

Cognition as the sensitive management of an agent's behavior

Mikio Akagi

John V. Roach Honors College, Texas Christian University

Abstract

Cognitive science is unusual in that cognitive scientists have dramatic disagreements about the extension of their object of study, cognition. This paper defends a novel analysis of the scientific concept of cognition: that cognition is the sensitive management of an agent's behavior. This analysis is "modular," so that its extension varies depending on how one interprets certain of its constituent terms. I argue that these variations correspond to extant disagreements between cognitive scientists. This correspondence is evidence that the proposed analysis models the contemporary understanding of cognition among scientists, without artificially resolving questions that are currently considered open.

Keywords

Cognition; Cognitive science; Analysis; Embodied cognition; Extended cognition; Mark of the cognitive.

This is a postprint. Please quote/cite the published version in *Philosophical Psychology* (published online 2021), DOI: 10.1080/09515089.2021.2014802.

1. The Problem of Cognition

Scientists in many disciplines disagree about the nature of their object of inquiry. For example, there is ongoing controversy about how to understand LIFE,¹ though this controversy rarely affects the work of contemporary biologists (Cleland, 2012; Machery, 2012). And there is little agreement on precisely what physical entities or events are (Montero, 1999; Earman, 1975), though physicists and engineers get on regardless. It is uncommon, though, that a scientific enterprise feature foundational conflicts about the *extension* of its object of inquiry, and therefore the bounds of its inquiry.

Nevertheless, since the 1980s the extension of COGNITION has been contested, and the cognitive science literature is replete with "border wars" over controversially cognitive phenomena including distributed cognition, extended cognition, embodied cognition, plant

¹ I use small capitals to indicate reference to a concept. Cognition is a phenomenon, "cognition" is a word, and COGNITION is a concept. In Sect. 3 onward, I indicate reference to uninterpreted parameters with italics.

cognition, and other controversial cases. There have been some attempts to adjudicate these disputes with various “marks of the cognitive” (notably Adams & Aizawa, 2001, 2008; Rowlands, 2009, 2010; Adams & Garrison, 2013; Buckner, 2015), but such proposals are generally as controversial as the judgments they vindicate. My goal in this paper is not to criticize these efforts nor to propose another criterion that seeks to settle which phenomena are “genuinely” cognitive. Rather, I aim to characterize the concept of cognition as understood by contemporary scientists, including its rough edges and indeterminacies. Such a characterization would be pluralistic (though perhaps not quite “relaxed” in Colin Allen’s [2017] sense). Put another way, I propose to explicate the concept *COGNITION* ecumenically, so as to represent its present controversies perspicuously, rather than to end them.

An ecumenical analysis of *COGNITION* will not settle the extension of cognition or adjudicate the border wars, but there are other purposes such an analysis might serve (see [Akagi, 2018] for more discussion). First, such an analysis could make explicit the way cognitive scientists think about their object of study, especially in contrast to pretheoretic conceptions of cognition or thought, perspicuously representing agreement and disagreement between scientists. Second, the scientific concept of cognition is poorly understood by laypeople and students, who are surprised that scientists take cognition to include phenomena like emotion, motivation, and motor control. An ecumenical analysis of *COGNITION* could make cognitive science more accessible to non-specialists. Third, an ecumenical analysis could inform philosophical research regarding topics that draw on the empirical sciences of the mind. These three purposes can be accomplished even by an analysis that leaves the controversies of the cognition border wars open. And by leaving those controversies open, an ecumenical account avoids the hubris of predicting what scientists will find or what form their theories will take (as feared by e.g. Cleland [2012]; Allen [2017]; Ramsey [2017]; Akagi [2018]).

My present task, therefore, is to defend a new analysis of the scientific concept of cognition that can serve these three purposes. I will begin by establishing landmarks for the contested extension of *COGNITION*, introducing test cases in three categories: some paradigmatically cognitive phenomena, some phenomena generally agreed not to be within the purview of cognitive science, and several controversial cases from the border wars (Section 2). I will describe a method I call modular analysis that models contested extensions (Section 3). Then I will describe the sensitive management hypothesis (SMH), a modular analysis according to which, in a slogan, cognition is the sensitive management of an agent’s behavior (Section 4). I will argue that the sensitive management hypothesis correctly classifies the test cases, thus providing strong evidence of its extensional adequacy (Section 5), before reflecting on the hypothesis’s potential to serve the three purposes described above (Section 6).

2. Test Cases

Even if we ignore the disagreements of the border wars, contemporary cognitive scientists use the word “cognition” in at least two distinct senses (see Greene et al., 2004, p. 389; Rowlands, 2009, p. 7; Akagi, 2018, pp. 3553f.) which I’ll call *highfalutin* and *inclusive* cognition. “Cognition” in its highfalutin sense (“higher cognition”) has an extension similar to that of the pretheoretic concept THOUGHT, and is generally understood in contrast to perception or affect. Highfalutin “cognition” is a scientific term, and features in scientific expressions like “cognitive therapy” and “cognitive control.” However, cognitive scientists often speak of cognition in a broader sense such that its extension also includes perceptual, emotional, motivational, and (some) motor phenomena. The extension of “cognition” in this inclusive sense is, even on a conservative reckoning, revisionary relative to a lay understanding of “cognition” or “thought.” Both highfalutin and inclusive “cognition” have contested extensions, but the purview of cognitive science definitely includes many perceptual and affective phenomena, which are happily called “cognition” by many. So insofar as the border wars are about the bounds of the cognitive sciences they must be about which phenomena count as cognitive in the inclusive sense.

The border wars notwithstanding, cognitive scientists generally agree about many paradigmatic cases of inclusive cognition. For example, face recognition—the capacity of humans (and perhaps some other animals) to distinguish and recognize individual faces visually—is a cognitive capacity. In humans this capacity is explained by cortical activity (often with a focus on the right fusiform gyrus; see e.g. Kanwisher, 2010). Other uncontroversially cognitive phenomena include a typical person’s capacity to navigate a familiar city, as well as the capacities to see color, make decisions, feel disgust, and control one’s limbs, at least when these capacities belong to animals not using external aids or other technology. Cognitive scientists also agree on many consensual cases of non-cognitive phenomena. A rock warming in the sun is not a cognitive phenomenon. Nor is the erosion of a stream bed, or the relationship between the time it takes light to reach the Earth and the distance of its point of origin.

But many phenomena are controversially cognitive, where contrary verdicts are associated with theoretical frameworks and empirical research programs. I will focus throughout this paper on six phenomena that each have vigorous proponents and opponents in various disciplines (though I will largely cite philosophers in what follows, since philosophers tend to be the most explicit about abstract commitments). Readers familiar with these cases could skip to Section 3.

First, proponents of *distributed cognition* contend that the coordinated activity of multiple agents sometimes constitutes a kind of cognition over and above individual cognition. For example, maritime navigation—the representation and planning of a ship’s course—is not attributable to any individual sailor, even with help. Rather, navigation is a cognitive task

accomplished by a system that consists of crewmembers, their equipment, and their implementation of various procedures (Hutchins, 1995).

Second, disputes about *robot cognition* are as old as cognitive science and older (Turing, 1950; Putnam, 1975) but recent discussions often invoke the robot Herbert, which moved autonomously, avoided obstacles, and pilfered unattended soda cans (Brooks et al., 1988). Herbert is sometimes regarded as a compelling model for cognition (Brooks, 1991),² and sometimes regarded as an elaborate tool with no cognitive capacities (Adams & Garrison, 2013).

Third, Otto is a fictional man who accommodates a memory impairment through use of a notebook. Proponents of *extended cognition* hold that Otto's cognitive processes occur partly outside his skin, in the notebook (Clark & Chalmers, 1998); opponents argue that while the notebook must be involved in the explanation of Otto's behavior, the notebook itself is not part of a cognitive mechanism (see e.g. Adams & Aizawa, 2001; Rupert, 2004).³

Fourth, proponents of *embodied cognition* argue that cognition occurs in parts of the body outside the central nervous system (Varela et al., 1991; Chemero, 2009). For example, female crickets have special adaptations allowing them to reliably locate mates. A system of tracheal tubes in their bodies amplifies the frequencies of conspecific mating calls (Webb, 1994). Proponents of embodied cognition claim that the tracheal tubes are components of the cricket's cognitive system, whereas opponents argue that the cognitive action, so to speak, is entirely in the cricket's neural tissue.

Fifth, proponents of *plant cognition* maintain that plants behave, though at longer time scales than animals, and that this behavior is coordinated by cognitive processes. John Haugeland (1991) once supposed that if a "super-sunflower" could track the direction of the sun in the absence of sunlight, it would possess cognitive capacities. Paco Calvo (Calvo Garzón, 2007) argues that *Lavatera cretica* is such a plant; it orients toward the sun even when the sun is obscured for several days. Calvo and others now suggest that plants are capable of learning, and flexible decision-making based on multiple cues (Segundo-Ortín & Calvo, 2019).

Sixth, there are defenders of *microbe cognition*. The example most familiar to philosophers is Fred Dretske's (1986): "magnetotactic" bacteria direct their locomotion, e.g. away from oxygenated water, through sensitivity to the Earth's magnetic field. This adaptive locomotion is sometimes regarded as cognitive behavior.

These six contentious cases of cognition—distributed cognition in maritime navigation, robot cognition in Herbert, extended cognition in the Otto-notebook system, embodied

² Brooks does not defend Herbert as a case of genuine cognition per se, but he suggests that cognition (or "intelligence") need consist in nothing more than the kinds of capacities Herbert possesses, produced by mechanisms like Herbert's.

³ In Clark and Chalmers' original paper, Otto is an example of the extended mind hypothesis, which is stronger and more speculative than the extended cognition hypothesis. But Otto is the most discussed example in the extended cognition literature, so I use it here.

cognition in cricket audition, plant cognition in *L. cretica* phototropism, and cognition in bacterial magnetotaxis—do not represent all of the disputes of the border wars. In particular, they do not represent disputes that concern the proper method of modeling or explaining cognition, e.g. disputes over dynamicism (van Gelder, 1998). But they illustrate a substantial amount of the flavor and variety of scientific views about the contested extension of COGNITION.

3. Modular Analysis

My aim is to produce an analysis of COGNITION that is adequate to these cases—not by settling whether the controversial cases are “genuinely” cognitive, but by correctly classifying cases as consensually cognitive, controversially cognitive, or consensually non-cognitive. Since these cases reflect informed scientific judgments, rather than lay judgments, the proposed analysis will be an empirically-informed one (as is urged by e.g. Kornblith [2017]). Traditional conceptual analyses consisting of necessary or sufficient conditions do not accomplish non-dichotomous classifications gracefully, so I will avail myself of a method called modular analysis.⁴

Consider the following toy analysis of the concept BRAVE:

An action ϕ , performed by an agent a , is brave iff ϕ is dangerous for a and ϕ is *admirable*.

Suppose that in our toy linguistic community there is general agreement concerning which actions are dangerous for their agents, but disagreement concerning which are admirable. We can model the usage of BRAVE in this community by understanding the analysis above as a traditional analysis (with necessary and sufficient conditions), except that the extension of ADMIRABLE varies based on the diverse ways that members of the community reckon admirableness for actions. I call an analysis so understood a *modular analysis*, since the function (from actions to semantic values) corresponding to the predicate “admirable” can be swapped out for a different function. Let us refer to swappable predicates like “admirable” as *parameters* (by analogy with the use of the term in mathematics, where assigning values to parameters transforms a schematic formula into a specific one, and since the term “variable” is normally reserved in predicate calculus for names that can be arguments of predicates). The different

⁴ For fans of Carnapian explication: we can consider the pretheoretic concept of cognition an explicandum (an imprecise, old concept), and the scientific concept of inclusive cognition (my analysandum) an explicatum of sorts (a new concept embedded in a theoretical context), in want of an explicans, or characterizing expression. Modular analysis is a nonstandard form for an explicans. In general the extensions of explicata can differ significantly from the extensions of their explicanda, but SMH is meant as a precise characterization of an existing but informal explicatum, so in order to be adequate it must describe the extant, contested usage of “cognition” in its inclusive sense.

functions correspond to different criteria for ascribing parameterized predicates, and can be called “values” of the parameter or, more colloquially, “interpretations.” Parameters often correspond to loci of disagreement between users of a contested concept. A modular analysis proper (that is, a modular analysis considered without a domain of parameter values) represents the extension of the analysandum as a relation to the various values of its parameters. Modular analysis as used here is meant to be a modeling strategy, not an insight into the secret, “true” meanings of some concepts. The benefit of a modular analysis is that it yields multiple extensions for its analysandum (e.g. BRAVE), so it can represent a state of disagreement about how to use the analysandum concept and license appropriate inferences about the ways various speakers use concepts. Yet modular analyses are easy to understand, and perspicuously represent both the commonalities and differences in the ways that members of the linguistic community use the analyzed term.

Perhaps, for example, Marilyn is an older woman who enjoys skydiving, a dangerous activity for its agent. Her friend Chandana thinks Marilyn’s skydiving is admirable, because she is inspired by Marilyn’s defiance of social expectations. Marilyn’s son Kasson, by contrast, considers skydiving irresponsible; he does not see skydiving as admirable. Marilyn herself has no opinion on whether skydiving is admirable. The toy analysis above may be considered an extensionally adequate modular analysis if and only if it licenses the correct inferences about who thinks Marilyn’s skydiving is brave: if Chandana considers it brave, Kasson does not, and Marilyn has no opinion. We may say of predicates like “brave” and “admirable” that they each have multiple *sectarian extensions*, i.e. the relatively classical extensions they have corresponding to Chandana’s judgments, to Kasson’s judgments, etc.). Some sectarian extensions of “admirable” contain skydiving, and some do not. However, each predicate has only one *ecumenical extension*, which corresponds to the set of its sectarian extensions.

The introduction of parameters into analyses creates some methodological hazards that must be managed. First of all, since analyses should provide criteria and not vague guidelines, it must be clear which predicates in the analysis function as parameters and which do not. And distinct parameters should be relatively independent of each other (the choice of interpretations across parameters should be at least logically independent). Second, the extensional adequacy of a modular analysis must be assessed relative to the values of its parameters, so counterexamples are only valid if they correspond to a consistent way of speaking. For instance, it is no objection to the toy analysis that skydiving is admirable (according Chandana) but not brave (according to Kasson). A licit counterexample requires that all expressions be interpreted coherently (e.g. if skydiving is brave according to Chandana but not admirable according to Chandana). In addition, a modular analysis cannot be thoroughly assessed for extensional adequacy unless we specify the permissible interpretations of its parameters. Ideally, we should be able to acquire evidence about the values of the parameters for particular subcommunities

of speakers (or contexts of use) independently of evidence about the usage of the analysandum. For example, in our toy scenario we determine whether Kasson sees skydiving as admirable and, independently, whether he sees it as brave. If we cannot gather independent evidence about the values of the parameters they are “free parameters,” which permit a degree of ad-hocness. It is a virtue for analyses and other empirical formulae to have as few free parameters as possible.

4. The Sensitive Management Hypothesis

Modular analyses have the right form for modeling the extensions of contested concepts like COGNITION. Consensually cognitive phenomena should belong to every sectarian extension of COGNITION, controversially cognitive phenomena should belong to some but not all sectarian extensions, and consensually non-cognitive phenomena should belong to no sectarian extensions. I am only concerned to accommodate first-order judgments of cognitive scientists (i.e. about which phenomena are cognitive), not higher-order judgments (i.e. general claims about what kind of thing cognition is). And unlike traditional Jackson-style analysis (Jackson, 1998), I ignore lay judgments. The test cases described in Section 2 reflect scientific judgments about cognition that are informed by the counterintuitive empirical findings and theories of the last half-century. An analysis that is adequate to those test cases and others like them will reflect an empirically-informed concept of COGNITION, not a pretheoretic one.

I contend that the test cases described in Section 2 are correctly classified by the sensitive management hypothesis (SMH), a modular analysis according to which, in a slogan, *cognition is the sensitive management of an agent's behavior*. More precisely, the sensitive management hypothesis is that

A process ϕ is a cognitive process if and only if ϕ is a component process of a mechanism m , where:

1. m belongs to an agent a ,
2. m is sensitive to the circumstances of a , and
3. m manages the behavior of a .

SMH contains three parameters: *BELONGING TO AN AGENT* (abbreviated as *AGENT*), *SENSITIVITY TO CIRCUMSTANCES* (*SENSITIVITY*), and *BEHAVIOR*. While some of these terms are neologisms, they each correspond to concepts that have received significant attention from philosophers of cognitive science. When the parameters are each assigned an interpretation, the analysis determines a sectarian extension. SMH is extensionally adequate only if, with different assignments of interpretations to these parameters, the sectarian extensions of this formula always include consensual cases of cognition, never include the

consensual non-cases, and variously contain or exclude controversial cases like the six test cases described in Section 2. I will discuss the numbered conditions and their parameters in the next section; first I will unpack some unparameterized features of SMH.

SMH provides criteria for identifying cognitive processes, but can be used to define a number of other specific expressions. *Cognition* is the manifestation of a cognitive process. *Cognitive mechanisms* are mechanisms that satisfy conditions 1–3. *Cognitive states* are functional states of cognitive mechanisms. A *cognitive system* is an exhaustive collection of cognitive mechanisms belonging to a particular agent. *Cognitive behaviors* are behaviors that are managed by cognitive mechanisms. *Cognitive capacities* are capacities to exhibit cognitive behavior, or functional capacities of components of cognitive mechanisms, and so on. These formulations are consistent with the judgments of most cognitive scientists, e.g. that “cognition” does not refer to any behavior itself, but to the processes that produce it (Shapiro, 2013; Aizawa, 2017). It is sometimes suggested that some enactivists identify cognition with behavior, rather than with a precursor of behavior. However, even unorthodox judgments like this might be accommodated if “cognition” per se is identified with what I call “cognitive behavior” instead of “cognitive processes.”⁵ And I expect that enactivists would accept that there are processes that explain cognitive behavior, though they will find them in unorthodox places—associated with non-animal agents, and often located outside of the bodies of those agents—or they might appeal to processes that are not straightforwardly decomposable or localizable.

I analyze cognition in terms of “mechanisms,” but intend this term to be less freighted than it may at first appear. Of course, it is in line with some recent trends in the philosophy of science to understand some theories in the biological and behavioral sciences as concerned with mechanisms. The new mechanists (Machamer et al., 2000; Bechtel & Abrahamsen, 2005; Craver, 2007; Bechtel, 2008) claim that the main explanatory project of many special sciences is not to articulate universal, exceptionless natural laws, but to describe naturally-occurring structures—mechanisms—that exhibit phenomena we want to explain. These mechanisms are typically understood as organized complexes of component parts and operations.

Some argue that models in cognitive explanations are not mechanistic in this sense, despite the many similarities between cognitive models and mechanistic models in the biological sciences, because “the way in which [cognitive models] *correspond* to the underlying modeled system is far less straightforward” (Weiskopf, 2011, p. 332), or because some cognitive models do not derive their explanatory power from strategies of localization and decomposition (Silberstein & Chemero, 2013). I take it that these arguments are basically sound, but we lack a good word to refer to the truth-makers of the diverse sorts explanations found in cognitive

⁵ I’m not sure this is the correct reading—enactivists like Chemero, Noë, and Hutto and Myin seem to regard cognition as an interactive process of which behavior is a proper part rather than a causal consequence.

science—unless that word is, confusingly, “mechanism,” which is the word that cognitive scientists often use (and which had a life in the philosophical literature before the new mechanists; cf. Wakefield, 1992; Cummins, 2000). So I propose to use the word “mechanism” here to refer indiscriminately to whatever is described by adequate explanations in cognitive science: biological mechanisms, generic mechanisms (Stinson, 2016), target structures of adequate Weiskopf-style cognitive models, network dynamics (Silberstein & Chemero, 2013), concrete instances of functional analyses (Cummins, 1975), complexes of organism capacities and Gibsonian invariants (Gibson, 1979), and so on. While this is a heterodox understanding of “mechanism,” it is anodyne. Mechanisms will, in general, be structures that are organized such that they produce phenomena of interest, which in cognitive science are typically either certain capacities of agents or context-sensitive variations in behavior (i.e. “effects”; see Cummins, 2000). But the components of mechanisms in this weak sense may be relatively abstract. The components of cognitive mechanisms might include representations, modules, network properties, perceivable invariants in an environment, or arbitrary realizers of a functional role. The organization of mechanisms is often represented by graphs (see Danks, 2014), but can be specified more or less completely by equations or by description. Of course some cognitive models are phenomenal, not mechanistic (Craver, 2006; Hochstein, 2012), but even phenomenal models describe phenomena for which mechanisms in this thin sense are presumably responsible.

At any rate, “mechanism” is not a parameter in SMH. Disagreement about the nature of mechanisms is primarily disagreement among philosophers about what scientific practice is or should be, not disagreement between cognitive scientists about the merits of particular research programs. The existing practice and research goals of scientists are the standards against which various accounts of mechanism are judged (see e.g. Chirimuuta, 2014; Ross, 2015). At any rate, I am not convinced that differing interpretations of “mechanism” in the above sense correlate with different judgments concerning the extension of *COGNITION*. More expressions could be parameterized on a finer-grained view that aimed to articulate points of controversy beyond those of the border wars, but since explanatory power is *inter alia* a result of effective parameter reduction, I limit the number of parameters to the three I describe below.

5. Reproducing Expert Disagreement with Parametric Variation

Since a modular analysis cannot be evaluated without a domain of parameters, I will describe several values for each parameter, describing how they can be manipulated in order to selectively include or exclude various controversial cases, while consistently including paradigmatic cases and excluding non-cases. In principle, an interpretation is licit so long as it corresponds to a scientifically legitimized way of thinking about cognitive processes (for

Table 1: Selected interpretations for parameters of SMH.

Parameter value	Gloss
AGENT _α	A mechanism belongs to an agent only if it is part of the body of a biological organism.
AGENT _β	A mechanism belongs to an agent if and only if it regularly and reliably promotes the goals of a biological organism.
AGENT _γ	A mechanism belongs to an agent if and only if it regularly and reliably promotes the goals of a goal-apt system.
SENSITIVITY _α	A mechanism is sensitive to an agent's circumstances only if represents some of those circumstances and the representational content is non-derived.
SENSITIVITY _β	A mechanism is sensitive to an agent's circumstances only if represents some of those circumstances.
SENSITIVITY _γ	A mechanism is sensitive to an agent's circumstances if it functions to transform information about the agent's circumstances.
BEHAVIOR _α	A performance is a behavior only if it is an observable moving of the body.
BEHAVIOR _β	A performance is a behavior only if it is an observable moving of the body or it is a tropism.
BEHAVIOR _γ	A performance is a behavior only if it is an observable moving of the body or it is a taxis.

instance, one exhibited in published literature by cognitive scientists). I cannot describe an exhaustive list of licit values for the parameters here, so I will enumerate only three interpretations for each parameter, assuming for the sake of illustration that no other interpretations are permissible. This little fiction will allow me to illustrate the flavor of SMH as a modular analysis. The interpretations I describe are sufficiently varied to account for the test cases in Section 2. I submit that correct classification of the test cases is good, though defeasible, evidence for the adequacy of SMH as an analysis proper, which may be supplemented by subtler interpretations. For each parameter I will begin by considering relatively restrictive interpretations, discussing more liberal interpretations in turn. The interpretations are summarized in Table 1.

5.1. *Belonging to an Agent*

Most parties to the border wars seem to agree that it is objectionable to ascribe “un-owned” cognitive processes (Adams & Aizawa, 2001; Rupert, 2009; Clark, 2010a; Rowlands, 2010), though Rowlands is unusual in discussing the details of “belonging” to a subject at length (2010, Ch. 6). Rupert calls the owner or container of a cognitive process a “cognitive system.” Rowlands calls it a “representational subject.” Let us call that critter, whatever it is, a “cognitive agent” or “agent.”⁶

Cognitive agents in this sense need not be the “organisms” of biological theory, though paradigmatic cognitive mechanisms like those in the brain are parts of biological organisms. So let us say that on one interpretation of *AGENT* (*AGENT_α*), a mechanism belongs to an agent only if it is a nondetached part of the body of a biological organism. (The requirement that such parts be nondetached excludes e.g. secretions and body parts that have been shed, e.g. hairs, feathers, skin, shells.) Now, Otto is a biological organism, but SMH countenances Otto’s extended memory only if (parts of) his notebook can be considered cognitive mechanisms for him. Otto’s notebook is not a nondetached part of his body, since the notebook is separable from him and is sometimes physically distant (presumably he bathes without it). So the Otto-notebook system does not belong to the extension of *AGENT_α*. On a second interpretation of *AGENT* (*AGENT_β*), a mechanism belongs to an agent if and only if it regularly and reliably promotes the goals of a biological organism. Since Otto’s notebook helps him to navigate to his desired destination, it can belong to him in the sense required by *AGENT_β*.

A more liberal interpretation of *AGENT* might allow that an agent be any goal-apt system: any critter to which we attribute goals, such as survival and reproduction, or perhaps other goals such as constructing shelter, mauling yonder antelope, or following an experimenter’s instructions. Anything to which we can apply Dennett’s (1987) intentional stance is an agent in this sense. Let a mechanism belong to an agent in the sense of *AGENT_γ* if and only if it regularly and reliably promotes the goals of a goal-apt system. Hutchins’ ship may count as an agent_γ, so long as we attribute goals to it like traveling safely to its destination. And the navigation team—along with their equipment and procedures—are parts of a mechanism that promotes this goal. Similarly, Herbert possesses mechanisms that promote its goals, such as avoiding collisions and appropriating soda cans. This third interpretation respects the judgments of those who, like Fred Adams and Rebecca Garrison, deny that Herbert is a cognitive system on the ground that while Herbert appears to be goal-directed, it does not really have goals of its own; Herbert’s apparent goals are really the goals of Herbert’s creators (2013, pp. 342, 347f.). If Herbert has no genuine goals, it is not an agent_γ and therefore possesses no mechanisms that satisfy SMH.

⁶ For suggesting this term, I thank many colleagues including Joshua Shepherd, Samuel Asarnow, Peter Aronoff, and Lewis Powell.

Since navigation teams, Otto's notebook, and Herbert's robotic gizmos are each included in some but not all of these interpretations (α - γ) of *BELONGING TO AN AGENT*, they are controversial cases of things that belong to an agent. The face-recognizing mechanism in cortex, however, consensually belongs to an agent. It is part of the body of a biological organism, it promotes the goals of a biological organism, and it promotes the goals of a goal-apt system (i.e. a person). By contrast, consensually non-cognitive systems are excluded. A rock warming in the sun is not a biological individual nor a goal-apt system that possesses mechanisms. Similarly, by interpreting the other parameters of the sensitive management hypothesis, we can account for the variation in expert judgment about the other test cases.

5.2. Sensitivity to Circumstances

Cognitive mechanisms must be sensitive to the circumstances of an agent. I do not intend *SENSITIVITY* to be understood in the sense generally meant by epistemologists (e.g. Nozick, 1981), but in the sense that (assuming normal operation) the mechanism's functional states depend causally on selected states of affairs, perhaps imperfectly, but with non-accidental reliability. The language of "sensitivity" is intended to include accounts that rely on some notion of representation, as well as accounts that deny that "representation" so-called is necessary for cognition (e.g. Chemero, 2009; van Gelder, 1995; Brooks, 1991). Even anti-representational accounts acknowledge that cognition involves sensitivity to an agent's circumstances; they simply deny that the forms of sensitivity we discover when we examine cognitive systems are always happily called "representation." Appeal to representation and its analogues is the most common strategy for adjudicating border war disputes, and the variable interpretations of the *SENSITIVITY* parameter are meant to subsume most disagreements that turn on such appeals (though of course there are more, and more subtle, extant interpretations than I will discuss).

"Circumstances" here refers to states of affairs that bear on an agent's goals, either by being consistent with those goals, inconsistent with those goals, or by being such that an agent must modulate its behavior according to them in order to pursue its goals. Agents are generally sensitive only to a proper subset of their circumstances (cf. Millikan, 2017, pp. 146f. on the relativity of signs). Sensitivity is paradigmatically achieved through perceptual, interoceptive, or inferential processes, though other kinds of processes may also exhibit sensitivity to circumstances. It is sufficient for condition (2) of SMH that a mechanism's operation be modulated distally according to such sensitivities. For example, "place cells" in parahippocampal cortex respond differentially to an agent's position in space (O'Keefe & Dostrovsky, 1971), though they are rarely reckoned to be part of a perceptual system in the way that cells in visual cortex are. Nevertheless, their activity is sensitive to the agent's circumstances. This sensitivity is of course mediated by the connections between place cells and other brain

regions (just as, for example, a neuron can have a “receptive field” in virtue of mediated causal dependency).

The most traditional understanding of *SENSITIVITY TO CIRCUMSTANCES* is that a mechanism is sensitive to an agent’s circumstances only if it *represents* those circumstances, whatever “represents” means. For example, consider the role of cortical areas in visual face-recognition. Since patterns of neural activity are thought to be involved in facial recognition, brain areas (like right fusiform gyrus) are often said to represent faces (Kanwisher [2010], for example, uses representational language freely). Natural representations typically involve manipulations of a representing medium with functional significance in a mechanism—e.g. a pattern of neural activity—where different states of the medium track differences in the representand—e.g. visually distinct faces.

Activity in the fusiform gyrus is furthermore taken to be a case of representation with “intrinsic” or “non-derived” content—the sort of content had by mental representations but not by conventional representations like maps or printed words. Representations with merely derived content are often supposed to have content in virtue of their relations to representations with non-derived content, like mental representations. The stipulation that cognitive sensitivity be cashed out in terms of representations with non-derived content is controversial, but nevertheless figures in many accounts of cognitive representation (e.g. Searle, 1980; Adams & Aizawa, 2001; Rowlands, 2010). So we have our first interpretation of the parameter, *SENSITIVITY_α*: a mechanism is sensitive to an agent’s circumstances only if represents some of those circumstances, and the representational content is non-derived. Representations in cortical face-recognizing mechanisms satisfy this requirement.

The writing in Otto’s notebook does not have non-derived content. Defenders of extended cognition usually deny that cognitive sensitivity must be understood in terms of representations with non-derived content, and they tend to be suspicious of the very notion (e.g. Clark, 2005). Andy Clark and David Chalmers claim that Otto’s mnemonic mechanisms are realized in part by the state of his notebook, which is understood to have merely derived content. By denying that cognitive representations must have non-derived content, Clark and Chalmers make space in their conception of cognition for Otto’s notebook to serve as part of a cognitive mechanism. Likewise, Hutchins’ maritime navigators rely on representations with derived content, such as charts. Let us capture this more liberal perspective with a second interpretation of the parameter (*SENSITIVITY_β*): a mechanism is sensitive to an agent’s circumstances only if it represents some of those circumstances (where the representational content may be derived or non-derived).

On a third interpretation of *SENSITIVITY* (*SENSITIVITY_γ*), a mechanism is sensitive to an agent’s circumstances if it functions to transform information about an agent’s circumstances, even if it has no representing medium. Since I’m not defending a specific account of sensitivity

I don't wish to favor any particular information concept (e.g. Shannon [1948] information; Millikan's [2017] natural information). Let us say that tracheal tubes in crickets satisfy $SENSITIVITY_{\gamma}$ since they modify sound waves so as to emphasize the egocentric direction of conspecific mating calls. Likewise, Herbert's mechanisms satisfy $SENSITIVITY_{\gamma}$. Many would be uncomfortable calling such sensitivity representational (e.g. William Ramsey 2007), since most physiological processes might exhibit sensitivity in this sense. Even so, no features of stones and stream beds meet this permissive criterion, since stones and stream beds lack goals and therefore circumstances.

Each of these interpretations of the $SENSITIVITY$ parameter corresponds to avowed commitments of various researchers. The variation in researcher commitments about sensitivity mirrors complementary variations in researcher commitments about which phenomena are cognitive. Neural representations of faces satisfy all these interpretations. They satisfy the analyses of $SENSITIVITY_{\alpha}$ and $SENSITIVITY_{\beta}$ because they are ex hypothesi representations with non-derived content. And they are sensitive $_{\gamma}$ because they are part of a mechanism that enables people to satisfy a goal: recognizing individual conspecifics. However, as I argued, several of our controversial cases satisfy some but not all of these interpretations, and our consensual non-cases of cognition do not satisfy even the most liberal interpretation.

If SMH is an adequate ecumenical analysis of $COGNITION$, then this is precisely the relationship we should expect. It illustrates the relationship that parameterized terms (like $SENSITIVITY$) have to variations in willingness to ascribe $COGNITION$. Furthermore, variation in interpretations of the $SENSITIVITY$ parameter incorporates variation in scientists' commitments about whether cognition is representational, and which varieties of representation (or near-relatives) are necessary for cognition. Thus the $SENSITIVITY$ parameter corresponds to a well-established strategy for adjudicating disagreements about the extension of $COGNITION$.

5.3. *Managing Behavior*

Finally, cognitive mechanisms must manage the behavior of agents. A mechanism "manages" behavior just in case a pattern of behavior—i.e. certain behavior-types in some circumstances and distinct behavior-types in other circumstances—depends causally on the mechanism's functional state, assuming normal operation (the circumstances in question being the same circumstances to which cognitive mechanisms are sensitive).⁷ The management of behavior by cognitive mechanisms should be relatively causally specific (Woodward, 2010; Haueis, 2018;

⁷ I use the awkward language of causal dependence in order to allow for the Dretskean view that system-internal "causes of behavior" are in fact parts of the behavior—which is a process—rather than events that are causes of bodily motions called "behavior" (Dretske, 1988).

though a causal specificity requirement means little without a way of individuating behavior-types). Path integration in desert ants is a cognitive phenomenon in which the ant's locomotion is managed. Some locomotive movements are caused rather than others, and the difference is ascribable to the ant's circumstances (e.g. its destination being at such-and-such a bearing). By contrast, causing an irregular muscle spasm is a poor candidate for the management of behavior if it cannot be explained by appeal to any adaptive character.

The behavior condition of SMH allows for mediated management of behavior, just as the sensitivity condition allows for mediated sensitivity to circumstances. So the on-line control of grasping movements (e.g. Jeannerod, 1984) is the management of behavior, and face-recognition mechanisms also manage behavior since the recognition of a face (e.g. as Sanna's rather than Christy's) has effects that, in some situations, manifest as differences in behavior (e.g. uttering "Hi, Sanna," instead of "Hi, Christy").

In much cognitive science, a mechanism's relation to behavior is an object of attention less often than its sensitivity. Nevertheless, such a connection is presumed for any putatively cognitive process, and the discovery that a process has no effect on behavior would entail that it is not a cognitive process (this requirement echoes "consumer-based" accounts of representation, e.g. Millikan, 1984). For example, consider a honeybee hive construed as a cognitive agent (a goal-apt system per $AGENT_{\gamma}$). It was controversial for some time whether foraging honeybees are guided by the waggle dances of their sisters (Munz, 2005). Despite robust correlations between the properties of bee dances and the location of a food source relative to the hive, it was not always clear that witnessing a waggle dance affected the subsequent behavior of foragers. If it had turned out that the dances did not affect the behavior of other bees, I submit that cognitive scientists would resist the claim that waggle dances are cognitive mechanisms for honeybee hives (although the dances are still behaviors of individual bees). Rather, the dances would merely be fascinating performances that served no function in the life of the hive. To take another example, Susan Goldin-Meadow argues that gesturing can improve learning and thinking (Goldin-Meadow & Wagner, 2005). Her evidence is often cited by proponents of embodied cognition, some of whom claim that physical gesturing is itself a part of a cognitive mechanism—not only a behavior. This is because gesturing manages the successful performance of other behaviors, e.g. the performance of memory tasks or the acquisition of new skills. Thus, some external performances are parts of cognitive mechanisms since they manage other behaviors (Goldin-Meadow, 2005; Clark, 2007).

Although psychology is said to be the scientific study of behavior, explicit criteria for counting a performance as a behavior are rarely discussed. There is an ample literature distinguishing intentional action from mere behavior, but little extant discussion of what distinguishes behavior from other performances of agents, or of how behavior-types should be individuated (a notable exception is Dretske [1988], who argues convincingly that behavior is

the endogenous causing of e.g. bodily movement, rather than endogenously caused bodily movement). Self-motion (e.g. walking but not being pushed) is generally regarded as behavior, but pumping blood and healing damaged tissue are generally not regarded as behavior (at least not as the behavior of agents rather than of their parts, or at least not by cognitive scientists rather than biologists).

Since the extant discussion is limited, I will describe three flat-footed interpretations of *BEHAVIOR* designed to mirror rather than rationalize common patterns of judgment. According to *BEHAVIOR_α* a performance counts as behavior only if it is an observable moving of the body, where by “observable” I mean that the movement is of the body’s surfaces (so changes of heart rate are excluded), it is macroscopic, and it occurs on roughly animal-like time scales (so slower processes like growth are excluded). *BEHAVIOR_α* captures the common judgments that microbes and plants do not behave, except perhaps dramatic plant performances like those characteristic of a Venus flytrap. But animals behave, as do naval vessels and robots like Herbert. Our paradigmatically cognitive cases also involve the management of behavior_α. My recognition of Sanna’s face in virtue of cortical activity is a cognitive process, since linguistic performances—e.g. saying “Hi, Sanna,”—are observable movings of the body that are managed in part by face recognition mechanisms. Likewise, a typical human’s navigation of a city is an observable moving of the body.

The extension of *BEHAVIOR_α* is a proper subset of the extensions of the other two interpretations. According to *BEHAVIOR_β* a performance is behavior only if it is an observable moving of the body or it is a tropism. On this interpretation, plant phototropism counts as behavior. This interpretation mirrors the judgments of those who characterize some plant performances as behavior (e.g. presumably contributors to the journal *Plant Signaling & Behavior*, established in 2006). Calvo (Calvo Garzón, 2007; Calvo Garzón & Keijzer, 2011) freely characterizes certain plant performances as behaviors, but offers no explicit criterion distinguishing behaviors from non-behavioral performances. He does offer a representation-in-absence criterion for distinguishing “cognitive” behavior from “reactive” behavior (2007, p. 210), but this seems to be a requirement on what I call “sensitivity,” rather than a criterion for distinguishing behaviors from non-behaviors.

Evan Thompson and others (Lyon, 2006; van Duijn et al., 2006; Thompson, 2010) have suggested that all (or almost all) living organisms have cognitive processes, so they must consider at least some microbial performances to be behavior. According to *BEHAVIOR_γ* a performance is behavior only if it is an observable moving of the body or it is a taxis at any time- and length-scale. On this interpretation, bacterial magnetotaxis counts as behavior.

Bacterial magnetotaxis and plant phototropism fall in the gap between the extensions of the more conservative *BEHAVIOR_α* and the more liberal extensions of *BEHAVIOR_{β-γ}*, which is appropriate for controversial cases of cognition. This set of three interpretations for *BEHAVIOR*

Table 2: Selected variation in the sectarian extension of COGNITION according to SMH.

Interpretations of parameters:	Sectarian extension of COGNITION includes:
AGENT _α SENSITIVITY _α BEHAVIOR _α	face-recognition in typical humans urban navigation in typical humans
AGENT _β SENSITIVITY _β BEHAVIOR _α	face-recognition in typical humans urban navigation in typical humans extended cognition in Otto-notebook system
AGENT _α SENSITIVITY _γ BEHAVIOR _α	face-recognition in typical humans urban navigation in typical humans embodied audition in crickets
AGENT _γ SENSITIVITY _β BEHAVIOR _α	face-recognition in typical humans urban navigation in typical humans robot cognition in Herbert extended cognition in Otto-notebook system
AGENT _β SENSITIVITY _γ BEHAVIOR _α	face-recognition in typical humans urban navigation in typical humans extended cognition in Otto-notebook system embodied audition in crickets
AGENT _α SENSITIVITY _γ BEHAVIOR _β	face-recognition in typical humans urban navigation in typical humans extended cognition in Otto-notebook system embodied audition in crickets off-line phototropism in <i>L. cretica</i>
AGENT _γ SENSITIVITY _γ BEHAVIOR _γ	face-recognition in typical humans urban navigation in typical humans distributed cognition in maritime navigation extended cognition in Otto-notebook robot cognition in Herbert embodied audition in crickets magnetotaxis in bacteria

is disappointing with regard to analytical clarity, but modular analyses aim to articulate seams of disagreement, and so much the better for the analysis if some of those seams have not yet received the attention they merit. If SMH is correct, then philosophers interested in vindicating or denying the possibility of plant and microbe cognition might fruitfully attend to the question of what behavior is, rather than focusing so exclusively on representation.

6. Applying the Sensitive Management Hypothesis

The sensitive management hypothesis is a modular analysis of *COGNITION*. It classifies each of the test cases correctly using the toy parameter values I discussed, including the consensually cognitive phenomena in every sectarian extension, excluding consensually non-cognitive phenomena from every sectarian extension, and including each of the controversially cognitive phenomena in some sectarian extensions but not others. See Table 2 for examples of variation in the extension of *COGNITION* based on different sets of interpretations for parameters; note that the consensual cases of face-recognition and unassisted navigation are in every sectarian extension, and consensual non-cases of cognition are in no sectarian extension. Thus, I have motivated the sensitive management hypothesis as a *prima facie* adequate analysis proper of the scientific concept of *COGNITION*, one that leaves open the border war disputes that cognitive scientists have not yet resolved.

Many authors have suggested that there is no need for a definition or analysis of *COGNITION* (e.g. Allen, 2017; Ramsey, 2017; Chemero, 2009, 212n8; Clark, 2010b, 62). Carol Cleland (2012) argues that in chemistry and biology, good definitions follow open-minded inquiry rather than preceding it. I agree that that an analysis is not a prerequisite for responsible cognitive scientific inquiry, and doubt that a compelling analysis will resolve disputes in the cognition border wars. However, an analysis can serve other kinds of purposes. I contend that SMH serves the three purposes I identified in Section 1 for ecumenical characterizations of *COGNITION*.

First, the sensitive management hypothesis makes explicit some ways cognitive scientists think about their object of study and perspicuously represents areas of agreement and disagreement. For example, it makes explicit that cognitive scientists explain by identifying and describing mechanisms (in the anodyne sense described in Section 4), that cognitive mechanisms are ascribed relative to an agent or containing system, and that mechanisms are considered cognitive only if their operations are sensitive to an agent's circumstances and they regulate an agent's behavior. The value of conceptual analysis generally consists in articulating the relations between the analyzed concept and a system of theoretical concepts; SMH makes claims about the relations between *COGNITION* and other concepts in the philosophy of cognitive science, such as *MECHANISM*, *AGENT/REPRESENTATIONAL SUBJECT*,

REPRESENTATION, and BEHAVIOR. These relations are less strictly defined than some more classical conceptions of cognition (e.g. the classical computational theory of cognition; Ramsey, 2007), but this is appropriate since cognitive science is less theoretically unified than it once was. And SMH represents disagreement by treating open questions as unsettled. Its parameters correspond to matters on which cognitive scientists provide conflicting judgments.

Second, the sensitive management hypothesis could clarify the concept of cognition in classrooms and public communication. Introductory science textbooks often begin with “definitions” of their subject matter that, while helpful for students, are philosophically unsatisfactory, and cognitive science introductions often invoke familiar slogans like “the mind is the software of the brain” or “the mind is what the brain does.” SMH can fulfill this sort of role, facilitated by its abbreviated slogan that “cognition is the sensitive management of an agent’s behavior,” which is less susceptible to obvious counterexamples than the familiar alternatives and more specific than others like “the function of the brain is sensorimotor processing.” My suggestion is not, of course, that the sensitive management view be described to novices along with a lengthy commentary about the values of its parameters. Rather, even without such commentary SMH is a less hand-wavy characterization of cognition, and one that scaffolds further understanding by incorporating reference to mechanisms, behavior, agents, and representation (i.e. sensitivity to circumstances). For more advanced students, examining the possible interpretations for the parameters of SMH is a scaffold for deeper understanding of extant theoretical disagreements.

Third, when the sensitive management hypothesis is considered along with its modular logical apparatus (i.e. including well-specified domains for licit parameters), it is a precise tool for representing various background commitments of cognitive scientists. This precision makes it valuable as a resource for appeal in philosophical arguments that draw on facts about cognitive scientific practice. Such appeals are common in disputes about the mind-body problem, multiple realizability, reductionism, and other philosophical topics. At present, philosophers often rely on functionalist formalisms, despite the fact that functionalism is known to have shortcomings as an expressive resource for empirical claims (Bechtel & Mundale, 1999; Shagrir, 2005; Sprevak, 2009). SMH may offer some fresh resources for engaging in old debates, especially since the values of parameters can be tailored for different argumentative or rhetorical purposes.

A further virtue of SMH is that it can be improved by increments. Philosophers and cognitive scientists can augment the set of toy interpretations I describe above. Disputes over the better and worse interpretations of one parameter are largely independent of similar disputes over the other parameters. Since parameters represent terms with determinable extensions, they will often correspond to strategies for adjudicating disputes about the extension of COGNITION. Some of these strategies may be well-worn, like the strategy of adjudicating

disputes about the extension of *COGNITION* by appealing to various conceptions of representation. However, a modular analysis may demonstrate its value by drawing attention to concepts like *BEHAVIOR* whose analysis might prove fruitful, but that have so far received little attention.

7. Conclusion

While many sciences feature disagreement regarding the nature of their object of inquiry (e.g. life, the physical, the chemical), cognitive science is unusual in that the extension of *COGNITION* is highly controversial. This fact makes cognitive science unusually difficult to understand and reason about—for students, laypeople, humanist scholars of cognitive science, and cognitive scientists themselves. These difficulties might be eased by an analysis of the cognitive scientist's inclusive concept of *COGNITION* that, unlike extant “marks of the cognitive,” aims to clearly represent controversies over the extension of *COGNITION* rather than to resolve them. The sensitive management hypothesis is such an analysis. Since it is modular, its extension can vary according to the various licit interpretations for its parameters. Using a restricted set of simple interpretations, I illustrated how the analysis handles a variety of test cases, and made a *prima facie* argument that the analysis proper is extensionally adequate, in that with a toy set of parameter values its sectarian extensions always include the paradigmatic test cases of cognition, never include consensual non-cases of cognition, and selectively include or exclude controversial cases of cognition depending on which set of interpretations is assigned.

Acknowledgments

I am grateful for generous feedback from many people, including Joseph McCaffrey, Robert Brandom, Edouard Machery, Alistair Isaac, William Bechtel, three anonymous reviewers, and colleagues at the University of Pittsburgh, the University of Edinburgh, and University College Dublin. A version of this paper appears in my PhD dissertation, “Cognition in Practice: Conceptual Development and Disagreement in Cognitive Science” (University of Pittsburgh, 2016). I received financial support through a Junior Faculty Summer Research Grant at Texas Christian University, and the Wesley C. Salmon Fund at the University of Pittsburgh.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Mikio Akagi is Assistant Professor of the History and Philosophy of Science in the John V. Roach Honors College at Texas Christian University.

References

- Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*, 43–64.
- Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition*. Blackwell.
- Adams, F., & Garrison, R. (2013). The mark of the cognitive. *Minds and Machines*, *23*, 339–352.
- Aizawa, K. (2017). Cognition and behavior. *Synthese*, *194*, 4269–4288.
- Akagi, M. (2018). Rethinking the problem of cognition. *Synthese*, *195*, 3547–3570.
- Allen, C. (2017). On (not) defining cognition. *Synthese*, *194*, 4233–4249.
- Bechtel, W. (2008). Mechanisms in cognitive psychology: what are the operations? *Philosophy of Science*, *75*, 983–994.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 421–441.
- Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: linking cognitive and neural states. *Philosophy of Science*, *66*, 175–207.
- Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.
- Brooks, R. A., Connell, J. H., & Ning, P. (1988). Herbert: a second generation mobile robot. *A.I. Memos*, *1016*, 0–10.
- Buckner, C. (2015). A property cluster theory of cognition. *Philosophical Psychology*, *28*, 307–336.
- Calvo Garzón, F. (2007). The quest for cognition in plant neurobiology. *Plant Signaling & Behavior*, *2*, 208–211.
- Calvo Garzón, P., & Keijzer, F. (2011). Plants: adaptive behavior, root-brains, and minimal cognition. *Adaptive Behavior*, *19*, 155–171.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press.
- Chirimuuta, M. (2014). Minimal models and canonical neural computations: the distinctness of computational explanation in neuroscience. *Synthese*, *191*, 127–153.
- Clark, A. (2005). Intrinsic content, active memory, and the extended mind. *Analysis*, *65*, 1–11.
- Clark, A. (2007). Curing cognitive hiccups: A defense of the extended mind. *Journal of Philosophy*, *104*, 163–192.
- Clark, A. (2010a). Coupling, constitution, and the cognitive kind: a reply to Adams and Aizawa. In R. Menary (Ed.), *The Extended Mind* (pp. 81–100). Bradford.
- Clark, A. (2010b). Memento's revenge: the extended mind, extended. In R. Menary (Ed.), *The Extended Mind* (pp. 43–66). Bradford.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.
- Cleland, C. E. (2012). Life without definitions. *Synthese*, *185*, 125–144.
- Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*, 355–376.
- Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press.

- Cummins, R. (1975). Functional analysis. *The Journal of Philosophy*, 72, 741–765.
- Cummins, R. (2000). ‘How does it work?’ vs. ‘what are the laws?’: two conceptions of psychological explanation. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and Cognition* (pp. 117–144). MIT Press.
- Danks, D. (2014). *Unifying the Mind: Cognitive Representations as Graphical Models*. MIT Press.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press.
- Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, Content and Function* (pp. 17–36). Oxford University Press.
- Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.
- Earman, J. (1975). What is physicalism? *The Journal of Philosophy*, 72, 565–567.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Lawrence Erlbaum.
- Goldin-Meadow, S. (2005). *Hearing Gesture: How Our Hands Help Us Think*. Harvard University Press.
- Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, 9, 234–241.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389–400.
- Haueis, P. (2018). Beyond cognitive myopia: a patchwork approach to the concept of neural function. *Synthese*, 195(12), 5373–5402.
- Haugeland, J. (1991). Representational genera. In W. Ramsey, S. P. Stich, & D. E. Rumelhart (Eds.), *Philosophy and Connectionist Theory* (pp. 61–89). Lawrence Erlbaum.
- Hochstein, E. (2012). Minds, models and mechanisms: A new perspective on intentional psychology. *Journal of Experimental and Theoretical Artificial Intelligence*, 24, 547–557.
- Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.
- Jackson, F. (1998). *From Metaphysics to Ethics*. Oxford University Press.
- Jeannerod, M. (1984). The timing of natural prehension movements. *Journal of Motor Behavior*, 16, 235–254.
- Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, 107, 11163–11170.
- Kornblith, H. (2017). A naturalistic methodology. In G. D’Oro & S. Overgaard (Eds.), *The Cambridge Companion to Philosophical Methodology* (pp. 141–160). Cambridge University Press.
- Lyon, P. (2006). The biogenic approach to cognition. *Cognitive Processing*, 7, 11–29.
- Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Machery, E. (2012). Why I stopped worrying about the definition of life... and why you should as well. *Synthese*, 185, 145–164.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.
- Millikan, R. G. (2017). *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford University Press.
- Montero, B. (1999). The body problem. *Noûs*, 33, 183–200.
- Munz, T. (2005). The bee battles: Karl von Frisch, Adrian Wenner and the honey bee dance language controversy. *Journal of the History of Biology*, 38, 535–570.
- Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.

- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–175.
- Putnam, H. (1975). The mental life of some machines. In *Philosophical Papers, Vol. 2: Mind, Language, and Reality* (pp. 408–428). Cambridge University Press.
- Ramsey, W. M. (2007). *Representation Reconsidered*. Cambridge University Press.
- Ramsey, W. M. (2017). Must cognition be representational? *Synthese*, 194, 4197–4214.
- Ross, L. N. (2015). Dynamical models and explanation in neuroscience. *Philosophy of Science*, 82, 32–54.
- Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology*, 22, 1–19.
- Rowlands, M. (2010). *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. MIT Press.
- Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy*, 51, 389–428.
- Rupert, R. (2009). *Cognitive Systems and the Extended Mind*. Oxford University Press.
- Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3, 417–457.
- Segundo-Ortin, M., & Calvo, P. (2019). Are plants cognitive? A reply to Adams. *Studies in History and Philosophy of Science Part A*, 73, 64–71.
- Shagrir, O. (2005). The rise and fall of computational functionalism. In Y. Ben-Menahem (Ed.), *Contemporary Philosophy in Focus: Hilary Putnam* (pp. 220–250). Cambridge University Press.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656.
- Shapiro, L. A. (2013). Dynamics and cognition. *Minds and Machines*, 23, 353–375.
- Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science*, 80, 958–970.
- Sprevak, M. (2009). Extended cognition and functionalism. *The Journal of Philosophy*, 106, 503–527.
- Stinson, C. (2016). Mechanisms in psychology: Ripping nature at its seams. *Synthese*, 193, 1585–1614.
- Thompson, E. (2010). *Mind in Life: Biology, Phenomenology, and the Sciences of the Mind*. Harvard University Press.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433–460.
- van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of minimal cognition: casting cognition as sensorimotor coordination. *Adaptive Behavior*, 14, 157–170.
- van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, 92, 345–381.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21, 615–665.
- Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- Wakefield, J. C. (1992). The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist*, 47, 373–388.
- Webb, B. (1994). Robotic experiments in cricket phonotaxis. In D. Cliff, P. Husbands, J.-A. Meyer, & S. W. Wilson (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior* (pp. 45–54). MIT Press.

Weiskopf, D. A. (2011). Models and mechanisms in psychological explanation. *Synthese*, 183, 313–338.

Woodward, J. (2010). Causation in biology: Stability, specificity and the choice of levels of explanation. *Biology and Philosophy*, 25, 287–318.