governor and Throttle-Valve.*

Figure 2

As the speed of the engine increases, centrifugal force elevates the arms of the flywheel, closing off a valve and restricting the flow of steam, thereby decreasing the engine speed. The issue is whether the angle of the arms represents the speed of the flywheel. Bechtel (1998, 2001) and Chemero (2000) think that a representational construal is appropriate; Ramsey (2007) and Shapiro (2010) think it is not. Another hotly disputed case is the toy car [Figure 2] described by Ramsey (2007, p.199).

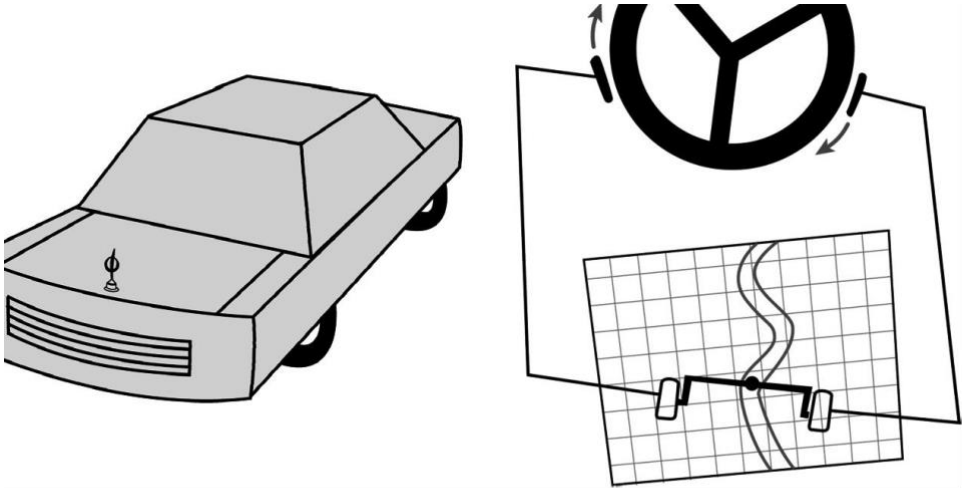

Figure 3

The car negotiates a tricky S-curve tunnel by making use of a groove-and-rudder system that guides the wheels of the car smoothly through the curve. According to Ramsey the system is representational because it uses a structure that is isomorphic to the curved tunnel. But the representational construal of the system is open to dispute. Whatever representational capacity the car has doesn't generalize – it can't negotiate other tracks. And Tonneau (2011) argues that, by Ramsey's measure, a key represents a lock.

We can identify at least three general motivations for resisting a representational construal of a cognitive model: (1) a too narrow, language-based construal of representation, in other words, the intuition that *only* models that posit interpreted symbol structures with a compositional syntax count as representational. It should be noted, however, that not all *public* representation involves such symbol structures – maps, for example, do not – so the intuition that internal representations must be quasi-linguistic is dubious; (2) the idea, popular among proponents of embodied and enactive approaches,

that representation is not necessary for cognition;[11] and (3) the worry, often expressed by proponents of enactivism, that a naturalistic account of representational content is simply not in the cards, and so invoking representations in a scientific account of cognition is indefensible. Hutto and Myin (2013) claim that representation-based theories "… are unable to account for the origins of content in the world if they are forced to use nothing but the standard naturalist resources of informational covariance, even if these are augmented by devices that have the biological function of responding to such information." (2013, xv), dubbing this the *hard problem of content*. They identify three options for the theorist of cognition: (i) give up content, and hence mental representation; (ii) hope that content can be naturalized in some other way; or (iii) posit content as an irreducible, explanatory primitive, in other words, embrace a kind of dualism. Enactivists propose (i), eschewing content and hence mental representation. But, as I have argued above, there is a fourth option for dealing with the 'hard problem': don't give up on content, but recognize that it is in part pragmatically determined, and confine it to an explanatory gloss.

Let's return to the central issue of this section: what is it for an internal state to function as a *representation*? I suggest that focusing on non-cognitive cases – the Watt governor, Ramsey's car – tells us very little about how representations function in accounts of cognition. Intuitions about these cases are not dispositive. It is more fruitful to focus on the typical explanatory context in which a theory in cognitive science is

[11] Rodney Brooks (1991, 81) famously claimed: "…explicit representations and models of the world simply get in the way. It is better to use the world as its own model." Despite the rhetoric, Brooks doesn't argue against representations per se, but rather against positing general context-free representation of the environment, and separate, explicit representation of goals. He is the father of 'action-oriented representations'.

developed – a manifest cognitive capacity such as seeing what is where in the scene, locomotion, manipulating objects in view, and so on – and ask under what conditions such a theory is committed to representations. This project is more modest – it won't tell us what it is to function as a representation *in general*. There may not be an interesting non-disjunctive answer to that question.[12] Rather, what we seek is an account of what it is to function as a representation in an explanatory account of a cognitive capacity. This would fall short of a general metaphysical account of representation, but it would be interesting nonetheless.

As it happens, our characterization of the adder [figure 1] provides the basis for answering the question. The mapping f_R isolates the causal structure relevant for the exercise of a given cognitive capacity. This will typically involve characterizing a set of states or structures, and the properties of these states or structures in virtue of which they play the distinctive roles they do in the exercise of the capacity. These states/structures will function as representations – in particular, as representational *vehicles* – just in case they are interpreted by a mapping f_I that assigns them contents. Given the content assignment specified by f_I the states or structures specified by f_R are not 'mere causal relays', as they would be without the semantic interpretation.

The representational vehicles specified by f_R are as real as states or structures posited in any well-confirmed scientific explanation of observable phenomena. An analogy may be helpful: genes are realized by physical/chemical structures; molecular biology groups these structures together by their causal powers to produce proteins ultimately responsible for particular phenotypical effects, abstracting away from some of

[12] The concept *representation* may not pick out a natural kind but rather be a *motley*, functioning differently in different contexts. This possibility can't be ruled out a priori.

their more basic physical/chemical properties. Similarly, the realization function (f_R) abstracts away from some of the properties of the realizing neural states and groups them together by their role in cognitive processing. In both cases, the states/structures may be multiply realized by states/structures characterized at the more basic level. In both cases, assuming that the theory is empirically well-confirmed, a realist attitude toward the posited structures is appropriate.

The upshot is that a cognitive model – whether so-called 'classical', connectionist, dynamical, embodied, or enactive – posits representations just in case it identifies representational vehicles, via f_R, and assign them contents in f_I. The kinds of states or structures that can count as representational vehicles – the kinds of objects and properties specified by f_R – is left open.[13] Intuitions grounded in our familiarity with public representational systems carry little weight here. A connectionist model that construes characteristic patterns of activation of hidden units to be causally efficacious in the exercise of a given cognitive capacity *and* assigns these patterns of activation contents in an appropriate gloss would thereby posit representations.

An implication of the view is that to determine whether an explanatory theory of a cognitive capacity posits representations it must be articulated at the level of structures and processes. Absent an account of the causal organization of the system given by f_R, we cannot determine the representational commitments of the theory. That said, my account of computational characterization is something of an idealization. A complete f_R mapping specifies precisely how the mechanism is realized in neural hardware. Many

[13] Since f_R specifies the causal organization of the system, the relevant objects and properties must be capable of having causal powers. *Abstracta*, therefore, cannot function as representational vehicles.

computational models are not fully articulated at the level of neural structure. The important point here is that to assess the representational commitments of the theory it must posit structures/states to serve as representational vehicles, and causal processes in which these vehicles are involved, even if the realizing neural details are yet to be supplied.

The proposed account of mental representation couples a realist account of representational vehicles and a pragmatic account of representational content. The resulting package is *deflationary* about mental representation. Contents serve a variety of heuristic purposes but are not part of what I have called the 'theory proper.' They are, strictly speaking, not necessary to explain the target phenomena and are best construed as part of an explanatory gloss. They are not determined by a privileged representation relation but are rather motivated by a variety of pragmatic considerations. A deflationary view of mental representation is not a species of *fictionalism*.[14] Fictional objects cannot play causal roles in cognitive processes, as representations are presumed to do. Neither is it a version of *interpretivism*, as that view is normally understood.[15] The states/structures that are interpreted in the gloss have their causal roles – though, of course, not their contents – independently of the interpretative practices of theorists.

5. Satisfying the Adequacy Conditions

Let us see how the deflationary account fares with respect to Ramsey's (2007) adequacy conditions.

[14] A fictionalist construal of neural representation has been discussed (though not endorsed) by Sprevak (2013).
[15] See, for example, Dennett (1987).

(1) Mental representations must serve a function sufficiently like paradigm cases of representation.

Considering the variety of functions served by public representations with which we are familiar – utterances, inscriptions, maps, photographs, graphs, and so on – it isn't clear that there is a single function shared by paradigm cases. However, representations are often said to *stand in* for the object or property specified by their content, where the relation of 'standing in' is left sufficiently vague to cover the central cases. But this is no less true for mental representations, as characterized by the deflationary account. Once a representational vehicle is assigned a content in an appropriate gloss, then it can be regarded, for all intents and purposes, as standing in (in the same vague sense) for the object or property specified by that content. The stand-in plays a characteristic causal role in the exercise of the target cognitive capacity.

(2) The content of mental representations should be causally relevant to their role in cognitive processes.

Many philosophers have made this point.[16] Dretske (1988) talks about "… content getting its hands on the wheel." Of course, since content is abstract, it cannot literally be a cause of anything. Rather, the requirement seems to be something like this: the content that a state has *causally explains* the role that the state plays in cognitive processing. So understood, the requirement puts the cart before the horse, and so should be rejected. Content captures a salient part of the causal nexus in which the state is embedded. For example, construing the frog's internal structure as representing *fly* emphasizes the causes of its tokening in the frog's normal ecological niche (its production); construing it as

[16] For a sample of the literature promoting this idea see Dretske (1988), Segal and Sober (1991), and Rescorla (2014).

representing *frog food* emphasizes downstream nutritional effects of its tokening (its consumption).) Thus it is no surprise that content *looks* to be causally relevant – one of its jobs, as noted above, is to characterize internal structures/states in a way that makes perspicuous their causal role in a cognitive process, again, given specific explanatory concerns. But content doesn't causally explain anything.

(3) The account should not imply pan-representationalism: lots of clearly non-representational things shouldn't count as representations.

Pan-representationalism is not a worry for the deflationary account, because it does not purport to offer a *metaphysical* theory of representation. It does not specify a general representation relation that holds independently of explanatory practice in cognitive neuroscience. This is one sense in which the account is deflationary. The view has no implications for Venus-fly traps, Watt governors, and some of the other things Ramsey cautions may turn out to be representations if the account is not sufficiently constrained.

(4) The account should not *under-explain* representational capacities; it should not, for example, presuppose such mental capacities as *understanding*.

The realization function f_R isolates the causal structure relevant for the exercise of the target cognitive capacity. It characterizes the states or structures that serve as representational vehicles and the properties of these states/structures in virtue of which they play the distinctive causal roles they do in the exercise of the capacity. Cognitive processes are not sensitive to any semantic or intentional properties that the vehicles may be assigned in the interpretation function f_I, so the theory does not posit or presuppose any intentional processes such as understanding. Moreover, as explained above, there is

no appeal to a representational relation in what I call the 'theory proper'. So the deflationary account does not under-explain our representational capacities, in Ramsey's sense. If anything, it seems at risk of violating his final condition:

(5) It should not *over-explain* representational capacities, such that representation is *explained away*.

According to Ramsey, if representations function as 'mere causal relays' then, in effect, the phenomenon of interest has disappeared. Causal relays are ubiquitous; surely not all of them are representations. I claim that representations are distinguished from *mere* causal relays by the fact that they are assigned contents by the interpretation function f_I, but since the content assignment is confined to the heuristic gloss, it might be argued that the phenomenon of interest – representation – has indeed disappeared. My response to this charge is to challenge the adequacy condition.

Cognitive neuroscience purports to give *reductive* accounts of cognitive capacities. This is what Fodor and Field, motivated by the conviction that intentionality is not fundamental, were asking for in the passages quoted above. The same conviction motivates the naturalistic semantics project. But a reductive account of a phenomenon – especially a *mental* phenomenon, with which we have an intimate, first-person acquaintance – will often tend to look like over-explaining. The phenomenon of interest may seem to have 'disappeared.' By the same token, biochemistry, in explaining the essence of life in terms of carbon-based molecular processes, may appear to have over-explained its target: the special *elan vital* that we know and value has, in effect, disappeared. But if our representational capacities are really to be explained – *naturalistically* explained – then at some point the notions 'representation' and 'content'

are going to drop out of the account given by the theory, and what is left may look like mere causal relays. The appropriate reaction is not to find fault with the reductive theory (assuming it is well-confirmed), or with the urge to subsume the phenomenon of interest under more fundamental processes that are better understood. Successful reduction, and the unification that it makes possible, is the hallmark of scientific progress.

Nonetheless, there is *something* right about the 'don't over-explain' requirement. A reductive account of a phenomenon that is both central to our way of understanding ourselves and also, pretheoretically, somewhat mysterious – as *life*, *intentionality*, and *consciousness* certainly are – creates an *explanatory gap* of sorts between the account given by the theory and the commonsense conception of the phenomenon with which we began, a gap, in other words, between the scientific and the manifest image, as Wilfrid Sellars (1962) would have put it.[17] This gap typically leaves the reductive theorist with an obligation to connect the theory with the pre-theoretically conceived explanatory target, and this is precisely the function served by an explanatory gloss. In the case of a reductive explanation of our representational abilities, what is required is an *intentional* gloss connecting the theory proper with the intentionally characterized phenomenon with which we are pre-theoretically familiar.

One needn't accept my pragmatic account of mental content to see the point. An intentional gloss would most likely be needed even if the naturalistic semantics project were to succeed in specifying sufficient non-semantic and non-intentional conditions for a mental state's having the meaning it does. There are at least two reasons for this. In the first place, existing naturalistic theories, at best, require further conditions to resolve

[17] The explanatory gap between reductive proposals for consciousness and phenomenal experience is, of course, the most famous example.

indeterminacy. Perhaps striking out in an entirely new direction is a more promising strategy. In any event, if there *are* non-semantic and non-intentional conditions that ground determinate content they are likely to be highly disjunctive or their specification otherwise very complex.[18] There is no reason to think that such conditions would be *explanatory* of intentionality, because they would not necessarily contribute to our understanding of intentional phenomena in any significant way. The job of connecting the naturalistic theory with the target phenomenon – *meaning* – would be left for a gloss. Secondly, a naturalized reduction of intentionality is likely to leave what is distinctively personal out of the picture. If there are naturalistic conditions for content, then what we think of as distinctively *mental* representations – thoughts and feelings – may turn out not to be special. The conditions may be satisfied by all kinds of *mindless* systems. For example, plants have circadian clocks, and it has been argued that they represent temporal properties. But plants are not thought to have what Morgan (2014) calls *mental-grade* intentionality. We need to consider the possibility that from a detached, naturalistic perspective there may not be any *distinctively mental* representation. But, of course, human minds don't just present themselves as objects for scientific study; we have direct first-person acquaintance with our own states of minds, and it is the phenomena with which we are intimately acquainted (thoughts and feelings!) that will seem to have

[18] A case in point is Fodor's ultimate (1990, 121) formulation of his *asymmetrical dependency* proposal, which requires three somewhat (in this reader's opinion) non-intuitive conditions, and yet still leaves the possibility of (at very least) Quinean indeterminacy, and so requires still further conditions.

disappeared. Reconciling these two perspectives – finding what the theory seems to have lost – is a job for a gloss.[19]

6. Is the deflationary account empirically accurate?

The deflationary account has recently come under attack as failing to accurately describe actual practice in cognitive neuroscience. The charge is that computational theories are fully committed to representations; the attribution of representational content is not a mere gloss. I shall consider arguments offered by William Bechtel and Michael Rescorla in turn.

Appealing to the development of theories of spatial representation in the rodent brain, Bechtel (2016) argues that:

> "… much neuroscience research is in fact directed at determining which neural processes are content bearers and understanding how they represent what they do. Content characterizations are not mere glosses on the research; the goal of the research is to determine what content the representations have." (1291)

The discovery in the 1970s of 'place cells' in the rat hippocampus prompted research on the role of these cells in navigation, which eventually led to the discovery of other types of neurons – grid cells, head-direction cells, boundary cells – whose firings correlate reliably with tokenings of various spatial properties in the local environment. These cells were shown to interact with place cells in the mechanism responsible for spatial navigation. Bechtel says of this and related work:

[19] Much more needs to be said about the relation between the cognitive contents posited in explanatory glosses of computational models and personal-level contents, but this issue is beyond the scope of the present paper.

> A strategy neuroscientists have employed with great success in attempting to understand the mechanisms that underlie cognitive abilities is to identify cells in which the rate of action potentials increases in response to specific stimulus conditions. They then construe such neurons as representing those features in the environment whose presence is correlated with the increased firing and attempt to understand how subsequent neural processing utilizes representations that stand in for those features of the environment in guiding behavior. (1288)

So, for example, place cells respond to particular regions of the local environment. They are said to *represent* that location. Head-direction cells are so-named because they respond to head direction, and are said to *represent* head direction. It does not follow, however, that these content attributions play an essential role in the theory, or that the goal of the research is to "to determine what content the representations have," as Bechtel claims.

The significant theoretical achievement here is specifying the distal conditions to which the cell's firing is responsive and determining its role in controlling subsequent behavior. That is the goal of the research, not determining the content that the posited representations have. Once the cell's role in the cognitive process has been characterized the theoretical heavy lifting is done. Talk of the cell's firing *representing* its distal stimulus conditions is a convenience – a gloss – that adds nothing of theoretical significance. Recall one of the functions of content ascription I identified earlier: to characterize internal structures/states in a way that makes perspicuous their causal role in

a cognitive process that typically extends into the environment. This is the main function of content ascription here.[20]

Arguing that representational content plays a fundamental role in cognitive neuroscience, Bechtel goes on to say:

> … an early and integral step in the investigation of how specific information is processed within organisms appeals to representational content to determine representational vehicles. Initial characterizations of the vehicles and attributions of content are then both subject to revision as more vehicles are discovered and the processing mechanisms that generate the relevant activity and respond to it are identified. What is especially important is that such additional inquiry is inspired and guided by the initial attributions of representational content and directed at fleshing out the account. The attribution of content is a first step in articulating an account of a mechanism for processing information. (1291)

Here Bechtel seems to recognize that the goal of the research is to identify the structures and processes responsible for the target capacity. He points out that content attributions can play an important role in their discovery, illustrating one of the functions of representational content I identified above: to serve as a temporary placeholder for an incompletely developed computational theory and to guide the discovery of mechanisms underlying the capacity. Characterizing to-be-discovered structures in terms of content

[20] William Ramsey would deny that place cells and other 'receptors' that are regularly activated by some distal condition are really representations. He would claim that they are 'mere causal relays'. But it is hard to draw a principled line between these cases and others that clearly do seem representational. My strategy is to agree that these cases qualify as representational – deploying a *deflationary* construal of representation, i.e. interpreted structures (vehicles) posited in the service of cognitive capacities – and then construe the assigned content as part of a heuristic gloss.

allows the theorist to formulate hypotheses about the causal roles of the structures she is investigating. To be sure, it is not appropriate to call such content ascription a *gloss* because at this early stage there may be little or no theory to gloss – representational vehicles have yet to be fully characterized – but the relevant point is that the content ascription serves an explicitly *heuristic* purpose, analogous to glosses deployed in developed theories.

In conclusion, the deflationary account I favor explains the rat navigation case quite well. And since Bechtel's argument depends on general features of neuro-scientific theorizing, there is good reason to think the account will handle a wide range of cases.

Another version of the empirical accuracy challenge focuses on a very different class of cognitive models. Michael Rescorla argues that my deflationary account is false of Bayesian psychological models. He claims that representational content plays a fundamental and essential role in Bayesian theorizing:

> Bayesian models individuate both *explananda* and *explanantia* in representational terms. The science explains perceptual states *under representational descriptions*, and it does so by citing other perceptual states *under representational descriptions*. For instance… the generalizations type-identify perceptual states *as* estimates *of* specific distal shapes…. Thus, the science assigns representation a central role within its explanatory generalizations. The generalizations describe how mental states *that bear certain representational relations to the environment* combine with sensory input to cause mental states *that bear certain representational relations to the environment.* (2015, 14, emphasis in original)

Rescorla claims that Bayesian perceptual models construe perceptual states as essentially representational; their distal content plays an essential role in specifying these states. In another recent paper, on sensorimotor models, he characterizes the Bayesian program as follows:

> Researchers adopt a two-step approach: first, construct a normative model describing how an optimal Bayesian decision-maker would proceed; second, fit the normative model as well as possible to the data…. Our model yields ceteris paribus generalizations relating sensory input, mental activity, and behavior. We evaluate through experimentation how well the generalizations describe actual humans. Hence, the basic explanatory strategy is to use Bayesian normative models as descriptive psychological tools. This explanatory strategy presupposes that the motor system largely conforms (at least approximately) to Bayesian norms. (2016a, 31-32).

I shall make two points about Rescorla's characterization of Bayesian psychological models. In the first place, and most importantly for the discussion of the empirical accuracy of the deflationary account of representation, Bayesian models are typically not developed at a level of description that allows us to assess their representational commitments. More accurately, they have no representational commitments, in the relevant sense. There is no computational implementation – no commitment to internal states or structures and causal processes defined on them – and so no commitment to representational *vehicles*, in other words, no commitment to *representations*, in the sense at issue. Bayesian models are the merest of mechanism 'sketches' (in the sense articulated by Piccinini & Craver 2011). It is not simply that we

don't know how the models are implemented in neural mechanisms. More relevantly, we don't have an account of the causal organization of the system at the level of abstraction specified by f_R.[21] If we had a computational implementation of a Bayesian mechanism then, but only then, could we determine whether the contents assigned to the posited states play an essential, individuative role in the theory, or whether they function as a gloss of the sort I have proposed. This is certainly not intended as a criticism of the Bayesian program. In the absence of a computational implementation, how else is the theorist to describe to-be-posited internal states and processes except in intentional terms, by reference to their presumed distal contents?[22] It is merely to note that assessment of the representational commitments of specific Bayesian models must await their further development.[23]

---

[21] Rescorla is at pains to point out that Bayesian models are not committed to what he calls 'formal/syntactic' computation, claiming

> The science… individuates mental states in representational terms *as opposed to* formal syntactic terms. [2016a, 25 emphasis in original]

He is right – Bayesian models are not articulated at the level of structures and processes, so they are not committed to syntax. But syntactic objects are just one type of representational vehicle. A theory is committed to representations only if it posits representational vehicles and assigns them content (setting aside for present purposes whether the content assignment is in the theory or in a gloss), so a characterization of mental states in terms of content does not obviate the need to characterize them in terms of their causal role in cognitive processes. Simply put: *no vehicles, no representations.*

[22] Thus, distal content ascription in Bayesian models, whatever else it may do, serves the placeholder function described above.

[23] It is worth noting the slide between "explanatory" and "descriptive" in the last two sentences of the Rescorla quote above. There is some dispute about the correct interpretation of Bayesian models: are they intended to *explain* actual psychological processes, or merely to *describe* them in a way that systematizes and predicts behavior? Colombo and Series (2012) argue for the latter view. They point out that current Bayesian models do not provide mechanistic explanations – they do not specify the structures and processes that implement Bayesian computations – and argue that at the current stage of theorizing an instrumentalist attitude toward the models is appropriate. An assessment of this instrumentalist conclusion is beyond the scope of the present paper, though see Egan (2017) for defense of the view that a characterization of the function (in

Secondly, to the extent that a realist construal of Bayesian psychological models is appropriate, they are committed to the claim that mental processes are probabilistic inferences, and that internal mechanisms compute probability distributions optimally, according to Bayes' theorem.[24] Under a natural interpretation, internal structures represent probability distributions.[25] In any event, Bayesian models, to the extent that they say anything about how the brain actually works, give what I have called a *function-theoretic* characterization; they specify the function, in the mathematical sense, computed by the mechanism.[26] The function is specified intensionally by Bayes' theorem.

Rescorla apparently thinks that the mathematical characterization is an artifact of our idiosyncratic conventions, rather than a central commitment of Bayesian psychology:

> Bayesian perceptual psychology offers intentional generalizations governing probability assignments to environmental state estimates. We articulate the generalizations by citing probability distributions and pdfs over mathematical entities. But these purely mathematical functions are artifacts of our measurement units. They reflect our idiosyncratic measurement conventions, not the underlying psychological reality. (2015, 32)

---

the mathematical sense) computed in the exercise of a cognitive capacity can be explanatory even absent an account of how the capacity is computationally (or neurally) implemented.

[24] Under idealization, of course, just as hand calculators and human subjects compute the addition function only under idealization.

[25] But, as Wiese (2017) points out, neither textbook Bayesian inference nor approximate Bayesian inference (in, for example, predictive processing models) requires representing probability values or values of probability density functions. He calls the problem of determining how the brain implements an approximation to Bayesian inference the *probability conundrum* and notes that different solutions to it have been proposed in the literature. Kwisthout & van Rooij (2013) argue that considerations involving computational tractability suggest that explicit representations of probability distributions are unlikely to be employed by the brain.

[26] See fn. 7 above.

This is very puzzling. To think that commitment to Bayes' theorem – a function defined on probability distributions – reflects an arbitrary choice of conventions is analogous to thinking that a claim that a device computes the addition function reflects a commitment to representing addends and sums in base 10.[27] *Contra* Rescorla, to the extent that Bayesian models are to be construed realistically – and if they are not, then disputes about the status of representational content in Bayesian models are pointless – such proposals should be construed as hypotheses about underlying psychological reality, committed, in particular, to the claim that the system is computing an approximation to Bayes' theorem.

To summarize my reply to the empirical accuracy objection, viz. the claim that theories in cognitive neuroscience and cognitive psychology make essential appeal to representation: (1) If the theory characterizes a cognitive capacity in terms of mechanisms, states, and processes (as in the account of rat navigation), then a deflationary reinterpretation of the representational talk employed by theorists is appropriate. Such talk is playing a gloss-like role. (2) If it does not characterize the capacity in terms of mechanisms, states, and processes (as in current Bayesian psychological models), then the theory has no representational commitments in the relevant sense, that is, no commitment to *representations*.

A final word on the so-called 'representation wars', currently raging over whether predictive processing models, enactivist accounts, and other recent approaches posit representations. The deflationary account is itself neutral in the representation wars. In

[27] Rescorla makes the same point in (2016b), arguing against my account of function-theoretic description that to characterize a device as computing a mathematical function is to commit to an arbitrary choice of measurement units.

particular, the idea that representational content functions as a kind of gloss has no implications for which broad classes of cognitive models, when the computational details are spelled out, carry representational commitments (other than 'classical' models, which undoubtedly do). But the view has implications for how the wars should be settled. A cognitive model posits representations just in case it identifies representational vehicles, via f_R, which play crucial causal roles in the exercise of the capacity, and assign these vehicles contents in f_I.[28]

References

Brooks, R.A. (1991), "Intelligence without Representation," *Artificial Intelligence* 47: 139- 159. (Reprinted in *Cambrian Intelligence*, Cambridge, MA:MIT Press 1999)

Chemero, A. (2000), "Anti-Representationalism and the Dynamical Stance," *Philosophy of Science* (67): 625-647.

Chemero, A. (2009), *Radical Embodied Cognitive Science*, Cambridge, MA: MIT Press.

Bechtel, W. (1998), "Representations and Cognitive Explanations," *Cognitive Science* 22: 295-318.

Bechtel, W. (2001), "Representations from Neural Systems to Cognitive Systems," in W.

[28] Thanks to Robert Matthews, the editors of this volume, and an anonymous referee for helpful comments on earlier versions of this paper. Thanks also to audiences at the Institute for Philosophy at the University of London, the Society for Philosophy and Psychology annual meeting at Duke University in June 2015, the conference on Mental Representations at Ruhr-University Bochum in September 2015, and the students in my Mental Representation seminar at Rutgers University in Fall 2017.

Bechtel, P. Mandik, J. Mundale, and R. Sufflebeam (eds.), *Philosophy and the Neurosciences,* Oxford: Blackwell, 332-348.

Bechtel, W. (2016), "Investigating Neural Representations: The Tale of Place Cells," *Synthese* 193:1287–1321.

Clark, A. (1997), "The Dynamical Challenge," *Cognitive Science* 21(4): 461-481.

Colombo, M. and Series, P. (2012), "Bayes in the Brain: On Bayesian Modeling in Neuroscience," The British Journal for the Philosophy of Science 63(3): 697-723.

Cummins, R. (1989), *Meaning and Mental Representation,* Cambridge, MA: MIT Press.

Dennett, D.C. (1987), *The Intentional Stance,* Cambridge, MA: MIT Press.

Dretske, F. (1981), *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.

Dretske, F. (1986), "Misrepresentation," in *Belief: Form, Content, and Function*, Radu Bogdan (ed.), Clarendon Press, 17-36.

Dretske, F. (1988), *Explaining Behavior*, Cambridge, MA: MIT Press.

Dretske, F. (1995), *Naturalizing the Mind,* Cambridge, MA: MIT Press.

Egan, F. (2014), "How to Think about Mental Content," *Philosophical Studies* 170: 115-135.

Egan, F. (2017), "Function-Theoretic Explanation and the Search for Neural Mechanisms," *Explanation and Integration in Mind and Brain Science*, David M. Kaplan (ed.), Oxford University Press, 145-163.

Field, H. (1975), "Conventionalism and Instrumentalism in Semantics," *Nous* 9: 375-405.

Fodor, J.A. (1975), *The Language of Thought*, New York: Thomas Y. Crowell.

Fodor, J.A. (1980), "Methodological Solipsism Considered as a Research Program in Cognitive Science," *Behavioral and Brain Sciences* 3: 63-109.

Fodor, J.A. (1987), *Psycho-semantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.

Fodor, J.A. (1990), *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.

Fodor, J.A. (2008), *LOT2: The Language of Thought Revisited*, Oxford: Oxford University Press.

Gallagher, S. (2008), "Are Minimal Representations still Representations?" *International Journal of Philosophical Studies* 16 (3): 351-69.

Gallistel, C.R. (1990), *The Organization of Learning*, Cambridge, MA: MIT Press.

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

Kirsh, D. (1990), "When is Information Explicitly Represented?" Vancouver Studies in Cognitive Science 340-365.

Kwisthout, J., & van Rooij, I. (2013). Bridging the gap between theory and practice of approximate Bayesian inference. *Cognitive Systems Research, 24*, 2-8. doi: http://dx.doi.org/10.1016/j.cogsys.2012.12.008

Marr, D. (1982), *Vision*, New York: Freeman.

Matthen, M. (1988), "Biological Functions and Perceptual Content," *Journal of Philosophy* 85: 5-27.

Millikan, R. (1984), *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.

Millikan, R. (1989), "Biosemantics," *The Journal of Philosophy* 86: 281-297.

Mollo, D.C. (2017), "Content Pragmatism Defended," *Topoi* https://doi.org/10.1007/s11245-017-9504-6

Morgan, A. (2014), "Representations Gone Mental," *Synthese* 191: 213-244.

Neander, K, (2006), "Content for Cognitive Science," in *Teleosemantics*, Graham F. Macdonald & David Papineau (eds.), Oxford University Press, 140-159.

Neander, K. (2017), *A Mark of the Mental: In Defense of Informational Teleosemantics*, Cambridge, MA: MIT Press.

O'Brien, G. and Opie, J. (2004). "Notes toward a Structuralist Theory of Mental Representation," in *Representation in Mind: New Approaches to Mental Representation*, H. Clapin, P. Staines, & P. Slezak (eds.), Oxford: Elsevier, 1-20.

Papineau, D. (1993), *Philosophical Naturalism*, Oxford: Blackwell.

Piccinini, G. and Craver, C. (2011), "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches," *Synthese* 183(3): 283-311.

Quine, W.V.O. (1960), *Word and Object*, Cambridge, MA: MIT Press.

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

Rescorla, M. (2014), "The Causal Relevance of Content to Computation," *Philosophy and Phenomenological Research* 88: 173-208.

Rescorla, M. (2015), "Bayesian Perceptual Psychology," in *The Oxford Handbook of the Philosophy of Perception*. M. Matthen (ed.), Oxford: Oxford University Press.

Rescorla, M. (2016a), "Bayesian Sensorimotor Psychology," *Mind & Language* 31: 3-36.

Rescorla, M. (2016b), "The Computational Theory of Mind," *Stanford Encyclopedia of Philosophy*.

Ryder, D. (2004), "SINBAD Neurosemantics: a Theory of Mental Representation," *Mind & Language* 19(2): 211-240.

Segal, G. and Sober, E. (1991), "The Causal Efficacy of Content," *Philosophical Studies*

63: 1-30.

Sellars, W. (1962), "Philosophy and the Scientific Image of Man," *Frontiers of Science and Philosophy*, Robert Colodny (ed.), Pittsburgh: University of Pittsburgh Press, 35-78. Reprinted in *Science, Perception and Reality* (1963).

Shadmehr, R. and Wise, S. (2005), *The Computational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning*, Cambridge, MA: MIT Press.

Shagrir, O. (2012), "Structural Representations and the Brain," *British Journal for the Philosophy of Science*, 63: 519-545.

Shapiro, L. (2010), *Embodied Cognition*, Routledge Press.

Shea, N. (2007), "Consumers Need Information: Supplementing Teleosemantics with an Input Condition," *Philosophy and Phenomenological Research*, 75(2), 404-435.

Shea, N. (2013), "Naturalizing Representational Content," *Philosophy Compass* 8: 496–509.

Sprevak, M. (2013), "Fictionalism about Neural Representations," *The Monist* 96: 539–560.

Tonneau, F. (2011), "Metaphor and Truth: A Review of *Representation Reconsidered* by W.M.Ramsey," *Behavior and Philosophy*, 39: 331-343.

van Gelder, T. (1995), "What Might Cognition Be, if Not Computation," *Journal of Philosophy* 91: 345-381.

Wiese, W. (2017), "What are the Contents of Representations in Predictive Processing?" *Phenomenology and the Cognitive Sciences* 16: 715-736.