COGNITION IN NATURE:
INFORMATION, EXPLANATION, AND EMBODIMENT

Ben Baker

A DISSERTATION

in

Philosophy

Presented to the Faculties of the University of Pennsylvania

In

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Co-Supervisor of Dissertation

_____

Gary Hatfield

Adam Seybert Professor in
Moral and Intellectual Philosophy,
Director, Visual Studies Program

Co-Supervisor of Dissertation

_____

Lisa Miracchi

Assistant Professor of Philosophy

Graduate Group Chairperson

Errol Lord, Professor and Chair of Philosophy

_____

Dissertation Committee

Gary Hatfield
Adam Seybert Professor in
Moral and Intellectual Philosophy

Lisa Miracchi
Assistant Professor
Philosophy

Daniel Singer
Assistant Professor
Philosophy

Dedication

For Martha, Sheela, and Kavya.

Acknowledgement

Thank you to my dissertation supervisors, Gary Hatfield and Lisa Miracchi, whose guidance has been vital to this project and my intellectual development broadly.

Thank you to friends and colleagues who helped me at various stages to explore the ideas that went into this project, including Grace Boey, Daniel Burnston, Clarissa Busch, Devin Curry, Martha Farah, Hilary Gerstein, Kate-Nicole Hoffman, Javier Gomez-Lavin, Max Lewis, Daniel Koditschek, Ben Lansdell, Artemis Panagopoulou, Sonia Roberts, Tiina Rosenqvist, Daniel Singer, Jordan Taylor, Eugene Vaynberg, Yosef Washington, and Younbin Yoon.

For supporting me through the doctorate and the dissertation, thank you to Anna Baker, Justin Bernstein, Rebecca Holsen, Rohan Keshwara, Dylan Manson, Raj Patel, Sima Patel, Ian Peebles, Pierce Randall, Brian Reese, and Fabian Schellhaas.

Abstract

COGNITION IN NATURE:
INFORMATION, EXPLANATION, AND EMBODIMENT

Ben Baker

Gary Hatfield, Lisa Miracchi

THIS DISSERTATION ADVANCES A NOVEL VIEW ABOUT HOW TO UNDERSTAND COGNITION AS A PHENOMENON ARISING FROM THE COORDINATION OF BRAIN, BODY, AND ENVIRONMENT. THE PROJECT STARTS BY ARTICULATING A QUESTION ABOUT HOW TO ACCOUNT FOR THE *INTENTIONAL* NATURE OF COGNITION IN A WORLD COMPOSED OF THE UNINTENTIONAL PROCESSES OF NEUROPHYSIOLOGY, CHEMISTRY, AND PHYSICS. IN CHAPTER 1 I CONTEXTUALIZE AND PREVIEW THE BROAD SORT OF ANSWER I WILL DEVELOP, AND THEN EACH OF THE PROCEEDING THREE CHAPTERS ADDRESS A MAJOR PART OF THIS PUZZLE.

CHAPTER 2 DEFENDS A VIEW OF NATURAL INFORMATION AS FACTIVE, THAT IS, ALWAYS INDICATING MATTERS OF FACT. I OBJECT TO A RECENTLY COMMON LINE OF REASONING IN FAVOR OF A NON-FACTIVE VIEW. I ARGUE THE REASONING COMMONLY OFFERED IN FAVOR THE NON-FACTIVE APPROACH CONFLATES TWO DIFFERENT PROBLEMS TO DO WITH INFORMATION AND INTENTIONALITY, ONE OF WHICH DOES NOT ACTUALLY WEIGH IN FAVOR OF NON-FACTIVITY, AND THE OTHER OF WHICH IS NOT RESOLVABLE JUST IN INFORMATION-THEORETIC TERMS. IN

CHAPTER 3, I DEFEND A VIEW OF THE EXPLANATORY SENSE IN WHICH SYSTEMS OF DYNAMIC RELATIONS "GIVE RISE TO" COGNITIVE PROCESSES. I ILLUSTRATE THE RELEVANT FEATURES OF DYNAMICAL SYSTEMS MODELS, AND I DESCRIBE CONSTITUTIVISM – A WIDESPREAD VIEW OF THE EXPLANATORY RELATION IN QUESTION. I ARGUE CONSTITUTIVISM DOES NOT PROVIDE A PLAUSIBLE ACCOUNT OF LEVELS OF ORDER IN CERTAIN DYNAMICAL SYSTEMS. CONSTITUTIVISM MAKES A COMMITMENT ABOUT THE LOWER-ORDER COMPONENTS BEING CONTAINED BY THE HIGHER-ORDER PHENOMENA THEY HELP EXPLAIN, AND DYNAMICAL LOWER-ORDER COMPONENTS NEED NOT BE CONTAINED IN THIS WAY. I SUPPORT AN ALTERNATIVE, GENERATIVE, DYNAMICAL VIEW OF THE EXPLANATORY BASIS OF COGNITION. IN CHAPTER 4, I ADVANCE A VIEW OF COGNITION AS EMBODIED. I ARGUE THAT THE EXAMINATION OF THE EXPLANATORY ROLE OF THE BODY HAS BEEN LIMITED BY ITS FOCUS ON PERCEPTION AND ACTION, AND THAT THE BODY DESERVES MORE THEORETICAL ATTENTION. TOWARD THIS END, I ARTICULATE AN ORIGINAL, FORMAL NOTION OF THE RELEVANT KIND OF BODY, DESCRIBED AS A SPECIFIC KIND OF DYNAMICALLY GENERATED ORDER. IN SUM, I EXPAND ON THE UNDERSTANDING OF COGNITION AS EMBODIED IN A WAY THAT CENTERS ON THE BODY ITSELF.

Table of Contents

List of Illustrations

1 Preamble

Somehow there exist in nature entities that, in a proprietary way, *recognize*, *recall*, *aim for, anticipate*, and *imagine* things. Such cognitive processes exhibit intentionality; they are "directed at" or "about" things in a way that is characteristic of the mental (Brentano, 1874; Jacob, 2019). Intentional relations differ starkly from the relations described by physical and chemical sciences. They can involve motives, mistakes, and make-believe – all at once, in fact, as in when a child pretends to read, holding the book upside down. One is apt to wonder how or even whether the mindless, concrete realm of these sciences and the intentional, perspectival realm of cognition can be fit into a single image of nature. Insofar as cognitive science and neuroscience investigate the same world, and insofar as cognitive neuroscience is an intelligible field of research, there must be some sense in which the investigation of thinking processes coheres with the investigation of the flesh-and-blood world we put under a microscope. This dissertation aims to contribute to a rich tradition of philosophical work striving to make sense of the way vivid and value-laden, intentional phenomena arise naturally in the organization of certain living animal bodies. There is a complex assortment of theories that have been commonly put toward this end, including theories about information-processing, computation, representation, dynamical systems, and the relationship between perception and action. This project critically evaluates and expands on work on these topics in order to provide a view of what cognition in nature basically involves. Specifically, this work aims to reveal an

underappreciated, foundational role that the whole body ought to play in our theoretical understanding of cognition.

A notion often put to fundamental work in this domain is that of *information*. Brains and bodies carry information about the world outside them, and one might hope to understand the "aboutness" of thinking partly in terms of the "aboutness" of informational relations. Dretske (1981) offered an approach along these lines that figures significantly in the background of many of the views I will discuss. For Dretske, information is communicated only about matters of fact and so cannot be mistaken or misleading, yet we know that thought processes can indeed be led astray, so it is a puzzle how the latter could derive from the former. This puzzle is at the forefront in Chapter 2. To jump ahead to part of the conclusion there, this puzzle is not to be solved in information-theoretic terms alone. Finding that some neural activity, *N*, carries information about some fact, *F*, says very little, on its own, about what sort of cognition might be occurring. Information transmission is occurring literally all over the place and at all times, so some systematic account is needed to describe how the relevant information is received and processed.

One option, and another aspect of the theoretical landscape in the background here, is to analyze the relevant information-processing in terms of symbolic operations in the brain, carried out according to a syntax inherent in the structure of our nervous systems (Fodor, 1980, 1975, 1987; McCarthy & Hayes, 1969; Pylyshyn, 1984; Simon & Newell, 1971). On this view, there is an instructive analogy between the way the inner states of the laptop I am writing on are meaningful with respect to my keystrokes, and the way the inner states of my brain are meaningful with respect to the objects I am perceiving and thinking about. (I discuss this symbolic view in more detail in Chapter 4). A major problem for this

sort of approach is that there seem to be deep differences between the functional structure of cognitive systems and the structure of any symbolic computational systems we can come up with, and these do not seem to be differences that would be resolved in terms of bigger, more complicated processes of symbolic computation (Dennett, 1984; Dreyfus, 1992; Fodor, 2000; Harnad, 1990; Searle, 1980; R. Wilson, 2004). One way to put the worry is that a symbolic, syntactic system is semantically cut off from whatever physical processes might embody the system – the system is *dis*embodied (Barsalou, 2008; Chemero, 2011; Clark, 1998; Gallagher, 2005; Glenberg, 1997; Johnson, 2017; Shapiro, 2019; Varela et al., 1991; Wheeler, 2005; M. Wilson, 2002). If this line of thinking is right, then the computer analogy does more to delude than instruct, and we need a different framework for thinking about cognition as part of nature.

Setting aside concerns about symbols and embodiment for a moment, consider the idea that cognitive processes occur by way of *mechanisms that use information* carried by the nervous system to achieve particular outcomes, where the functioning of the whole mechanisms can be understood in terms of the coordinated functioning of its components (Baumgartner & Gebharter, 2016; Bechtel & Abrahamsen, 2005; Couch, 2011; Craver, 2007; Harbecke, 2015; Machamer et al., 2000)**.** This formulation adds to the notion of information-processing the idea of a characteristic method of achieving a certain result, and the idea that lower-level processes together give rise to cognition. Further, this suggestion does not require that these lower-level processes be symbolic computations. For example, take the connectionist models, which, for a time, were defended as alternatives to symbolic computation (Dawson, 1998; Hatfield, 1991; Smolensky, 1988), or other sorts of neural network models (Cichy & Kaiser, 2019; Hintze et al., 2017). However exactly

these models relate to symbolic computation, they arguably do describe information-processing mechanisms. I take this general idea of multi-level, information-using mechanisms to express a widely held view of the sense in which cognition happens by way of the brain and its interactions with its surroundings. However, neatly unpacking the idea of "information-using mechanisms" is not straightforward. Part of my overall project here is to reveal certain inadequacies with the understanding cognition as arising in this mechanistic fashion, and to point a way toward a model that can better guide the future science of cognition.

Many of the critics of a symbolic, computational view of cognition, who endeavor to shed light on the *embodiment* of cognition, endorse some form of dynamical systems modeling as a part of their approach (Beer, 1995; Chemero, 2011; Gelder & Port, 1995; Hurley, 2001; Kelso, 1995; Kelso et al., 2013; M. Lewis, 2005; Schöner, 2008; Shapiro, 2019; Spivey, 2008; Warren, 2006). Somewhat paradoxically, a key insight of this line of work has to do with how processes *outside* of the body contribute to cognition. This presents a source of tension for thinking about cognition in terms of constitutive mechanisms. One would probably have thought the location of cognitive mechanisms would be largely confined to the head or body, but mechanisms are partly located wherever their constitutive components are, which appears to extend far outside the body. Extant literature, so far as I find, has not clearly articulated what I think is the right way to untangle this matter. Others have either ignored the tension I just described, argued that dynamical explanations should be given a mechanistic interpretation, or accept the conclusion that cognitive mechanisms are broadly extended outside of what we canonically identify as a body (Bechtel & Abrahamsen, 2013; Clark, 2008; Kaplan & Craver, 2011; Noë, 2005; R.

A. Wilson, 2014). I argue for the alternative I favor in Chapter 3. I contend that we need a non-constitutive understanding of the relationship between levels of order in dynamical systems to make sense of embodiment, and that the notion of a generative relation, due to Miracchi (2017), meets this theoretical need.

So, my project here broadly supports and builds on an effort to reveal the dynamic and embodied nature of cognition. A central idea within this family of thinking is to foreground an essential interdependence that exists between processes of perception and processes of intentional action and, in light of this, to suggest ways of explaining how cognitive processes arise in perceptual-motor engagement with the world (Clark, 1998; Hurley, 2001; Thelen & Smith, 1994; Varela et al., 1991). This brings the body into our understanding of cognition, because perceptual-motor capacities are determined by the structure of the body. However, this sort of thesis about perception and action does not tell us precisely what we mean by "the body" itself, and so does not explore whether that body is relevant to understanding cognition in a way beyond its contribution to the structure of perceptual-motor capacities. I propose, in Chapter 4, to attend to the body in a more general sense than as a perceptual-motor machine. Drawing on basic principles from dynamical systems modeling that will already have figured in the discussion, I articulate an elementary account of a "Body" as a self-organizing boundary that meets certain formal conditions. These formal conditions involve characteristic asymmetries between parts of the system on either side of the dynamically generated boundary, and characteristic transactions across this boundary. This Body is a theoretical construct designed to figure in a larger view of perception, action, and cognition as embodied. My account thus means to develop the foundations of this view of cognition's embodied, dynamic nature – to

extend of this line of thinking at its roots, so to speak, providing for a deeper sense of cognition as embodied.

At the end of the dissertation, I will not have answered how specific, intentional contents are determined by specific, brain-body-environment relations. So, for instance, if we have competing descriptions of the information that figures in a cognitive or sub-cognitive process, my account does not say precisely how to settle the dispute. But I hope that it serves to reorient the discourse so that it is headed toward a philosophically well-founded, empirically tractable notion of how cognition arises from the brain, body, and environment.

Here is the plan: In Chapter 2 I argue for a view of Natural Information. I refine a Dretskean picture, wherein Natural Information is Factive, in opposition to a strand of non-Factive, probabilistic or correlational views of the information content relevant for understanding the workings of a cognitive systems. I show how the reasoning behind the non-Factive approaches conflates different shortcomings of a simple, Dretskean view. I propose one problem can be addressed by relying on a more precise understanding of natural laws and law-like invariance than Dretske did, while the other problem – the reference class problem – is not something to be solved in terms of a theory of information. In Chapter 3, I defend a view of the sense in which dynamic relations among brain, body, and environment together "give rise to" cognitive processes. In other words, I account for the relation between a higher-level phenomenon and its *explanatory basis*, in dynamical models in particular. I spell out the relevant aspects of dynamical systems models, and spell out a mechanistic, Constitutivist view of the explanatory basing relation, which I take to be widespread. I argue that Constitutivism fails to accommodate the way lower-order

components can be more global than then phenomena they give rise to, in dynamical systems. I demonstrate this issue of locality in detail and conclude, on its basis, in favor of understanding the explanatory relation in question as one of dynamic generation. In Chapter 4, I advance a view of cognition as embodied in the basis sense of arising with respect to a particular body, and toward this end I offer an original, formal notion of the relevant kind of Body. I contextualize and motivate this account by distinguishing important shared insights and disagreements among other approaches to the embodiment of cognition. I argue that the examination of the explanatory role of the body has been limited by its focus on perception and action, and work to develop deeper view of what the embodiment of cognition involves.

2 Natural Information, Factivity, and Nomicity

1. Introduction

Natural information is information that can be found in the observable world – something carried by objective events or properties in the world, or relations among them. I take it to be generally agreed that cognition is to be understood partly in terms of the way certain systems process or pick up natural information. Natural information is to be distinguished from information as represented mentally or linguistically, both of which involve the kind of intentionality that natural information is invoked to help explain. Current philosophical discourse about natural information is heavily indebted to Dretske's (1981) account, however it has been widely noted that his view runs into some problems. Specifically, this account is alleged to be *too strict* and to face a *reference class* problem. Part of Dretske's proposal was that Natural Information is only ever carried about actual events – about facts – and not about probability distributions among events. That is, Natural Information is Factive. A common thought about how we ought to depart from a Dretskean view of information is to hold that Natural Information is Non-Factive, in part as a response to worries about the reference class problem and strictness. I argue that this approach is misguided. I argue the move to Non-Factive information does not resolve the reference class problem and that the objectionably strict nature of a Dretskean view can be remedied

without giving up on Factivity. Rather, Dretske's claim that informational relations must be based in laws of nature is the culprit when it comes to strictness. Drawing on a more nuanced and far-reaching understanding of "lawlike," that is, Nomic relations as information-bearing, I argue for a Factive and Nomic view of the kind of Natural Information that plays an important role in perceptual and cognitive processes.

I proceed as follows: in section 2, I outline a basic, Dretskean picture of Natural Information, which understands informational relations to be Factive and based in laws of nature. I then describe several approaches according to which information is non-Factive, noting that this approach is motivated partly by the observation that information-users (whose behavior we ultimately hope to better understand) sometimes are led into *error* by their information-carrying states. In section 3 I turn to take a closer look at how we can properly understand talk of "laws of nature" and "Nomic" regularities in nature. I argue that an appropriately nuanced understanding of Nomicity should replace Dretske's more blunt appeal to laws of nature as a way of distinguishing information-carrying covariance from accidental covariance. I show that all parties to this discourse agree that such a distinction is necessary, but others have not recognized that being more precise about Nomicity addresses the main problem with the Dretskean account of information. In section 4, try to show that the critical reaction toward a non-Factive notion of information is based on a faulty line of reasoning; cases of information-users making errors do not support the conclusion that information is non-Factive. I conclude by offering a definition of Natural Information as both Factive and Nomic, and by clarifying how Factive information involves uncertainty and probability.

2. Dretskean Thesis and the Anti-Factivity Response

2A Factivity and Lawfulness

Dretske (1981) developes an account of how information about the environment flows it into the brain of perceivers and knowers, aiming to provide an understanding how perception and knowledge arise in the natural world. This notion of information expanded on the notion of information developed by Claude Shannon and Warren Weaver (1949). The Shannon-Weaver model essentially provides mathematical means for calculating stochastic entropy, which has deep ties to the considerably older, thermodynamic notion of entropy. The entropy of information theory is usefully glossed as a measure of "uncertainty reduction." (Barwise & Seligman, 1997; Kugler & Turvey, 1987). Consider the sense in which the result of a standard coin-flip measurably *reduces uncertainty* (from two possibilities to one) in a lesser degree than the result of tossing a six-sided die does (from six possibilities to one); this is the sense in which the Shannon-Weaver model says the former result is less informative than the latter. Grasping the math of their model (why information is measured logarithmically) is not important for this discussion. The philosophical questions to do with understanding information as *Natural* are beyond the scope of the Shannon-Weaver model. Since Shannon-Weaver information is purely formal, it does not describe a kind of information that is, per se, to be found in the world and not just in our thoughts and speech. Dretske offered a theory of information flow as a natural phenomenon, hoping to show how it gives rise to intentionality. (Going forward I will start to simply refer to "information" without the "natural" qualification, trusting my meaning will be clear from context).

The Dretskean view of information-carrying signals involves two important conditions, **Factivity** and **Lawfulness**. Factivity can be concisely expressed using Dretske's favored term; *indication*. Dretske's thinking here also parallels Grice's (1957) notion of "natural meaning." A signal carrying information about *F*-ness *indicates F*-ness. Smoke is an informational signal of fire insofar as it indicates a fire. By contrast, linguistic utterances and conventional signs are non-Factive. The utterance "there is a fire" and a ringing fire-alarm may not indicate a fire; the speaker of the utterance could be mistaken or lying, and the alarm could have been triggered by a prankster or an electrical malfunction. In such misleading cases the Dretskean view says that the utterance and alarm do not carry information about a fire. In Dretske's words, "*false* information and *mis*-information are not kinds of information—any more than decoy ducks and rubber ducks are kinds of ducks" (1981, 45). Of course, we sometimes speak of false information or mis-information, as in "he informed me that there was a fire even though there wasn't one," but that would be to use a non-Natural sense of "information." The contrast between the Factivity of information and the non-Factivity of intentional states is what sets up *the* central philosophical challenge for Dretske's information-based approach to intentionality. He saw the puzzle as one of accounting for how a Factive, brain-world relationship could be the basis of a non-Factive, mind-world relationship (Dretske, 1986). That puzzle is not my focus here though – here I am just concerned with what Natural Information is. So, I will work from the following statement of the Dretskean Thesis (DT) on information, which slightly rewords what can be found on pages 65 and 76-77 on Dretske (1981):

> **DT**: A signal, *b*'s being *N*, carries the information that *s* is *F* if and only if (i) the conditional probability of *s*'s being *F*, given that *b* is *N*, is 1, and (ii) this conditional probability relation is fixed by some law(s) of nature.

Some part or process in nature, *s*, is a *source* of information, which can manifest various states, among which one is *F*. For example, *s* might be particular part of space-time that can either exhibit a fire, *F*, or a non-fire, *G*. The information-carrying" event, *b*'s being *N*, is also a source of information itself, but carries information about *s*'s being *F* if conditions (i) and (ii) are met, according to DT. For example, *b* might be part of space-time that can manifest smoke, *N*, or non-smoke, *O*. More relevantly, *b* might be a part of the brain that can manifest a particular pattern of neural activity *N*, or a different pattern *O*.

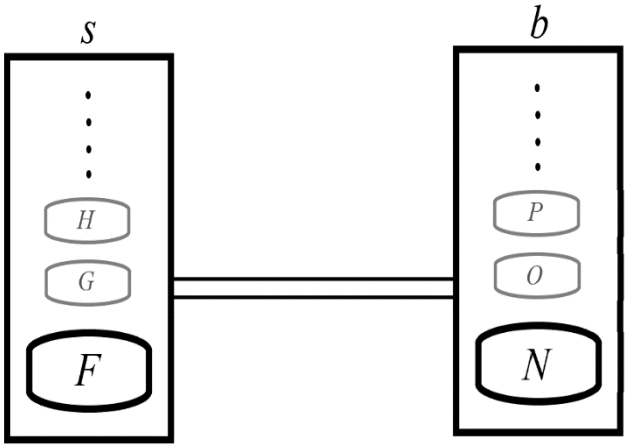

Figure 1

Diagram of information transmission: *b*'s being *N* carries information about *s*'s being *F*. Smaller, grey boxes and vertical ellipses represent possible counterfactual manifestations of *s* and *b*. The double-line between *s* and *b* represents the non-accidental relation between them; according to DT, it represents a constraint of some law(s) of nature. (The two sets of possible events are also here depicted as similarly sized and proximately connected but that need not characterize information transmission).

Factivity is entailed by DT's condition (i), in the required probability of 1. If *N* occurs and carries the information that *s* is *F*, then (with probability 1) *s* is *F*. Since a signal is to be construed as one among a definite set of possibilities, it is more precise to say "*b*'s being *N*" carries information, and somewhat misleading to say "*N*" (all by itself) carries information, but I will sometimes use the less precise phrasing, for brevity. Condition (ii) in DT expresses Lawfulness. More generally, (ii) is a version of a *non-accident* condition. That is, Lawfulness offers a way to account for the idea that there are some mere coincidences; accidental correlations that are not informative in the relevant sense. To illustrate, suppose that there is an ordinary coin that, when tossed many times in an ordinary way, happens to land heads on every single toss. Call this coin

"s," and use "is *F*" for the predicate "is tossed," call the event of this coin landing "*b*," and use "is *N*" for the predicate "lands heads." In such a case there is an observable, perfect correlation between the coin being tossed and the coin landing heads. This is an example of an accidental correlation because, insofar as it involves fair coin tosses, the correlation does not reflect the actual "fairness" of the coin – what one might call the natural disposition or physical tendency of such a coin's behavior. A Dretskean view holds that accidental correlations do not reflect the information-carrying properties of events in the world. The reference to "conditional probability" in DT is meant in the sense in which one would say that the probability of a fair coin landing heads is .5, conditional on it being tossed in the air, and that the probability that two entities are gravitationally drawn together is 1, conditional on them having mass. No matter how many times a fair coin has landed heads, there is no non-accidental relation between its being tossed and its landing heads, and so no channel via which information is transmitted from one event to the other.

Pointing to paradigmatic examples of accidental correlations like streaks of lucky coinflips (and so rejecting a Frequentist notion of probability), as I have just done, does very little to characterize the correlations that *do* amount to informational relations. So, as (ii) makes explicit, a Dretskean view looks to the laws of nature for this purpose. Actually, Dretske uses different phrases to connote non-accidentalness in different places, some of which are weaker-sounding that (ii). For instance in one place he makes this claim in terms of "nomic dependence" (75 and 76), and he relies on an example where informational relations are fixed by conventions among regular partners in a card game (70), which seems not to involve "lawful" correlations in as strong a sense as DT suggests. Still, I represent a direct appeal the "laws" in DT because he often makes such explicit appeal and because it

serves clarity of exposition here, as it reflects what others in the discourse I join have taken Dretske's view to be. Admittedly, it is not self-evident what should be counted among the laws of nature, or which conditional probabilities these laws determine, but the appeal to laws of nature at least seems to say more than an appeal to "non-accidentalness" alone. For example, it is prima facie plausible that some natural law(s) ensure a correlation between smoke and fire, and that no natural law underwrites the fact that a particular coin lands heads every time it is tossed.

It is worth briefly noting one thing Dretske says about why he invokes the laws of nature. He claims that the important thing laws do (that accidents do not) is determine the truth of certain *counterfactuals* that he claims must accompany informational relations. He says "laws have a *modal* quality (they tell us what *must* be the case or what *cannot* happen) that is absent from simple statements of exceptionless correlations." (1981, 77, italics in original). To sum up, according to DT there is an informational connection between *b*'s being *N* and *s*'s being *F* when and only when the connection between these events is Factive and is modally guaranteed by laws of nature. I now turn to some dissenting views.

2B Garden Variety Correlations

A prevailing criticism of Dretske' account of information transmission is that it is too strict (Eliasmith, 2005; Godfrey-Smith, 1992; Kraemer, 2015a, 2015b; Millikan, 2001; Scarantino & Piccinini, 2010; Shea, 2007). These critics suggest that it is easy to think of counter-examples to DT, and in fact that DT rules out exactly the sort of informational

relations on which we hope to base our hypotheses about intentional content. The informational signals most pertinent to the intelligent behavior of people and animals do not inhere in perfect correlations underwritten by laws of nature, so goes the objection.

Millikan (2004) gave clear expression to the reasoning that underlies the popular accounts of information that diverge from DT. She argues that, insofar as people and animals seem to cognitively rely on informational signals, these seem *not* to be signals that carry a lawfully guaranteed indication of some state of the world. She considers an example wherein a rabbit represents that there is a predator (a fox) in its immediate environment. This a paradigmatic case for present purposes, since all parties to the debate agree that information as to the presence of the fox is carried by its body, transmitted through various structures in the air – light patterns, sound waves, airborne chemicals, etc. – and eventually transmitted to some neural activity in the rabbit, such that the rabbit can have some intentional states about the fox. Millikan has this to say about such a case:

> "[N]o natural law can require it to *be* a predator that causes [a rabbit's] predator detectors to fire. Whatever information channel she uses, it is always nomically possible that non-predators should exist who would activate it. Suppose for the sake of the argument (though very implausibly) that there are unbreakable natural laws that concern the effects of foxes on rabbit sense organs. Still, there surely are no laws that nothing *else* could possibly produce these same effects on rabbit sense organs." (2004, p. 33)

In Millikan's view, taking *N* to be any state of a rabbit brain you like, it must be possible for *N* to be caused in the absence of any fox, say, by something that looks, sounds, or smells like a fox. Millikan's basic conclusion is that DT is untenably strict because it apparently cannot substantiate basic cases like that of a rabbit's brain carrying information as to the presence of foxes.

Nicholas Shea (2007) enlists Millikan's reasoning in arguing that it is "correlational information" that underlies intentional content, as opposed to information as defined by DT. Importantly, the relevant "correlations" can be imperfect ones. That is, Shea's notion of information is non-Factive. He makes this clear, for example, when he says "instances of a type [] R which carries correlational information about C can be tokened even when C does not obtain" (2007, 420). (To translate, change "R" to "*b* is *N*" and "C" to "*s* is *F*"). So according to Shea, our rabbit's brain-state *N* can carry information about a fox being around even when, in fact, there is no fox around, provided the relevant correlation exists. This is supposed to help explain how it is possible for the rabbit to have a (false) intentional state whose content is that there is a fox around when there is not one. Importantly for my later discussion, Shea's view also includes a non-accident condition. He does not refer to "laws of nature," but says the relevant correlations are fixed by some "common natural reason." This seems to be a weaker requirement than (ii) in DT, but Shea does not offer a detailed account of what a common natural reason is.

The main point here is that the correlations Shea refers to are not meant to be mere frequencies, as in the earlier case of a lucky streak of coin-flips landing heads. Rather, Shea wants to rely on "objective" probabilities, "like the 50% chance that a lump of 4.5 billion atoms of uranium-238 will emit an alpha particle in a year" (ibid). Unfortunately, even assuming the probabilities associated with atomic radioactivity are universal, it is much less clear how to make sense of more relevant cases. What is the "objective probability" that some non-fox causes an *N* in our rabbit's brain? Is the relevant objective probability determined by *all* the non-foxes in the universe capable of making it the case that *b* is *N*, under all possible circumstances? That would not make sense, since such a calculation

would not yield the information that we antecedently suppose the rabbit relies on; there are surely an incomprehensibly large number of non-foxes with that capacity, and most of them seem not to "commonly" or "naturally" interact with rabbit brains. However, if we are to relativize to some non-universal domain in order to determine the relevant probability – say, relativize to events on Earth, or in the rabbit's local habitat, or in the habitat of its progenitors – it raises the difficult question of why *that* relativization fixes the objective probability pertinent to the rabbit's intentional state. Shea does not endorse any particular account of objective probability, and so his "correlational information" is not precisely-defined enough to assess whether it supports plausible conclusions about the informational properties of states like *N*.

This sort of worry about the appropriate relativization or *reference class* is notorious, and is going to repeatedly crop up in this discussion (Hájek, 2007; Harman, 1983; Ruth Garrett Millikan, 2007). In section 4 I will clarify the relationship between the reference class problem and the account of information I defend here, and why we should not look to the latter to resolve the former.

2C Probabilistic Information

Scarantino and Piccinini (2010) offer the most sustained and direct argument I have found against Factivity, Kraemer (2015b, 2015b) also rejects condition (i) and defends an account of probabilistic information, and Scarantino (2015) presents a detailed theory of information as a "probabilistic difference maker." Like Shea's, their views are explicitly

motivated by observing that information-users (like rabbits) respond to signals (like sounds) that do not always indicate what the response is a response to (immediate danger). On that score, Scarantino and Piccinini say "[o]rganisms do make mistakes, after all. Some mistakes are due precisely to the reception of probabilistic information about events that fail to obtain" (2010, 319). They distinguish "all-or-nothing" information from "probabilistic" information, where the informational content of the latter sort is not the fact of some event occurring, like "*s* is *F*," but is rather the *extent* to which an event is more or less probable. Kraemer similarly points to the problematic strictness of Dretske's account and proposes that "many natural signs carry information about the *probabilities* of certain occurrences" (2015a, 145, emphasis in original). Roughly, these philosophers suggest that our rabbit's brain-state, *N*, might just carry the information that there is a fox around with some probability, *p*, whether or not any fox is actually there.

Scarantino and Piccinini's reference to probability also contains a non-accident condition. Not just any correlation will do for probabilistic information, but only "reliable" correlations. Further, Scarantino's (2015) use of the term "difference maker" suggests a requirement on informational relations that shares at least a family resemblance with condition (ii) of DT. Exploring the "difference maker" idiom in depth is beyond the present scope, but note that, intuitively, whether there is fire *makes a difference* to whether there is smoke, whereas whether a fair coin has landed heads many times in the past does not make a difference to whether it will land heads on a particular future toss. Like Shea, Scarantino and Piccinini avoid appealing to laws of nature but do not precisely articulate what distinguishes the non-accidental (reliable) correlations as such. Scarantino and Piccinini say that reliable correlations are the sort that "information users can count on to

hold in some range of future and counterfactual circumstances" (2010, 318), and essentially leave it at that.

In Scarantino's (2015) theory, an informational signal implies an *incremental change* in the probability of an event, *relative to a prior* probability that is fixed by certain background data. So, rather than carrying the information that there is likely (with probability *p*) a fox around, a signal carries the information that it is *more* (or less) likely that there is a fox around, compared to the period before the signal. This will sound a lot like Bayesian confirmation theory to those familiar with it, because it is. Scarantino aims to incorporate the resources of Bayesianism alongside a Shannon-inspired measure of information. Here is his formulation (2015, 423):

> **Incremental Natural Information** (**INI**): *b*'s being *N* carries information about *s*'s being *F*, relative to background data *d*, if and only if p(*s* is *F* | *b* is *N*, & *d*) ≠ p(*s* is *F* | *d*).

The vital role played by background data in INI clarifies the kind of relativization that is obscured in Shea's unadorned appeal to objective probability. According to INI, what information the rabbit's *N* carries depends on how the background data is spatio-temporally restricted (e.g., to Earth, or to the rabbit's home forest, or…). Of course, how to identify the right background data is non-trivial. Again, I leave further delving into this reference class issue for section 4.

2D Signalling Games

There are deep parallels between the puzzle about the intentional content of cognitive states and one about how to determine the content of the apparently communicative behavior of living things. Animals, plants and even bacteria manifest behaviors that seem designed to structure the behavior of other creatures (especially conspecifics) in various ways such as helping them find food, realize reproductive opportunities, or avoid predators (Millikan, 1989; Maynard-Smith & Harper, 2003). These behaviors are standardly described in both scientific and folk contexts as *signals* that communicate things roughly like "food is over here," or "I am your potential mate," or "danger from above." Supposing these are communicative signals, what determines their content? Although this content-determination puzzle is analogous in significant ways to the one about intentionality, there are also important differences between them, so it is a substantive question whether we need the same notion of information for both, and I will not address that question here. Still, I think it is worth briefly discussing the notion of information at play in this other arena, which is a non-Factive one.

Skyrms' (2010) theory of the information content of biological signals adverts to a game theoretic framework derived from Lewis (1969). Lewis was analyzing of the meanings associated with actions in a "signaling game." In a rudimentary form, such a game involves two players, "Sender" and "Receiver," who each get certain payoffs (or incur costs) that depend on what the exogenous state of the world and on their actions. One round of the game goes as follows: only Sender observes the state of the world, then Sender chooses an action, Receiver observes Sender's choice and then must choose an action, and the state of the world and Receiver's choice together determine the payoffs. Lewis's key

insight is that sometimes, given the right arrangement of possible actions and payoffs, Sender's rational (payoff-maximizing) strategy might be constrained in a way that *allows Receiver to learn something about the state of the world* on the basis of Sender's choice. It might be to Sender's benefit to correlate her actions with what she observes, allowing Receiver to do better than randomly guess the state of the world (assuming the available options and rationality of the players are common knowledge). Lewis tries to use this framework to understand the meanings of conventional, especially linguistic signals, and Skyrms adapts it to offer a theory of the content of communications in the biological realm generally.

One oft-discussed example of animal communication is that of a vervet monkey alarm calls, which differ depending on the type of predator that a monkey sees and allow monkeys in earshot to flee to bushes or treetops or wherever is most appropriate given the kind of predator around (Seyfarth et al., 1980). In short, vervets seem to *warn* each other as to which predators are around. To simplify and describe this case as a signaling game as per Skyrms (2010), we can suppose the range of possible states of the world is "Snake," "Eagle," and "No Predator," and that Sender – the monkey who makes the call – can take at least three different actions (calls), and that Receiver (the monkey who hears the call) can take at least three different actions (behavioral responses), and that each world-state is associated with a single high-payoff option for Receiver. Without delving any further into the details of the actions and payoffs, we can see that we might model the vervet interactions as a game where Sender's strategy involves correlating her calls with world-states such that Receiver can adopt a strategy that lets him reliably make the high-payoff choice. Such a process would presumably help vervets to survive and reproduce.

Skyrms' view is that biological signals like alarm calls carry information insofar as a Receiver's actions warrant attributing to the Receiver an (implicit) *probability assignment* as to the state of the world, conditioned on the signals. Skyrms's information content is thus a *vector* of probabilities over possible states. For example, [*p*, *q*, (1-*p*-*q*)] could be a vector describing the information in an alarm call, where *p* is the probability assigned to Snake and *q* is the probability assigned to Eagle. Importantly, this model does not require a perfect correlation between Sender actions and states of the world, so these vectors standardly involve probabilities less than 1. This occurs when Sender employs a "mixed strategy," where she sometimes takes different actions for the same state of the world, rendering the "message" in her action indeterminate. For instance, Sender's strategy might be to have a call that is highly correlated with Snake but that she occasionally uses in the No Predator world-state, and since Receiver could know this, the relevant information content vector would have *p* slightly less than 1. There are significant details I am skipping over, but for present purposes it is enough to see that, on Skyrms' approach, an information signal need not indicate the state of the world that Sender, in fact, observes.

On the above approach, information content in a biological signal is determined by the details of the game used to model the phenomenon, that is, by the specific model's assumptions about what actions and payoffs there are. Information is thus relativized to the details of the game model in a way analogous to how, in Scarantino's account, it is relativized it to a particular set of background data. Choosing a game model to use is analogous to choosing a reference class. Also, in specifying the players' strategies, we effectively stipulate the relevant modal facts, so the question of whether accidents carry information does not immediately arise on this game theoretic approach.

I have now illustrated several non-Factive views information, and noted how each relies on a kind of reference class and on the exclusion of merely accidental correlations. I now shift focus to "non-accidentalness," to work toward the view of information I want ultimately to defend here.

3. Non-Accidentalness as Nomicity

3A What Matters is Modality

Recall that Dretske's appeal to laws of nature was motivated by the need to establish relevant modal truths. Similarly, Scarantino and Piccinini (2010) expound on the "reliability" of a correlation by adverting to what *would* happen in some counterfactual cases, and Shea (2007) also suggests that "common natural reasons" underwrite relevant counterfactuals. So, there is consensus that the important difference between accidental and non-accidental covariance involves modal implications. Everyone here agrees, for instance, that if smoke carries information about fire, the non-accidentalness of smoke-fire covariance implies that if there were no smoke, there would be no fire (or at least, fire would be less probable than previously). By contrast, the fact that a fair coin has landed heads one hundred times in a row does not imply that, if it were tossed slightly differently the last time, it still would have landed heads. The common assumption is that such modal implications are required for information to be (potentially) *guiding* or *revealing*. This idea is connoted by Dretske's claim that information is *what one can learn* from a signal, and in Scarantino and Piccinini's claim that reliable correlations can be "counted on," and in

the much older idea that objective probabilities must be a "guide to life" (Butler, 1736). The basic point seems to be this: nature is structured in a way that has modal implications, some covariations manifest or constitute this structure and others do not, and information only inheres in the covariations that *do* manifest or constitute this structure. Further, unless information implies this modal robustness, it cannot help explain the way information-users accurately aim at things, reliably predict things, correctly infer things, etc.

One might hope to avoid worrying about non-accidentalness altogether by construing information directly in terms of the relevant counterfactuals, making no reference to probabilities. Cohen and Meskin take this approach; their central claim can be stated as follows (2006, 335):

> **Counterfactual Theory of Information** (**CTI**): *b*'s being *N* carries information about *s*'s being *F* if and only if the counterfactual "if *s* were not *F*, then *b* would not be *N*" is non-vacuously true.

The qualification "non-vacuously" is meant to rule out information always being carried about all necessary conditions. Without it, CTI would imply that every signal carries the information that 2+2=4, for instance. Note that CTI is a Factive notion of information, given that "*b*'s being *N*" and "*s*'s being *F*" are to be construed as actual events. Further, CTI is implied by DT, given the modal implications of laws of nature. However, CTI does not commit to laws of nature as determining the truth of the relevant counterfactuals, as it is silent about *what makes* the relevant counterfactuals non-vacuously true. This silence makes CTI unsatisfying in the present context, because the truth of counterfactual statements does not reveal or explain the informative, "guiding" property that informational relations are supposed to have. For example, suppose we accept that it is non-vacuously true that if there were no fox, the relevant neural activity in the rabbit's

brain would be different. We have not said what in the world connects the fox event and the rabbit-brain event such that information about the former is carried by the latter. To ask *in virtue of what* these counterfactuals hold is essentially to ask about non-accidentalness, of which CTI offers no particular view. If it turned out that, in order for a counterfactual to be non-vacuously true it must be determined by laws of nature, then CTI would come out approximately equivalent to DT. If not one based on laws of nature, still some notion of non-accidentalness seems to be playing an important, implicit role.

I propose we understand non-accidentalness in terms of Nomic relations. These are law-like invariances in the spatio-temporal ordering of nature, which are not themselves full-fledged laws of nature. I spell this idea out further below, but to set up the discussion of Nomicity, consider three variations of the example I call "Thermometer Information," below. It seems that we can make relatively uncontroversial judgments about the relevant informational relations in these cases (stated at the end of each variant), but what we want to know is why they are the right judgments.

Thermometer Information

(a) In a typical thermometer, a volume of mercury, *V*, covaries at rate *C* with the temperature, *T*, of the surrounding space. (*V* carries information about *T)*.

(b) A volume of mercury, *V*, covaries at rate *C* with the temperature, *T2*, of a location on a different planet. (*V* does not carry information about *T2*).

(c) In the vacuum of space there is a volume of (quite solid) mercury, *V*, which does not covary at rate *C* with the surrounding temperature, *T3*. (*V* does not carry information about *T3*).

I take it that the covariation in (a) is non-accidental and is a typical example of Nomic covariance. Case (b), on the other hand, is one of merely accidental covariation, assuming there is not a very peculiar, interplanetary connection. Case (c) appears to involve

the same kind of events that figure in (a) – there is some mercury and there is a temperature of the surrounding space – but the covariation at rate *C* is missing because of the way the internal structure of solid mercury differs from that of liquid mercury. This seems to suggest that whatever Nomic relationships fix the covariation in (a) do not hold in (c). However, one might have thought that the Nomic relationships governing nature and hold *everywhere*, so what gives? If we can find a counterexample to a generalization, such as "the volume of mercury exhibits covariance at rate *C* with the surrounding temperature," does that imply that the generalization is not a law of nature or Nomically guaranteed? Note that we cannot even say that the generalization in (a) holds "usually" or "on average," at least not if those terms are just interpreted spatio-temporally, since it appears much more of space-time is below mercury's melting point than is above it. If we said "usually on Earth" that would be another matter, but then we arrive again at the reference class problem. Why should relativization to Earth in particular fix these putatively objective informational relations?

To get clear about how to account for the standard conclusions on what information is transmitted in the range of Thermometer Information cases, we need a more nuanced understanding of non-accidentalness. Theories of information (as underlying intentionality) generally invoke non-accidentalness in the form of laws of nature, common natural reasons, or the reliability of certain correlation, but without spelling out the details of how this non-accidentalness operates between various kinds of events.

3B Law-Like Invariance, Nomicity

I propose to use Woodward's (2017) discussion of the relationship between physical modality, laws, and counterfactuals to guide the inquiry here. A key feature of Woodward's account that has not been adequately appreciated in the philosophical literature on information is the role that *initial conditions* play in the invariances that scientists probe, even those often dubbed a "Laws of Nature." Woodward suggests that "invariance" is actually a more felicitous term than "law" to designate the non-accidental relationships we are interested in here. This is because laws are, in general, paired with a range of initial conditions, and only within that range does the law guarantee the invariance in question. One could also say that a law can only be "counted on" to hold given the requisite initial conditions. The basic point here is that a generalization does not need to be exceptionless in order to be a law, because exceptions exist outside the range of initial conditions corresponding to a law. Woodward helps illustrate this with the examples of, first, General Relativity, which he notes is thought not to apply over the smallest measurable distances due to quantum effects, and second, the strong, weak, and electromagnetic forces, which appear, at very high energy scales, not to behave remotely like our models of them say they will (2017, 8).

This feature of the natural laws that practicing scientists employ is obscured by an old-school manner of articulating laws as logical schema, in particular as universally quantified conditional statements. For example, a candidate statement of a law of nature, using the classical philosophical formulation, would be "all smoke comes from fire," and another would be "if ambient temperature increases by X, the volume of nearby mercury increases by $X \cdot C$." Such expressions fit with a view of laws as universally applicable, and

as having *instances* wherever the properties they involve are manifest. Wherever in the universe there is smoke, the candidate law I just mentioned (if it were really a law) would be instantiated and would guarantee that a fire is there too. Woodward makes clear that this way of thinking about laws is deeply problematic.

Laws are typically formulated by scientists as (systems of) differential equations, which differ from universally quantified conditionals in several important ways. A law-describing differential equation involves a set of interrelated variables, each of which represent some aspect of nature, each of which can take a wide range of values while still exhibiting the invariance specified by the equation. That is, *what a law says is invariant* is the relationships among the variables specified by the equation(s), not generalizations about entities. Such generalizations, like "smoke follows fire," might loosely be thought of as described by certain combinations of certain values of the variables that figure in a law. Thus, a lawful invariance can be realized in starkly different ways, which are represented by starkly different sets of values and functions that yield a *solution* to the differential equation(s). To refer to all of these possible solutions as "instances" of the law makes little sense because of how they range over phenomena that are so qualitatively diverse. While it is relatively easy to group together the phenomena (instances) described by "there is smoke and there is a fire," it is not remotely easy to group together the phenomena that are aptly described by the differential equations that represent lawful relations among the strong, weak, and electromagnetic forces. In fact, exactly *what* physical phenomena such differential equations describe are so abstract and distant from everyday life that providing an illustrative example (to any non-physicists) is basically impossible without extensive explanatory work. The main point here is that such

differential equations only *have* solutions when their concomitant variables fall within a certain range of values, that is, initial conditions.

Therefore, if there were a law concerning the relationship between the volume of mercury and the surrounding temperature, it might specify that the two variables, *V* and *T*, *do* perfectly covary at rate *C*, but only when they fall within certain ranges. The fact that those ranges are exceeded in the vacuum of space would then explain why the law does not apply in case (c) of "Thermometer Information." However, our best scientific theories do not posit such a specific law about the relation between mercury and temperature. Rather, the law(s) relevant to that case involve highly general equations with variables whose values can describe both solid and liquid mercury, and countless other substances. The relevant laws hold both in typical households and in the vacuum of space, and so the invariance described by the relevant differential equations subsumes *both* the observable covariance in (a) *and* the lack thereof in (c). However, to formally represent what is invariant across these two cases requires a level of abstraction and precision that is difficult or impossible to achieve using everyday language (this is one reason calculus is important).

Now let us consider the fact that laws of nature do not advert to entities like "smoke" or "mercury." Most of the regularities that we talk about, and that inform how we live our lives, are like the one in (a) in the following way: they are *law-like* insofar as they are invariant given certain initial conditions, but compared with the invariances we call "Laws of Nature" they are much less general (i.e., have a much narrower domain of application) and are less theoretically relevant for the empirical study of other very broad generalizations. The difference here is essentially a matter of degree (2017, 9):

> "Laws are those invariant generalizations whose range of invariance is sufficiently large…and which (at least in many cases) are integrated with other laws as part of a coherent theory. One consequence of this picture is that the boundaries of the notion of a law become vague and there is a sort of continuum between laws and generalizations that are…narrow enough that [they] are not regarded as laws."

Following Woodward, I take it there is no fundamental principle that distinguishes an invariance as a "law" from the less general invariances that are familiar everywhere outside of basic science. (At least, no fundamental principle that does not appeal to the practicality of certain linguistic conventions). Going forward, by "law-like invariance" I mean to refer to all generalizations on the continuum that Woodward describes in the quotation above. It is in these terms I propose we understand the Nomicity of information.

Recall that condition (ii) of DT required that informational relations be "fixed by some law(s) of nature." Among the dissents to DT that I reviewed, Millikan's was particularly critical of (ii), but the others also at least tacitly rejected it, avoiding the reference to laws in favor of some weaker-sounding non-accident condition. I agree that we should abandon condition (ii) because it rules out the cases of information transmission we are most interested here, as there are not laws that specify correlations (perfect or otherwise) between, for example, foxes and rabbit brain-activity. Given the preceding discussion, I propose we adopt the following definition of **Nomicity**: covariation is Nomic if and only if it realizes a law-like invariance. Since there *are* law-like invariances that hold between, foxes and rabbit brain-states, requiring that informational relations are Nomic does not force us into an unacceptably restricted view about what informational relations exist. This reveals that the problemetic strictness of DT can be addressed by giving up Lawfulness in favor of Nomicity. Now I want to turn back to reconsider Factivity.

4. Nomic, Factive Information

4A Mistake-cases and Factivity

To briefly review, condition (i) in DT says that if *b*'s being *N* carries information about *s*'s being *F*, the conditional probability of *s*'s being *F*, given that *b* is *N*, must be 1. Less technically, DT says *b*'s being *N* indicates that, in fact, *s* is *F*. Recall that recent accounts of information have rejected this condition, partly on the basis of considering cases involving *mistakes*, such as a rabbit who responds to some sound as if it indicated a predator when in fact no predator is nearby. So, the observation that information-users make mistakes is presented as constraining our information theory in such a way that we must posit non-Factive information. I aim to show that this reasoning on the basis of mistake-cases is fallacious. To help illustrate my points, I rely on the following two variants of the case of Fleeing Rabbit:

Fleeing Rabbit

(a) A pattern of air-vibration caused by a stalking fox reaches a nearby rabbit's ears and causes a brain state *N* in the rabbit. *N* then figures in further brain processes that normally generate the rabbit's fleeing behavior, and it does flee.

(b) A pattern of air-vibration caused by a fallen branch reaches a nearby rabbit's ears and subsequently causes a brain-state *N** in the rabbit. *N** then figures in the same processes as in (a), and the rabbit flees. No predator is nearby.

Let us stipulate that the pattern of air-vibration and brain activity are morphologically exactly similar across (a) and (b), such that the everything after the fox's step is indistinguishable from everything after the branch's landing – say, indistinguishable from the viewpoint of a well-positioned observer who has superbly high-resolution devices for distinguishing patterns of air-vibration and rabbit brain activity. This will mean at least that *N* and *N** are physically constituted by the same rates of neural firing in the same

neurons, they are proximally caused by the same pattern of energy impinging on the rabbit's sense organs, and their proximal effects are also the same.

Now I will reconstruct the line of reasoning that I take to be implicit in the accounts of Shea (2007) and Scarantino and Piccinini (2010). To start, we assume *N* carries some predator-related information, since we are assuming the information in *N* supports ascribing predator-related intentional content to the rabbit. Now add the following premise; *N* and *N** must carry exactly the *same information*. One might think this premise is supported by the fact that *N* and *N** are morphologically identical and have the same proximal causes and effects, and the fact that the rabbit presumably has the same intentional state across (a) and (b). This is often explicitly a part of why theorists support a non-Factive view of information; an information-carrying state like *N* is thought to carry probabilistic content – something like "90% predator, 10% other" – because that way we can understand *N* and *N** to be information-carriers of the same basic kind. Insofar as one is convinced that *N* and *N** have the same relevant information content, and given that *N* carries predator-related information, we can conclude that *N** also carries predator-related information even though no predator is around. Thus, the non-Factivity of information follows.

However, the stipulated similarities between *N* and *N** do *not* entail that they carry the same information. Nothing in the discussion so far suggests that morphology and proximal causal relations fully determine information content. All parties here agree that covariation must be non-accidental in order to be information-transmitting, but the non-accidental relations that *N* and *N** stand in are not exhausted by proximal causal relations. The critics of DT endorse non-accident conditions that are weaker than Lawfulness, and I

contend that these critics should be happy to accept the Nomicity requirement articulated above, as it does the job that notions of "reliability" or "common natural reason" are supposed to do. However, understanding non-accidental covariation as tied to initial conditions in this way undermines the argument against Factivity. We cannot, in general, assume that *N* and *N** carry the same information, since *N* and *N** stand in different Nomic relations to the rest of the world.

The event of *b*'s being *N* stands in all sorts of Nomic relations and carries all sorts of information. It carries information about the presence of a fox, about patterns of air movement in the rabbit's ear, about the presence of a working rabbit brain and much more, all specified with respect to the initial conditions of one or another Nomic regularity that *N* stands in. This fits with the idea that information, understood as a natural, objective commodity, is what a hypothetical observer can learn from a signal, where such a hypothetical observer could have any perspective we can imagine. In other words, one might be interested in any *s* that stands in some Nomic relation to *b*, and this *s* could be described in terms of any manner of *F*-ness, *G*-ness, etc. Thus, when we talk of *the* information that a state like *b*'s being *N* carries – all of the information – we are talking about an incomprehensibly huge and diverse multitude of facts, each guaranteed by one in a multitude of sets of initial conditions. It can be easy to lose sight of how copious the informational structure of the world is, partly because we are so good at filtering relevant from irrelevant information. Nomic *and Factive* covariation is very easy to come by, and certainly is present between rabbit brain-states and foxes. Under some range(s) of initial conditions, *N* does Nomically guarantee that a fox is nearby. This is not just a technical matter, because it might turn out that *N*'s role in giving rise to a cognitive achievement like

fleeing is not determined just by its proximal causal properties. Rather, it might be how *N* figures in a complex system of dynamic relations that structures the rabbit's fleeing behavior, in which case *N*'s causal and modal relations to distal entities not present in case (b) could be relevant a theory of what intentional content is associated with *N*. In the subsequent chapters I develop a view of how such broad, dynamical relations are essential to the understanding of cognition – a view that is ruled out by the assumption that *N* and *N** have the same relevant informational properties.

Recognizing the copiousness of information allows one to ask, in essence, "which information is it," carried by a particular pattern of brain activity, *N*, that is relevant for understanding some intentional state or activity. That is, having hypothesized a role for some information-carrying event – like a role in predator-avoidance behavior – on what basis does one pick out the Nomic relationship and attendant initial conditions that define the information relevant to that role? This is just to state the reference class problem in terms of information. By assuming a certain kind of answer to this question, we can be compelled by the idea that *N* and *N** must carry the same (relevant) information. We come to this judgment if we assume a reference class that roughly reflects our own epistemic position as human observers. That is, when we look at a rabbit brain across various contexts, *N* and *N** look to us like the same state – *we* can learn the same sorts of things from each of them. For instance, we can learn from each that the rabbit is about to exhibit predator-avoidance behavior. But appealing to what we can tell apart brings our own reference frame into the picture, so to speak. It smuggles our own intentionality into what was supposed to be an account of how a brain, body, and environment interact to give rise to intentionality via information transmission.

So, to classify *N* and *N** as information-carriers of the same kind requires some justification for selecting reference classes. However, what justifies picking out a particular class of informational relations is not something to be answered in information-theoretic terms; something besides information theory must help us tell which information is relevant to a particular intentional process. A common thought is that something about how *N functions* in a larger process of predator avoidance determines which is the relevant information *N* is carrying. Dretske (1988) developed this thought in terms of states having a function to indicate states of the world, where functions derived from evolution and learning, broadly speaking. Informational Teleomantics names a family of theories that try to account for the information-carrying functions of bodily processes and communicative behaviors by appeal to natural selection (Godfrey-Smith, 1992; Millikan, 1984, 2017; Neander, 2017, 2018; Papineau, 1987; Shea et al., 2018). While *N* and *N** do not have identical informational properties, they might both have the function to indicate (carry Factive information about) the presence of a predator. It could be that *N** has the function of carrying information about a predator even though no predator is around, because *N** is failing to achieve its function – it figures in a malfunctional process.

Allow me to briefly recapitulate. The rabbit's avoidance behavior is similar across (a) and (b), and *N* and *N** are assumed to play a similar role in those behaviors. If this similarity in behavior is to be accounted for by the similarity of information between *N* and *N**, non-Factivity of information follows. But *N* and *N** do not have identical informational properties. Once we grant that *function* is needed to determine which informational relations are relevant to the behavior, we can appeal to the similarity in function rather than a similarity in informational properties to explain the similarity in behavior across (a) and

(b). In other words, once we have helped ourselves to a notion of information-carrying functions or ways of *using* information, we can describe mistake cases like (b) in terms of misfunctioning or misuse of information. There is thus no need to think the rabbit's mistake means its brain is carrying information as to a probable predator when, in fact, there is none. Instead we can suppose, roughly, that a rabbit's perceptual-motor system is using *N** as if it carried information about a predator, even though it does not.

It is a controversial philosophical matter just what determines the function of part of an organism, but it is beside the point here whether the a notion of selected functions to carry information can stand up to scrutiny. If not by appeal to functions, something other than information theory must support a hypothesis about which information figures in an explanation of some perceptual or cognitive process. And going in for non-Factive information does not bring us closer to resolving the reference class problem. In fact, it seems the reference class (or background data, in the terms of INI, above) must be more specific for non-Factive information; not only must we distinguish predator-related information from irrelevant information also carried by *N*, say about the shape of the fox's foot, we must also account for the particular probabilities that figure in the relevant information content – say "90% predator" rather than "95% predator." In any event, some theoretical resource outside of information theory is doing important work in the picture of information transmission as it figures in Fleeing Rabbit. Whatever this theoretical resource is, it can allow us to make sense of the similarity between Fleeing Rabbit (a) and (b) *without* holding that *N* and *N** must carry the same information. Thus, we should not be moved to endorse a non-Factive notion of information based on the observation that animals sometimes make mistakes.

4B Nomic, Factive Information

One can maintain that processes of perception and cognition importantly depend on information flows within a nervous system and environment, and yet acknowledge that it is a substantive question which information flows are relevant in that regard. This is to acknowledge that the reference class problem cannot be resolved just with the tools of information theory. Having done this, and also recognizing the extreme heterogeneity of Nomic regularities (given initial conditions), we can understand information, as carried between natural events, as Nomic and Factive. Here is how we can define Nomic, Factive Information (NFI):

> **NFI:** Given the event of *b* being *N*, this event carries the information that *s* is *F* if and only if (i) *s* is *F*, and (ii) this covariation between *b*'s being *N* and *s*'s being *F* is Nomic.

NFI is similar to DT, but it is distinct in two important ways. First, it incorporates the above notion of Nomicity based on Woodward's (2017) view of laws and law-like invariance. Second, NFI makes no explicit reference to probability. The requirement of Factivity entailed by DT's condition (i) can be readily expressed without mentioning probability, as I have done in NFI's condition (i). I suspect that Dretske's appeal to conditional probabilities (of 1) has been a source of confusion in this discourse, in part because it suggests that the most straightforward way to address DT's problematic strictness is to lower the required probability in (i) to something less than 1, when, as I argued, Lawfulness is really the source of the problematic strictness.

Because NFI makes no explicit reference to probabilities, one might worry that this notion of information has nothing to do with probability or uncertainty. That would be an incongruous result given the roots of thinking about information in this way. However, NFI

is still connected to uncertainty in that the *measure* of information carried by an event depends on what happens in counterfactual scenarios where the event does not occur. That is, given that *N* carries information that *s* is *F*, the amount of this information depends on what other ways *b* might have manifested (as *O*, *P*, etc.) and how likely it is that *s* would have been *F* (rather than *G*, *H*, etc.) in those counterfactual scenarios. If *s*'s being *F* is quite common across all possible states of *b*, then *b*'s being *N* does not carry much information about *s*'s being *F*. By contrast, if *s* is rarely *F* except when *b* is *N*, then that signal carries much more information – it removes uncertainty to a greater extent. Recall that the counterfactual manifestations of *s* and *b* are literally part of the picture of information transmission in Figure 1 (p.12). For information transmission to be measurable, *F*, *G*, *H*, etc. must have determinate probabilities that sum to 1. Probabilities are a part of Nomic, Factive information, not in the information's content but in its measure.

Having some measure of the strength of an informational relation is essetial. For one, the amount of information a signal carries in related to what one can do with that information. Some reasons to endorse or abandon the hypothesis that *N* functions to carry the information about a nearby predator will pertain to how much information about a nearby predator *N* carries, how much it communicates elsewhere in the brain, and how this measure of information compares to the strength of other signals of the predator. I have not focused here on a formulation for quantifying information, but requiring that information be Factive and Nomic does allow for measuring different amounts of information transmitted depending on the Nomic regularity in question. For example, given a certain reference class restricted to the rabbit's environment (however defined), *N* would plausibly both carry the information that a predator is nearby and the that a mammal is nearby, but

carry more information as to a nearby predator, that being the more ulikely of the two events. So, even though NFI does not mention probabilities, this is still a notion of information that is fundamentally tied to the idea that events reduce uncertainty about various the state of the world, in varying degrees. By properly appreciating how manifold are the law-like, modally robust relations between actually existing parts of nature, and by also acknowledging that accounting for intentionality requires resources outside of information theory, we can recognize that Nomic, Factive information is perfectly adequate for helping us understand how intentional content might arise.

Conclusion

My main goal has been to highlight a misstep in a common line of thinking of information as non-Factive. The prevailing critical responses to Dretske's (1981) theory conflate the problematic strictness of this account with the fact that it does not explain how information-users sometimes make mistakes. In the hopes of solving that latter problem with information theory alone, the misstep relies on a false assumption that certain (brain-local) states must have the same informational properties across similar instances of behavior, though the behavior is appropriate in some cases and mistaken in others. Lacking a precise understanding of non-accidentalness, recent discourse on Natural Information has failed to appreciate how a careful understanding of Nomicity makes Factive and Nomic information broad enough to avoid the strictness of DT. I have tried to show how Natural Information – the kind of thing that perceivers and cognizers generally depend on their brains to process in order to engage intentionally with their world – should be understood as Factive and Nomic.

3 Dynamic, Generative Bases of Cognition

1. Introduction

Much of our scientific interest in brains and neural networks aims to improve our understanding of cognitive activity. What is this supposed explanatory relation between the neural and the cognitive? In what sense are cognitive processes "based" in neural and other processes in the brain, body, and environment? Stepping back to relate this question to other pieces of the larger philosophical puzzle here, keep in mind that cognitive processes are described in intentional terms. Seeing, believing, wanting, imagining, and other activities associated with cognition are activities *about* external objects and possibilities. By contrast, interactions among neurons are described in a physical or formal way – in a vocabulary that is removed from what cognition is about. My broad goal in this chapter is to defend a view of the sense in which we can understand cognition as based in lower-level processes.

More specifically, I offer a view of cognition as generated by dynamic relationships between the brain, body, and environment. A goal of the larger project here, in the background of this chapter and the foreground of the next, is refining and supporting a view of cognition as *embodied*. Part of that larger view involves drawing on principles from dynamical systems modeling to understand the lower-order relationships in terms of

which to try to better understand cognition. How to characterize the explanatory relationship between lower- and higher-levels in such systems is not straightforward, and this complicates the more general controversy about the explanatory relationship we take there to be between the neural and cognitive. Getting clear about this issue, I argue, involves displacing a common view of the explanatory relation in question. Constitutivism is the view that lower-order processes are explanatorily relevant to a higher-level, cognitive phenomenon insofar as they are constituent parts of it. I argue against this view on the grounds that higher-order phenomena in dynamical systems can be more local than their lower-order components are. I support an alternative – Generativism – for characterizing the sense in which cognition is based in the dynamics of brain, body, and environment. I proceed as follows:

In section 2, I describe the basic aspects of dynamical systems models relevant to this inquiry, and the major lines of thought that suggest they are ca serve our understanding of cognition. In section 3, I articulate and defend a view of cognitive processes being dynamically and generatively based in lower-level processes. I do this by considering some basic features of different kinds of explanation, describing Constitutivism as a candidate kind of explanation, then arguing against Constitutivism and for Generativism based on what I show about dynamical systems. In short, I argue that lower-order dynamics can give rise to entities that they are not contained in, so this "giving rise" must be seen as a matter of generation rather than constitution. In section 4, I illustrate my main points in more detail by revealing their application in a formal, dynamical model of a simple animal behavior.

2. Dynamic Levels of Order

2A Dynamical Systems Models: What?

What is important to know about dynamical systems for the purpose of modeling cognitive phenomena? I want to keep this discussion fairly compact and non-technical, thus I will overlook certain ways that dynamical systems models can vary from the kind I describe here. Readers familiar with dynamical systems models will find parts of this section elementary – I ask for your patience as I set up the discussion of what is involved in a dynamical systems-based explanation of cognitive processes. For readers who would like a more detailed overview of dynamical systems models as applicable to the domain of intelligent behavior, I point to the following sources: (Gelder & Port, 1995; Jirsa & Kelso, 2004; Kelso, 1995).

A basic feature of dynamical systems models is that the systems they model are inherently undergoing a process of change. Elements in these models are typically functions of time and of each other. At any given point in time, the activity of each element is shaping and being shaped by the other elements to which one is dynamically related. The system is understood in terms of the character of its constantly shifting from one state to the next – modeling the system means representing how it changes from moment to moment. By default there is *no pre-stimulus state or inactive mode for such a system*. The dynamical system does not await some input before moving or, alternatively, one could say that its current state is also its input.

Insofar as dynamic interrelations among all parts of the system structure its process of change, the causal contributions of any of its parts are *global*. That is, the impact of one

part of the system is not, in general, isolated to some local region of it, but rather is potentially implicated in every aspect of the overall change to the system; any part of such a system may constrain and be constrained by the entire remainder of the system. I will look more closely at the globality or locality of the components of dynamical systems later in the discussion. To briefly illustrate by contrast, we would not treat something as a dynamic system if describing how the system changes is done separately from describing the state of the system at a time. In that case the system's global state would not be essentially bound up with how it transitions into the next state. A minimal example of such a non-dynamic system is a logic gate, whose "activity" occurs only under a discrete input condition, and whose rules of behavior are specified in a abstract list of input-output relations. Nothing interesting happens to a logic gate *over* time.

A more nearby case worth briefly discussing is Conway's Game of Life and other cellular automata (Gardner, 1970; Izhikevich et al., 2015). The relevant system in the Game of Life is a grid of pixels, and its global state is constituted by the combination of the "on" or "off" states of all off the pixels. The system transitions from one state to the next as pixels blink on or off at each timestep, and each pixel's change (or non-change) happens according to a function of how many pixels immediately around them are "on." The Game of Life is a dynamical model in that the system's occurrent state determines a distinct temporal evolution of the system. Also, the Game of Life has the interesting property of many dynamical systems of being able to manifest higher-level entities as patterns in the collective interactions governed by the lower-level (pixel-level) interactions of the system. Given the right initial conditions, "Blinkers" and "Gliders" and other macroscopic pixel-structures can persist and interact over long sequences of time (ibid.). These are basically

cyclical patterns among pixels within a certain area, which just derive from how the state-to-state transition rules play out over multiple timesteps.

An important respect in which cellular automata standardly differ from some dynamical systems is that the causal contributions of each of the parts (pixels) of the system are local, rather than global. How each pixel changes is fully determined by the states of neighboring pixels, and changes further away in the system make no difference to the activity of the individual part, at a time. Correspondingly, the individual pixel does not, at a given moment, make a difference to the behavior of disparate parts of the system, but just to the behavior of neighboring pixels, and so a transition in one part of the system can be fully predicted and described without reference to all parts of the system. I will be focusing here on systems whose lower-level parts contribute in a global way, so I describe cellular automata as a way of bringing out, by contrast, the kind of globality exhibited by components of the systems I will discuss below. Systems with globally contributing lower-order parts are of special relevance because, as I will show, they pose a problem for a common, Constitutivist view of the sense in which higher-order entities depend on them.

The *state-space* of a dynamical system is basically the space of all of its possible configurations, which encompasses all the initial conditions in which we might find the system. More technically, one can depict the state-space as a vector space, where each vector includes every aspect of the system's position and the rates of change of those aspects (and possibly rates of change of those rates of change…etc., for higher-than-second-order dynamical systems). For example, for certain purposes one could effectively depict a basketball's state-space with a six-dimensional vector, three for its X, Y, and Z position, three more the rates of change of those positions, i.e., velocity in each direction.

Because the values for rates of change are included in this vector, it encodes the ball's instantaneous direction of movement along all of its dimensions, thus the system's state determines a unique evolution or *trajectory* of the system's state over time. One could complicate this, for instance, by adding stochastic rates of change, or by allowing that the system is not closed and receives exogenous influences on its states, or by depicting the dynamic relationships themselves as functions of time. There is no need to explore these complications here though. The point is that describing a dynamical system is standardly done by representing the space of all of these trajectory-determining points, and thereby representing the possible ways the system can evolve over time.

In addition to letting one project the system's possible trajectories from any given initial state, having this state-space description often supports the identification of interesting *higher-level features* of the systems behavior, in terms of some pattern or order in the state-space, or in the trajectories through it. These are typically called "order parameters" or "collective variables." A higher-level feature is not just a particular collection of vectors or trajectories, but rather it is some recognizable heterogeneity or partition in these trajectories, sometimes called a feature of the "phase flow." A higher-level feature is higher-order with respect to the relatively lower-order variables that make up the dimensions of state-space itself, which are typically called "control parameters." Some of the simplest and most vivid examples of such higher-order features can be found in the realm of harmonics and standing waves; the Chladni plate experiment offers a perspicuous demonstration and can easily be found online. In this domain, typically an important control parameter is the frequency of some cyclical motion (oscillation), and certain changes in frequency are associated with changes in the order of the whole system.

I want to briefly highlight an example of a common, higher-order feature of dynamic systems that I will refer back to later – an *attractor*. An attractor is, roughly, a place in the state-space toward which the system's state trends from a wide range of initial conditions. Probably the most familiar example of this is gravitational attraction. Consider a system of two massive bodies in outer space. The varying parameters that make up state of this system – including the twelve dimensions of locations and velocities for both bodies – are dynamically interrelated such that, under a broad range of initial conditions, the system reliably comes to fall within a particular part of its state-space. The attractor might be represented by a point or volume of the state space (depending on the dimensions in which the state-space is represented). In the two-body, gravitational system, an attractor is constituted by the set of points in state-space that represent the two bodies having collided. An attractor also determines a "basin of attraction;" part of the state-space outside of the attractor, from within which all trajectories lead into the attractor. At an attractor, the system's dynamics work to keep the system there, so there are no trajectories leading out of it. In this sense the attractor represents a stable part of the state-space. However, this stability does not imply stillness. The system might be drawn to a self-sustaining pattern of activity, sometimes called a "limit cycle," where it undergoes a regular pattern of change. This is exemplified by the fact that, under the right conditions, rather than being drawn into an attractor where the two massive bodies come into contact, they might get stuck orbiting one another, continually and cyclically shaping each other's motion. To emphasize the main point, an attractor is higher-order with respect to the system's control parameters and trajectories through the state-space. The attractor is a qualitative feature of

all the ways the system can move through the state-space; it serves as a basis for classifying whole families of states and trajectories of the system.

Mathematically, dynamical systems are most standardly described by differential equations (sometimes multiple at once). Parameters representing different parts or aspects of the system figure as the variables and expressions in the equations. These parameters are typically functions of one another, or of one another's time derivatives, or their derivative's derivatives, etc. This is illustrated by the way velocities and distances are interrelated in gravitational systems. For instance, the faster a satellite is moving parallel to the edge of earth, the faster it must also be falling toward earth (and vice versa) in order to maintain its orbital distance. Thus, the differential equations describing such a satellite system would specify this relationship between rates of change. More generally such differential equations, if they accurately describe the target system, present a formalism specifying the states that the system can possibly occupy as it evolves from each state. That is, these equations determine a state-space. Aside from differential equations, dynamic systems can also be described by directed graphs, especially cyclic graphs (Hintze et al., 2017; Koller & Friedman, 2009). These kinds of models all effectively describe a space of possible, state-to-state changes the system can undergo, and they determine the course of such changes or "updates" that ensue from any given starting condition.

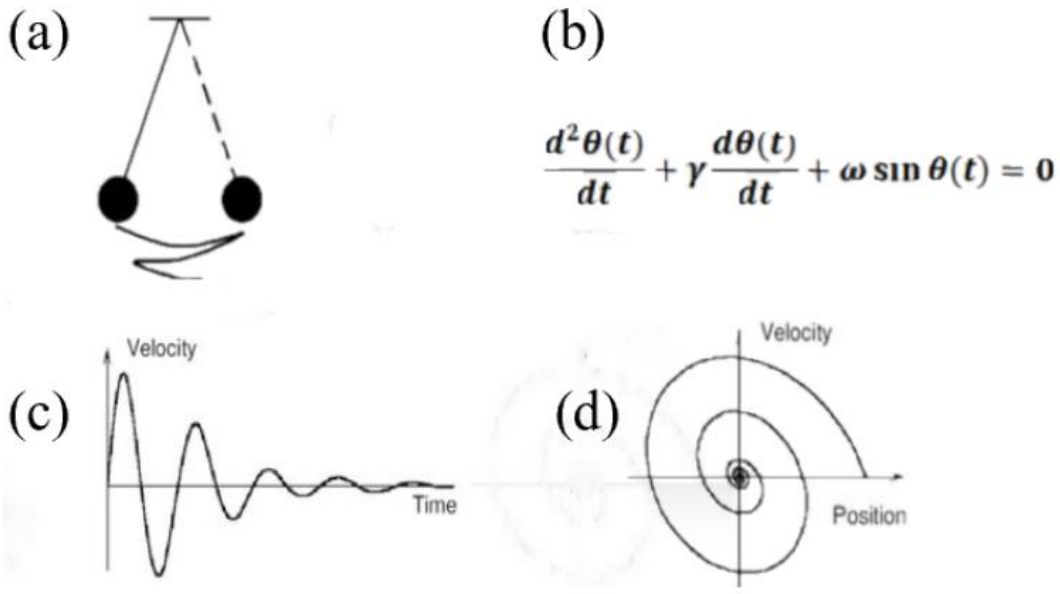

Figure 2
Representations of the movement of a pendulum over time. (a) Visual schematic (b) Differential equation: θ for the angle from vertical of the pendulum arm, γ for the friction constant, and ω for the natural frequency (c) Plot of velocity versus time (d) Temporal trajectory through state-space. Illustrates fixed-point attractor where velocity and position are zero. (Image from Hasse & Bekker, 2016).

2B. Dynamical Systems: Why?

Dynamical systems models are used to improve our understanding of the macroscopic behaviors of the target system in various ways. Often the impetus for coming up with dynamic systems model is the observation of some high-level behavior. One hopes to better understand that behavior and the system itself by identifying the lower-order, dynamic relationships in virtue of which the behavior arises. Also, we might look for the lower-order dynamics that underlie causal and predictive relationships we find at the higher-level in order to refine our grasp of those relationships. If we are seeking a more

thorough understanding of a causal relationship between *P*s and *Q*s, coming up with a model of the lower-level dynamics underlying that causal relation could allow us to furnish new, testable hypotheses about the exact conditions under which *P*s do and do not cause *Q*s. Dynamical systems models also provide a framework for intuitively comparing the high-level behaviors of diverse complex systems, especially if tools are available to represent many trajectories through state-space at once (to represent the "phase flow"). For example, a solar system and a system of marbles on a suspended, elastic surface can exhibit similar attractor dynamics, which can be visualized as a similar clustering of the trajectories of the system. Higher-level features and the way they arise from lower-level dynamics is a central part of what is interesting and useful about dynamic systems modeling.

This is an active and growing area of research; there are many philosophers, cognitive scientists, neuroscientists, and engineers proposing why and how we should use the language of dynamical systems to understand intelligent behavior (Beer, 1995; Gelder & Port, 1995; Kelso, 1995; M. Lewis, 2005; Rockwell, 2005; Schöner, 2008; Spivey, 2008; Warren, 2006), and accounting for particular capacities using specific dynamical models (Braud et al., 2018; Frank et al., 2015). A general idea behind these approaches is that dynamic relations among sensory, motor, and environmental components give rise to higher-order, cognitive behavior. This idea is made especially compelling in the domain of activities that mainly involve controlled movement of one's body, where the relevant dynamics can be clearly described and observed. In these cases, initial conditions vary continuously and there are complex but measurable dynamic relationships between limbs, sensory organs, the structure of how they change, and the structure of the environment. With the right access to the brain, data about these dynamics can be usefully compared to

patterns in an organism's neural dynamics. For a classic example, the influential Haken-Kelso-Bunz dynamical model is used to describe, inter alia, rhythmically patterned finger movements (Haken et al., 1985; Kelso, 2008). More recently, Krishna Shenoy et al., (2013) offers a theory of controlled arm movement arising from neural, dynamic "pattern generators." A related idea is that the brain ubiquitously employs "predictive coding" schemes, continuously working to minimize a mismatch between predicted experience and actual experience (Clark, 2016; K. Friston, 2010; Gładziejewski, 2016; Hohwy, 2013). This idea does not strictly require a dynamic systems-based approach, but it is well suited to one with dynamic relations between what is predicted and what is experienced, as some of its proponents have recognized. Moving from the biological toward the engineering realm, the idea that intelligent, autonomously controlled behavior could rely crucially on feedback relationships giving rise to stable behaviors over time goes back at least to W. Ross Ashby, whose famous, "Homeostat" machine achieved the behavioral "goal" of minimizing its interactions with its environment (Ashby, 1953). More recently, some AI and robotics researchers have used a dynamic systems approach to model and test minimally cognitive or proto-cognitive agents (R. Beer, 1995; R. D. Beer & Williams, 2015), agent behavior in forced-choice tasks (Bogacz et al., 2006), and task prioritization by locomotor robots (Reverdy & Koditschek, 2018).

In short, there are numerous promising routes to understanding cognitive phenomena in terms of dynamical systems models. Specifically, there are numerous lines of support for the following, broad thesis, which I call "Dynamical Basis:"

> **Dynamical Basis:** Cognitive processes are based in dynamic relationships spanning brain, body, and environment, in the sense that higher-order features of dynamical systems are based in the lower-order components of such systems.

I have tried to word this thesis in a way that does not obscure the distinction between the elements that populate a dynamical systems *model*, like parameter values and attractors, and the parts of the natural world that are dynamically interrelated in ways that our models aim to describe, like patterns of impinging physical energy and neural activity. Dynamical Basis is a quite general idea and thus leaves some large questions unanswered. For one, Dynamical Basis might be interpreted either as a claim about some cognitive processes, or about all cognitive processes. Is the point to understand cognition per se as arising dynamically, or just certain cognitive activities? I leave this question aside here, as the weaker, "some cognitive processes" version of the hypothesis is enough to warrant the present investigation. Dynamical Basis also invites one to ask what kind dynamic relations give rise to a cognitive system. Complex systems can exhibit all sorts of patterns, and if Dynamical Basis is right then we should hope to be able to specify what sorts of patterns in what sorts of systems serve to model cognitive processes. This question is beyond my current scope, but I will try to make some initial progress on it in the next chapter. For now, bearing in mind the kind of systems I have described here, I turn to investigate what it means to say that lower-level components are explanatory bases of higher-level, cognitive processes.

3. Generation vs. Constitution

3A Constitutivism

I started this chapter by wondering about the sense in which cognitive processes can be understood as higher-level processes that arise from or are "based" in lower-level processes. The idea that it is important to distinguish between levels of analysis in complex, functional systems has been prominent for a long time (Craik & Lockhart, 1972; Cummins, 1975; Darden & Maull, 1977; Marr, 1982). However there is not broad agreement about how we ought to understand the relevant relationship between these levels. I aim to challenge a common way of thinking about this relationship as a *constitutive* relation and advance an alternative, *generative* understanding of the relationship. I illuminate the reason to adopt this alternative by appeal to what I have said about levels of order in dynamical systems. First though, I will outline some key features of this explanatory "basis" relation in terms of the kind of explanatory role it is supposed to play.

One question we can ask about cognition is "what is it?" That question calls for an identity relation, which does not involve a distinction between a lower- and higher-level. To state an identity of a cognitive process is to re-describe that process itself – at its own level, so to speak – rather than describing something that the cognitive process depends on but is different from. I take it that cognitive neuroscience mostly tries to discover processes that are relevant to cognition but whose properties fundamentally differ from the features of cognitive processes per se. Cognitive processes have properties like being *about* things (even potentially non-existent things), being *rational* (or irrational), and being something *attributable to an agent*, whereas neural processes do not appear to have such properties.

Thus, the explanatory relation between cognition and whatever its lower-level components are is not one of identity.

Another question we can ask about some cognitive activity is "what causes it?" That question is standardly answered by reference to something that occurs *before* the relevant cognitive activity. Also, causal relations do not imply any significant distinction in spatio-temporal scale. A causal explanation of my visualizing my sister, for example, might just be that someone in earshot said her name. A causal explanation of my cognitive activity could describe neural activity, for instance, in a case where my sister-visualizing activity were elicited via electrodes applied to my brain. However, in general, our investigation of the brain is not aimed at uncovering what causes cognition, because it is aimed at uncovering something that stands in a synchronic explanatory relation to it. My line of inquiry here is premised on this thought that cognition is based in brain activity that occurs at the same time the cognition occurs.

We can also ask, of some cognitive activity, *"how* does it occur?" This question does seem to invite an answer in terms of the coordinated interactions of various processes occurring during and described at a finer grain than the cognitive activity itself. For instance, were we to ask how it is that I am able to visualize my sister, one would expect a plausible answer to appeal to various neural or computational processes that occur while I am thinking of her. My visualizing evidently happens *by way of* certain neural processes (in part). So, another way of stating the relevant sense in which a higher-level phenomenon like cognition can arise out of lower-level phenomena like neural activity is to say that what one's brain does (in part) answers how cognition occurs. In what precise terms should

we understand this idea of the lower-level being an answer to a 'how' question? What clear examples of this interlevel explanation are there outside of the neural-cognitive domain?

At this point a short examination of the term "emergence" is in order, to clarify what connection there is between the views I am investigating and claims that the mind emerges from a continuous interaction between a body and its world. The term "emergence" appears to be used in a stronger and a weaker sense, although there is some disagreement even about these two meanings (Chalmers, 2006; J. Wilson, 2015; Winning & Bechtel, 2019). I take it that a *strongly* emergent phenomenon *cannot* be understood in terms of the coordinated activities of its lower-level basis. The only hope of relating a strongly emergent phenomenon to the domain of entities it emerges from, if there is any hope, is in terms of laws outside of any of the physical sciences – laws specifying the conditions for emergence in systems described by physical laws. An attractor in the state-space of a dynamical system, for example, is not a strongly emergent phenomenon, because it *is* understood in terms of (complex relations among) the lower-order components of the system. Strong emergence is not relevant here, because we are considering a basis for explanation that is, by definition, not achievable in the case of strongly emergent phenomena.

Weak emergence, I take it, is marked by unexpectedness. We find weakly emergent phenomena in a system when it exhibits patterned behaviors that surprise us even though we know everything that is happening in the system at a fine level of grain. Thus, weak emergence characterizes certain high-level entities *with respect to us as observers,* rather than characterizing the sense in which a higher-level entity arises from or is explained by its lower-level components. An attractor is only an example of a weakly emergent

phenomenon if it is not apparent from the description of the system's lower-order dynamics that it exhibits that attractor. Even though the lower-order description, if accurate, entails that the system exhibits whatever attractors it exhibits, these attractors may be entirely unrecognizable in that lower-order description. Thus, a new, "higher" (more abstract) level of description is introduced to represent the attractor and any relations it might stand in to other higher-order entities. But again, whether such a higher-order entity is weakly emergent depends on whether we could tell the system would exhibit that feature just from the lower order description. This means not all higher-order features of dynamical systems are weakly emergent. For example, the moon's orbit, because it is intuitive and familiar, is not normally considered a weakly emergent phenomenon.

As with strong emergence, weak emergence seems to connote the *absence* of explanation or understanding; to the extent that some higher-level feature is well understood in terms of a complex interaction among lower-order processes, it appears less deserving of the label "emergent." Admittedly, the term is sometimes used in an even weaker sense – one that applies to well-understood phenomena just so long as they are higher-order in some respect. In this, weakest sense, "emergent from" seems to say no more than "based in," in which case the term does not shed light on the relation in question – it just changes the question to that of asking what is distinctive about the "emergence" of cognition. What I mean to investigate here is the kind of explanatory relation that exists when we *do* gain some understanding about how a higher-level entity arises from lower-level processes – the kind of relation that we lack in the paradigm cases of emergent phenomena.

The hope is that there is a workable account that is less vague than explanatory "explanatory basis," that applies consistently across differing examples of higher-level features and that plausibly extends to what is going on in cognitive and neuro-physiological systems. A leading view of this explanatory relation is Constitutivism, which I state as follows:

> **Constitutivism**: Lower-level processes partly explain a cognitive process insofar as they are constituents of it; when properly coordinated, lower-level processes jointly constitute higher-level, cognitive phenomena.

This view is supported by philosophers who propose to analyze cognitive activities and other interesting objects of scientific observation as *multi-level mechanisms*, who are thus occasionally branded "New Mechanists" (Baumgartner & Gebharter, 2016; Bechtel & Abrahamsen 2005; Couch, 2011; Craver, 2001, 2007; Harbecke, 2015; Machamer et al., 2000). A mechanism is essentially comprised of parts, where the organized activities of these parts serves to explain the activity of the whole in the sense that they constitute it. To elaborate this framework in the context of a simple example, consider a knife as a mechanism for cutting things. One thing worth highlighting immediately is that what we are explaining is an activity, role, or capacity (of cutting), not the knife considered as a bounded physical object ("activity" and "capacity" are more or less interchangeable for the purposes of this discussion). To illustrate, if one were to crumple the knife into a ball and throw it into the sea, while there may be a coherent sense in which the "same physical object" as the original knife would be sinking into the ocean, the knife-qua-cutting-mechanism would have ceased to exist. (There is surely more that needs to be said about the boundary conditions of mechanisms, but not here). To this effect, Craver says "[t]here are no mechanisms simpliciter. There are only mechanisms *of* behaviors" (2007, p 11).

What we want is an explanation of these behaviors by appeal to lower-level processes, and this approach says to understand these lower-level processes as constituents.

So, continuing with the example, we might describe the knife as constituted by two components, a blade and a handle. Like the knife itself, these components are identified by what they do to contribute to the knife's cutting, not by their physical state or morphology. Let us suppose it is the blade's "being sharp" and the handle's "being holdable" that, when combined in the proper way, constitute the knife's capacity to cut. Of course, other conditions external to this mechanism are required for it to manifest the activity of cutting – in particular, someone needs to wield the knife – but under those conditions, the higher-level activity of "cutting" is constituted by the blade and handle together doing what they do. What we might learn about knives or cutting from this multi-level analysis is perhaps unclear, but the purpose here is just to reveal the generality of the Constitutivist framework. The example shows that mechanisms and their constitutive parts are ubiquitous. The capacities we are most interested in explaining surely involve many more components interacting in much more complex ways, but the basic notion of constitution applicable in the knife case is supposed to serve our understanding of those more interesting capacities.

To define this idea somewhat more formally, let us label an arbitrary mechanism *M*, its to-be-explained, higher-level activity *A*, its lower-level components P_1, P_2, …P_n with P_i standing for an individual component, and the activities of these components are B_1, B_2, …B_n with B_i standing for P_i's activity (see Figure 3). The Constituvist holds in general that *M*'s *A*-ing is partly Constitutively explained by P_i's B_i-ing, and wholly Constitutively explained by the coordinated activities B_1, B_2, …B_n of components P_1, P_2, …P_n. In the simple example above, the knife's (*M*'s) cutting (*A*-ing) occurs in virtue of the blade's

(P_1's) being sharp (B_1) and the handle's (P_2's) being holdable (B_2). The thesis at issue, for which I will raise trouble below, is that cognition is to be understood analogously in terms of cognitive mechanisms constituted by lower-level components.

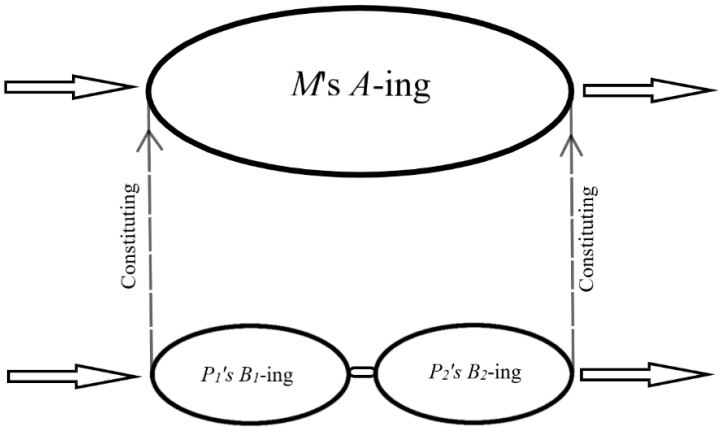

Figure 3
Diagram of Constitutivism. Large arrows represent the causal preconditions and effects of *M*'s *A*-ing. Double-bar between P_1 and P_2 represents an ordered, causal relation, which could be complex and feedback-involving. Vertical, dashed arrows represent the constitutive relation. These vertical arrows also signify physical locality, as if space-time proceeds left to right.

Note that the component activities are not merely aggregated but *ordered* so as to constitute the higher-level *A*-ing. Thus, it is possible for B_1, B_2, …B_n to occur out of order, in which case they do not constitute *A*-ing and so no *A*-ing occurs (assuming the containing system does not have a redundant mechanism for *A*-ing). You might have a sharp blade and a holdable handle – so P_1 is B_1-ing and P_2 is B_2-ing – but the handle is not affixed to the blade or is affixed in some strange way such that the would-be-knife is not actually usable for cutting. In order to Constitutively explain a mechanism, it is not enough to just list its components, rather one must also specify the relations those components must stand in so as to constitute the mechanism. This necessary ordering means that the components' contributions to the higher-level activity are dependent on one another – whether one

component does its part to help constitute the whole is bound up with whether the others do. Thus, the higher-level activity cannot be identified with a particular set of lower-level activities, and in this sense stands apart from them. This goes along with the idea that cognitive processes are multiply realizable – multiply constitutable, on this view. That is, *A*-ing can occur by way of different P_i's B_i-ing, and also, the same component processes B_1, B_2, …B_n might have multiple orderings that can constitute the same *A*-ing. Like a knife's cutting, the thought is, a cognitive mechanism's doing whatever it does can be made of somewhat different parts organized in somewhat different ways. All this is to attest to something important Constitutivism achieves; it specifies a sense in which cognition could be distinct or *sui generis* – really a "higher level" phenomenon than neural processes, even though it does depend on those neural processes.

3B. Against Constitutivism, For Generativism

A critical feature of the relationship between a mechanism and its constitutional components is that those component processes occur *within* the spatio-temporal boundaries of the higher-level processes they constitute. One way to put this is say that each P_i's B_i-ing must be "causally between the inputs and outputs" of *M*'s *A*-ing (Craver, 2007, 13). So, *M*'s *A*-ing, understood to have causal precursors (inputs) and effects (outputs), spatio-temporally includes all of its lower-level components. One could also say that, since the properly organized components make up the mechanism, the mechanism must extend wherever and whenever its components do. In short, *constitution entails containment*. What

the knife's blade does, qua component of a cutting mechanism, is contained in what the knife does. Suppose I am prompted to visualize my sister upon hearing someone say her name, and then I go on to think of the last time I saw her. It accords with Constitutivism to assume that whatever neuro-physiological processes that give rise to my visualizing activity occur while I am visualizing her, thus, after I hear her name and before I recall our last meeting.

The requirement that the lower-level processes are contained in what they give rise to is problematic because it is not met in certain dynamical systems. That is, a higher-order feature of a dynamical system can be spatio-temporally more local than the dynamics that give rise to it. A rough example of this can be depicted in a case of gravitational dynamics slightly more complicated than the case I described earlier. Instead of just two masses, imagine a cluster of gravitationally intertwined masses. Given the right initial conditions, it could occur that one mass moves in a repeating pattern or even stays motionless with respect to the disorderly meanderings of the masses around it. In such a case, the anomalous, cyclical movement of the one mass is a higher-order feature of the system that is shaped by gravitational relations spread throughout the system, though this movement itself is confined to a small part of the system. The lower-order components of this system – the masses and their gravitational acceleration towards each other – exert global influence on the system. If you remove a mass, it can affect the paths of masses all over the system, and if you pick a particular mass or spatio-temporal region to observe, what you observe is liable to be affected by a change anywhere else. In such cases the higher-level phenomenon does not contain, and so is not constituted by the lower-level processes that explain how it occurs.

A more relevant example might be recurrent patterns of activity in a neural network continuously impacting and getting feedback from an environment. Within an appropriately complicated network of this sort, we might find a few nodes collaborating in a repeating sequence of activity. While this pattern of activity is spatially local, it could be that connections between distant nodes or between nodes and parts of the environment are involved in the maintenance of the local cycle of activity. Again, in such a case, the combination of lower-order interactions that give rise to a higher-order phenomenon are not contained in it. In the next section I closely examine another such case, involving animal behavior. Whether the animal behavior in this case should be called cognitive is not something I want to debate here. In order to show that we need something other than a Constitutivist understanding of how cognition arises, it is enough to provide reason to think that some cognitive processes might have lower-order components whose spatio-temporal span exceeds the location of the cognitive processes themselves. literature that supports Dynamical Basis points in this direction, because cycles of feedback between internal processes and external effects of those processes are supposed to play an important role in explaining cognition. The case I discuss in section 4 points in this direction by analogy, because it demonstrates how global dynamics can give rise to complex adaptive behavior, cognitive or not.

It is not generally easy to show, for a given cognitive activity, whether it arises partly due to dynamics of processes outside of the brain, or rather occurs by processes entirely in the head. Settling this for any particular activity would require a near-complete description of the lower-level processes involved in it, which is not something we have for any interesting cases of cognition. That said, a general lesson from the research cited in

previous section is that the brain avoids doing work where it can. One could also say "offloading" onto processes in the environment as a rule, rather than an exception (Berry et al., 2019; Clark, 2008; Costa et al., 2011; Kirsh, 1995; Krueger, forthcoming; Risko & Gilbert, 2016). ("Offloading" has a specious, intentional ring to it, to my ears, but that is another matter). The point is, the processes that give rise to cognition significantly involve the structure of the brain's interaction with the body and environment. This means that if we accept Dynamical Basis then we should also accept Uncontained Dynamical Basis:

> **Uncontained Dynamical Basis**: Some lower-order components of the dynamical basis of a cognitive process are spatio-temporally more global than the cognitive process itself.

This notion of what the dynamical basis of cognition is like is incompatible with seeing the relevant lower-order components as constituents of cognition. If the dynamics of brain, body, and environment (at least sometimes) give rise to cognition in the above, uncontained way, then Constitutivism mischaracterizes the relationship in terms of which lower-level processes explain how cognition happens.

Below I will spell out an alternative view of this between-level explanatory relation, but first I want to briefly consider an objection to the argument I have offered against Constitutivism. If the dynamical basis of some cognitive activity partly occurs outside of the organismal body, perhaps this just means that the cognitive process itself is located partly outside of the body. Rather than concluding cognition does not contain all of its lower-level bases, this suggestion says one ought to conclude that the cognitive system is *extended* throughout the environment, thus containing the relevant lower-order processes there (Clark, 2008; Noë, 2005; R. Wilson, 2014). This view of cognition as literally extended in the environment indeed provides a way for Constitutivism to answer the charge

I presented, though only in an attenuated sense of "containment," as both higher- and lower-order processes are here taken to extend globally, throughout the system.

Additionally, there are problems with this view of cognition as extended. At first glance, some higher-order features of dynamical systems appear *not* to extend throughout the entire system even though they have global components. For instance, that the phenomenon of there being high- and low-tide extends to the moon, on this line of reasoning, since the moon's gravitational pull figures in the dynamical basis of the tides. That said, if one is ready to accept that the mind extends into our tools and environment and perhaps even others' behavior, then one may not find it so strange that the tides would be partly located a quarter of a million miles away. Anyway, this shows how radical a rethinking of the locations of physical phenomena is entailed by treating higher-order features of dynamical systems to extend wherever their explanatory bases do.

Further, seeing cognition as extended conflicts with idea of a single cognitive system that remains constant across scenarios that differ in terms of features of the environment but include roughly the same brain and body (Shapiro, 2019; M. Wilson, 2002). If we take an extended, "cognitive" system to be constituted by processes located in the environment as much as in an organismal body, then distinctions between cognitive systems look nothing in particular like distinctions individual, embodied entities; it no longer appears that we are talking about the kind of cognitive system we set out to investigate. Those who have tried to incorporate dynamical systems-based explanations within Constitutivism have not grappled with the problem I describe here, to my knowledge; they either do not appreciate the apparent locality of these higher-order phenomena with respect to their components, or they endorse claims about cognition being

broadly extended without critically reflecting on what more this entails (Bechtel & Abrahamsen, 2013; Kaplan & Craver, 2011). I have more to say, in the next chapter, about the role of an organismal body in our understanding of cognition. At this point I do not want to further discuss the problem of spatio-temporally locating cognitive processes. If my conclusions here hinge on whether better reasons can be offered elsewhere to reject the view of cognition as extended, so be it. I hope to have provided sufficient reason to think that some higher-level phenomena arise from dynamical bases that they do not contain, and that some cognitive processes are plausibly higher-level phenomena of this kind. This means that the explanatory relation at issue here is not one of physical parthood, that is, constituency. Lower-level components are explanatory parts in a different sense. I am now in a position to concisely articulate this different sense.

An express alternative to Constitutivism comes from Miracchi (2017), whose notion of *generative* relations partly inspires my account here. She argues that the kind of explanatory "basis" relation that figures centrally in cognitive science and neuroscience is generative explanation, which she delineates by contrasting generation with identity, causation, correlation, grounding and constitution. She describes generation as a species of *difference-making* relation (Strevens, 2019; Woodward, 2003). The basic idea is that science is generally about investigating "what makes a difference to what." Often scientific investigation uncovers what causes what, causal relations being one species of difference-making relation. Miracchi points to the way that science sometimes discovers what generates what – discovers what, synchronically and described at different scales, makes a difference to what. This generative relation makes no commitment about metaphysical necessity – The proposal is that, when it comes to cognitive processes, our search for

explanations in terms of lower-level components is a search for the generative bases of cognition. Call this view Generativism:

> **Generativism**: Lower-level processes partly explain cognitive processes insofar as they partly generate them; when properly coordinated, lower-level processes collectively generate higher-level, cognitive phenomena.

Just as Constitutivism is not a view about how cognitive processes are constituted, Generativism is not a view about how cognitive processes are generated. Constitutivism and Generativism are both views about the sense in which a cognitive process is explicable in terms of interactions among lower-level processes. An important difference between generative relations and constitutive relations is that the former do not imply anything about physical parthood, and so do not imply that the relevant lower-level processes are contained in what they help to generate. Recall the earlier case of single mass moving cyclically amid a collection of meandering masses, all of whose gravity effects this patterned movement. In the terms I have now introduced, we can say that such a higher-order phenomenon is generated by components spread throughout the system, even though phenomenon itself is local. More generally, higher-order features of dynamical systems are generated by the coordination of lower-order components, whether or not they are constituted by them.

There is no reason generative bases cannot be constituents at the same time; returning to an earlier example, we can now say the blade's "being sharp" partly generates the knife's cutting capacity, and is also constitutively involved in it. At the same time, an ornamental engraving on the blade might be constitutively relevant without being generatively relevant; various interventions on this engraving would make no difference to whether the knife cuts, but the engraving is physically involved in the cutting nonetheless.

Looking back to the start of the chapter, we can now say of the example of a cellular automaton that it is a dynamical system wherein the interactions among processes in a local area of the system are sufficient to generate the pattern of interest in that area. Generative relations can be the coordination of processes that neatly divide into spatio-temporal parts, or the coordination of processes whose influence is spread throughout an entire system, or combinations of both. We should not be worried about comparing the locality of higher-level phenomena to that of their lower order bases, because locality is not part of the explanatory relation in question. The relation is difference-making, at a time, from a lower-order to a higher-order description, that is, Generation.

I have tried to show the importance of this difference between Constitutivism and Generativism and argued that should adopt the latter in favor of the former because cognitive processes plausibly have uncontained dynamical bases. I will now try to illustrate the main points of my discussion more precisely in the context of an example of a dynamical systems model of a high-level, animal behavior.

4. Dynamic, Generative Bases of the Knifefish JAR

The details of this example are drawn primarily from work by Madhav et. al. (2013), which is part of a larger body of research on the Jamming Avoidance Response (JAR) of glass knifefish (*Eigenmannia virescens*). Madhav et. al. offer a dynamic systems-based model of the knifefish's behavior, and by representing their model in terms of the

view I have developed here, I hope to make the notion of dynamically generated cognition more tangible.

Glass knifefish emit an electric discharge of a variable frequency and are sensitive to certain features of the electric field in their immediate surroundings. The interaction between their electric discharge and objects around them affords them information about the location of those objects. The knifefish use this interaction between electric fields to navigate around objects in their environment, some of which are neighboring glass knifefish emitting electric discharges of their own. If two or more knifefish are close in proximity and the frequencies of the electric charge they generate are similar, the way that the charges interfere with one another obstructs the fish's ability to navigate. A knifefish is able to successfully navigate partly because it is able to shift the frequency of its electric discharge in the direction that increases the difference between its frequency and that of a neighbor, thereby avoiding the problematic "jamming." This shift is called the "Jamming Avoidance Response" or JAR.

Madhav et. al. (2013) put individual knifefish in an enclosure with recording electrodes and an electric signal generator. This allowed experimenters to register the frequency of knifefish's discharge and then feed an incoming charge into the enclosure, controlling the difference in frequency between the knifefish's outgoing charge and the incoming charge that it senses, closing the loop, so to speak. The researchers used the setup to probe the dynamics of the JAR and to develop a differential equation model of it, where the parameters of the equation describe lower-order processes they can observe and intervene on. According to their model, the rate of change of the knifefish's discharge frequency is equal to a sum of two terms (see Equation 1); one term is a function of the

currently emitted discharge frequency, the other is a function of the difference between that outgoing discharge and the incoming frequency. The equation relating these functions to the derivative of outgoing frequency specifies how the that frequency changes over time. (No deep grasp of the math below is necessary to follow my line of reasoning. I will refer to the important elements of the equation in verbal terms, that is, in terms of frequencies, the magnitude of difference between frequencies, and rates of change of frequencies).

Equation 1 $$\frac{dY(t)}{dt} = -Y(t) + E(Y(t) - U(t))$$

A simplified version of Madhav et. al.'s (2013) global non-linear model of the knifefish JAR. The output frequency of the fish is given by *Y*, the input frequency the fish senses is given by *U*, and *E* represents a function specific to this model system. The equation here is missing elements that would scale the values appropriately and represent the anatomical upper and lower limits on *Y*(*t*), which the reader can take to be implicit.

I want to point out two higher-order features of the dynamic system modeled by Madhav et al. First, there is the JAR behavior itself. This investigation of the knifefish is prompted by the observation of the knifefish's macro-scale capacity to "avoid jamming." This phenomenon is not picked out by any particular elements of the description of the system given by Equation 1. The JAR can be described in terms of a feature of phase flow, specifically by a *repellor* (the opposite of an attractor). In this case, the repellor is a part of the system's state-space where the difference between incoming and outgoing charges is sufficiently small, which we label "jamming," from which all nearby trajectories veer away. Equation 1 logically implies that the system's trajectory veers away from the "jamming" parts of the state-space (given a wide range of initial conditions), but it does not furnish us terms to specify the difference between "jamming" cases and "non-jamming" cases. Thus, to identify "jamming" cases as such, we use a level of description that picks out the commonality in an array of veering, counterfactual trajectories.

Another interesting, higher-order feature of this system is what Madhav et al. (2013) call a "snap-through." To appreciate this, first note that, in principle, the knifefish can avoid jamming by either emitting a frequency sufficiently greater, or sufficiently less than the incoming frequency it senses – it can dodge up or down, so to speak. This is reflected in the fact that, given an incoming frequency, the dynamical system typically has two, stable equilibria and Equation 1 has two stable values for $Y(t)$. However, because there are upper and lower bounds on the frequencies that a knifefish can emit, there are some incoming frequencies close to these bounds where the system only has one equilibrium solution. Thus, as the knifefish's outgoing frequency is continuously driven up (or down) by an incoming frequency that continuously ramps up (or down), the system's trajectory follows the nearer of the two equilibria until eventually that equilibrium disappears, and the system exhibits a sudden shift to the remaining one. In other words, the system's state "snaps-through" to the other side of the repellor. Like the JAR itself, this characteristic feature of certain JARs-over-time is a higher-order feature, not something represented in terms of Equation 1 or its components. In fact, the snap-through was not initially observed by the authors, but rather was a prediction following from their model, which they experimentally confirmed after coming up with the model (ibid., 4279).

What are the lower-level components of the JAR? Equation 1 amounts to a lower-level description of the whole system, so we can identify elements in the equation as potential components. The authors of the model explicitly identify two components of the JAR; the two terms that are summed on the right side of Equation 1. (Perhaps a useful, finer-grained set of components could be identified, or perhaps we should think the addition/subtraction operation marks a privileged decomposition – in any event, this two-

component picture will serve present purposes). According to Madhav et al., these two terms represent competing "sensory" and "motor" components. The sensory component is the function of the difference between immediate incoming and outgoing frequencies, $E(Y(t) - U(t))$, which they call the "escape" term, and it reflects the tendency for outgoing frequency to change in proportion to its proximity to incoming frequency. The motor component, $-Y(t)$, which they call the "return" term, reflects the tendency of the outgoing charge to settle toward some pre-stimulus equilibrium.

Note that the contributions of these components to the behavior of the whole system are global. Whatever state the system is in, both components' values go into the determination of the rate of change of outgoing frequency, and so also into the determination of the rate of change of the components' values themselves. However, the JAR itself is not similarly spread throughout the entire state-space of the system. The JAR occurs specifically when incoming and outgoing frequencies, $U(t)$ and $Y(t)$, are close enough together that the knifefish needs to adjust its outgoing frequency. When $U(t)$ and $Y(t)$ are very distant, $E(Y(t) - U(t))$ becomes a negligible value, so $Y(t)$ does not change appreciably in response to an incoming signal, so there is no threat of "jamming" to be avoided and we observe no JAR. This suggests the JAR does not contain the sensory and motor components as spatiotemporal parts, and so is not constituted by them. Rather, the JAR is generated by the interaction of the components under certain initial conditions. It is perhaps even clearer that the snap-through is more local than its explanatory bases. The snap-through only occurs when outgoing frequency reaches an upper or lower limit, but *how* it occurs is determined by the two basic components that are operative at all times

over all states of the system. The dynamical, generative bases of the snap-through extend beyond the location of the snap-through.

Conclusion

Dynamical systems models offer a powerful lens for investigating the workings of cognitive systems and have important implications for the kind of explanatory relation we rely on to understand cognition in terms of lower-order components. We must attend to the implications about locality that come along with a view about the sense in which cognition "arises" from a system of dynamic relations. Often we deal with systems whose macroscopic functional properties *can* be decomposed along neat, spatio-temporal lines, where constitutive relations and generative relations overlap – often but not always. I have tried to support the thought that a cognitive system is liable not to be explicable in terms of constituent parts, but rather in terms of dynamic, generative bases.

4 The "Body" in Embodied Cognition

1. Introduction

I defend a version of the claim that cognitive processes are to be understood partly in terms of the body of the cognitive agent. The term "Embodied Cognition" names a school of thought that broadly tries to reveal the way cognition is shaped by the body (Wilson & Foglia, 2017). Much of the research associated with this school focuses on the way perceptual and motor processes are involved in cognition. Embodied Cognition theorists broadly agree that the body is explanatorily relevant to cognition insofar as sensory-motor machinery (on the body, outside the brain) plays an important role in cognition. Whether the body is relevant in any further sense and what exactly we mean by the "body" are questions that have not garnered attention in this discourse. I hope to remedy this by setting down an account of the body that is designed to figure in a larger understanding of cognitive processes. I frame this account by articulating the key insights of the other approaches to the embodiment of cognition, which I aim to expand on. I then describe a basic model of the "Body" – a theoretical construct for use as part of a theory of cognition as embodied. The Body, on this view, is a kind of dynamically generated, self-organizing boundary that meets three formal conditions; Size Asymmetry, Contribution Asymmetry, and a Transaction Condition. I argue that the body is explanatorily relevant to

cognition in that the conditions for the arising of a cognitive system include the conditions for the arising of a body.

I proceed as follows: in section 2 I clarify the principal claims about the embodiment of cognition that set the stage for the proposal here. These include a critical stance toward a "disembodied" view of cognition, and a view about interdependent processes of perception and action as giving rise to cognition. In section 3, I turn to the explanatory role of the body, looking beyond its role in sensory-motor mechanisms. After clarifying my question and my methodology, I articulate a theory of the Body in terms of characteristic patterns of self-organization. I then show how this notion of Body is relevant to understanding cognition as embodied. In sum I offer a view of how we can understand perception, action and cognition as higher-order activities of certain kinds of Bodies, and I argue that doing so helps reveal the basic character of, and relationship between perception, action, and cognition.

2. In Pursuit of Embodiment

2A Disembodied Cognition and Internal Symbols

The sense of "embodiment" I want to spell out here is standardly construed in a way that opposes a view of cognition in terms of computational processing over internal symbols. To provide some context, I want to briefly state the main ideas to which Embodied Cognition theories critically respond. I will refer to the contrasting view based

on internal symbols as the "classical" one. (Fodor, 1975, 1980; Haugeland, 1978; Simon & Newell, 1971; Sternberg, 1969; Stich, 1983)

The computational role of a classical symbol – what processes it figures in and how – is determined by the set of rules governing transitions between configurations of symbols, that is, the syntax. Such symbols are therefore "internal" in the following sense: what a symbol putatively does to enable cognition can be understood just in terms of the system of symbols and syntax it figures in. Any facts about the world outside of this computational system are extraneous to understanding how a symbol allows the system to reach its proper output states from input states. Thus, a hallmark of such symbols is that a symbol's form is arbitrarily related to whatever it is a symbol for. A computational process is *abstract* in the sense that its elements and the relations between them are independent from the structure of whatever physical processes might realize the computation. In other words, one can understand how a computational process works without understanding how it might be concretely manifested in the world (Kaplan, 2011; Piccinini, 2015). That said, an important part of the classical symbolic view of cognition concerns the physical basis of the symbolic computations that putatively make up cognition; the relevant computations are thought to be physically realized by brain activity (Churchland, 1989; Dayan & Abbott, 2001; Eliasmith & Anderson, 2003; Gauthier et al., 2019).

Thus, classical symbols are internal in a second sense – in the sense of being "in the head." These doubly internal symbols are at the root of the view of cognition that seems objectionably "disembodied" to some philosophers. These two senses of the internality of internal symbols typically come as a pair, but it is important not to conflate these two senses, because they introduce distinct challenges. Summing up here, the point is that the

"disembodied" view of cognition that forms a backdrop for this investigation of embodiment asserts the conjunction of *Functional Internality* and *Physical Internality*:

> **Functional Internality**: Cognition happens by way of operations over abstract symbols whose relevance is determined by the syntax and state of a computational system.
>
> **Physical Internality**: The physical realization of these symbols occurs in the brain.

The internal symbols at issue here are, by definition, *amodal*. The body and nervous system have various modes of interaction with the external environment, and an internal symbol is amodal if none of these modes of interaction determine the structure of the symbol. We can think of the relevant modes as distinct channels or dimensions that make up one's sensory-motor engagement with the world (or crudely, we can just think of the canonical five senses). Processes that essentially involve sensory or motor modalities are, on the classical view, peripheral to cognition per se. Sensory and motor systems may rely on symbol-processing as well, but the symbols that figure in cognition are thought to abstract away any information to do with modes of sensory-motor engagement; see, for example, Fodor's (1983) distinction between "Input Modules" and "Central" systems. In other words, the structure of sensory-motor interaction with the world is external with respect to the syntax that governs the symbol-processing underlying cognition, given Functional and Physical Internality.

Functional Internality and the amodal nature of classical symbols are implicated in the *symbol grounding problem*, which a proper understanding of embodiment should solve or avoid (Barsalou, 2008; Glenberg & Robertson, 2000; Harnad, 1990; Müller, 2009; Searle, 1980). Because symbols are arbitrarily related to the environment of the physical system that realizes them, the problem is to account for how any of these symbols could be

about anything about the world that a cognizer acts in. In a computational systems we design, we can interpret internal symbols as referring to objects in the world according to our own purposes or linguistic conventions. However, these methods of fixing symbols' meanings do not work in the case of a cognitive system, because the symbols there are supposed to have meaning for the system itself. Thus, Functionally and Physically Internal symbols appear cut off from the kind of semantics that cognitive states have, because those symbols are cut off from the body's engagement with the world. That said, some views that seem broadly "classical" try to address the symbol grounding problem, partly in terms of symbols in sensory processing (Fodor, 1995). It therefore seems to me that a focus on "embodiment," insofar as it suggests a significant shift in the landscape of cognitive science, must go farther than saying that the extra-cranial nervous system is a non-negligible part of the relevant symbolic computer.

2B Interdependence of Perception and Action

In the above, critical reaction to the view of cognition as computation over internal symbols, one finds an emphasis on the way cognition is bound up with processes of perception and action. Perceptual-motor engagement is a major focus of work on the explanatory relevance of the body and the use of dynamical modeling approaches to cognition. Specifically, some have developed the insight that the *interdependence* of perception and action is important to understanding cognition. I describe this idea here in order to build on it further below.

Some actions obviously depend on perception in a weak sense. For example, if I look around the grocery store and I do not see the apples (and do not then ask anyone to point them out), I will not be able to buy them. But from this we would be wrong to conclude that apple-buying generally requires visual perception. Even if we are convinced that I must have *some* (perhaps distant or vicarious) perceptual access to the apples in order to buy them, this seems a far cry from saying that my action "is bound up with" or "essentially involves" my perception. Similarly, perception obviously depends on action in this weak sense. If I do not go to the grocery store, I will be unable to see the apples there (barring visual communication technology). The notion of dependence at play in these examples is causal and is contingent on several quite specific features of the scenario. The notion of dependence at play in the claim that perception and action are interdependent must be a deeper and more general one than this, for the claim to be of any consequence (Adams & Aizawa, 2001; Aizawa & Journal of Philosophy, Inc., 2007; Block, 2005). For the moment, let us just consider the vague claim that perception and action "fundamentally structure" each other and must therefore be understood in terms of each other. This idea has been spelled out in several ways and there does not appear to be consensus among Embodied Cognition supporters about how widespread this binding of perception and action is, and how important it is to understanding cognition generally.

One way this interdependence might be spelled out is in terms of the representations a cognitive system uses, and what determines the content of these representations. This notion of *representations* is itself controversial in this context and merits its own brief discussion. Having rejected internal symbols as the basic processing unit of a cognitive system, one might still think that there is another, properly "embodied" way of

understanding representations as crucial to how cognition works (Barsalou, 2008; Clark, 1998; Damasio et al., 1994; Glenberg & Robertson, 2000; Grush, 2004; Shapiro, 2019; Thompson, 2007; Wheeler, 2005; M. Wilson, 2002). On the other hand, it is central to some views of embodiment that cognition is not to be understood in terms of representations at all (Brooks, 1991; A. Chemero, 2011; Degenaar & Myin, 2014; H. L. Dreyfus, 2002; Hutto & Myin, 2013; O'Regan & Noë, 2001; Van Gelder, 1995). Complicating matters further, some describe an important role for a certain kind of "maps" or "schema" or even "symbols" in the brain but deny that these entities should be understood as representations (Gallagher, 2005; Johnson, 2017; Varela et al., 1991). This, I take it, is a serious point of dissent within Embodied Cognition; the theoretical camp is divided about whether and in what sense representations are a part of explaining cognition. This is a complicated topic and it is not one of my aims here to mend this divide. In order to describe certain lines of thought within Embodied Cognition I will sometimes use the term "representation" meaning to pick out whatever version of the notion is compatible with this framework, if there is one.

Getting back on track, one way to express how perception and action might be interdependent is via the following claim, which I call "Interdependent Contents:"

> **Interdependent Contents**: The contents of perceptual representations refer partly to the perceiver's capacities for action, and the contents of the representations of actions refer partly to the agent's perceptual capacities.

Millikan presents a clear version of this claim. She distinguishes between "descriptive" and "directive" representation (Anscombe, 1957/2000) and contends that certain representations "face both these ways at once" ((Millikan, 1995, 187). She calls

these "pushmi-pullyu representations." By way of example, she points to how the food call of a hen both describes the location of food and directs her chicks to come and eat the food. Her claims may not imply any version of Embodied Cognition, but the basic idea of representations fulfilling dual, descriptive-directive roles brings us toward a general way of seeing how perception and action might be essentially bound together.

Those who argue for the interdependence of perception and action tend to draw on James Gibson's approach to understanding visual perception (1966, 1979). Gibson develops an influential notion of *affordances*, which are, in short, possibilities for action constituted by features of an organism's environment. For example, for humans and other organisms with sufficiently similar bodies, an apple typically affords grasping and eating, and a rigid surface typically affords standing-on. But an animal with a different body and different capacities for sensation and locomotion will be met with different affordances in its environment; whether anything at all is graspable, edible, or stand-on-able depends on the animal. Thus, affordances also seem to "face both ways" in the sense I introduced above, in that they describe certain facts about an organism's environment while they are understood relative to capacities or interests of the organism. Gibson himself offers an anti-representationalist understanding of the role of affordances in perception and action (1979), as have others (Chemero, 2001; Noë, 2005; O'Regan & Noë, 2001; Turvey, 1992). On the other hand, some have supported representationalist accounts, wherein affordances make up part of what organisms perceptually represent (Clark, 1998; Hatfield, 1991; Kelly, 2010; Siegel, 2014). These approaches similarly maintain that perceptual capacities cannot be understood apart from the possibilities for action, and this places bodily motion at the fore of the investigation of the mental. That said, note that Interdependent Contents expresses a

thesis about perception and action, not about *cognition* per se. More must be said to connect this claim about interdependence to a view of cognition as embodied.

2C Perceptual-Motor Basis Loops

Consider the claim that cognitively guided behavior is a result of the following three-stage process: first, receive information about the environment, second, internally process that information to reach a conclusion as to the appropriate action, third, execute the appropriate action. On this picture, the intentional movements of a cognitive being are the outputs of internal computations that received sensory information as inputs. Less formally, this view says that *thinking* is essentially what happens after seeing and before moving. Hurley (2001) poignantly names this the "classical sandwich" view and defends an alternative that centers on the interdependence of perception and action. She argues that cognition emerges as a higher-order feature of processes that systematically "loop" from perception to action, back to perception, again to action, etc. In other words, complex, feedback-involving interactions between relatively sensory and relatively motor processes are the basis of perception and action simultaneously, and cognition is not freestanding from perception and action but rather arises from their interplay. Thelen and Smith (1994) and Clark (1998) propose a similarly foundational role for "action loops" of this sort. An example I find usefully illustrates the idea is that of performing a choreographed dance. Even though there is a sense in which the entire sequence of moves is known ahead of time (if one has practiced enough), performing is generally not just a matter of implementing

pre-planned movements. The dancer is constantly adjusting the force and direction of their motions in light of where and when they find themselves, relying on their own proprioceptive feedback. Visual and auditory feedback are also crucial – one may be entirely unable to perform without seeing a mirror or hearing the music. Thus, the thinking is not done when the dance begins, but is rather bound up in the coordination of perceptual and motor processes that the performance involves. See Kirsh (2011) for a more thorough discussion of cognition in choreography, and Montero (2016) for an account of expertise that describes similar forms of perception-action interdependence in phenomenological and neurophysiological terms.

This approach has a close analogue in control systems theory, in engineering. In order to make an artificial agent with the ability to navigate around obstacles and reach goals in its environment, it can be quite efficient for that system to have actuators whose activity is a function of the activity of sensors, and vice versa. A relatively simple archetype of this is put forward by Beer (2003), whose model agent relies continuously on its own movement to discriminate between objects to be avoided and objects to be "caught" (see also, Brooks 1985). This line of research in engineering is closely tied to the use of dynamical systems theory to model the agent-environment interactions in question. Beer (1995) spells out broadly how a dynamical systems approach can guide the design of state-of-the-art, mobile robots. Reverdy and Koditschek (2018) develop a dynamical systems-based approach to prioritizing and coordinating the goals of a mobile robot, and further derive the conditions (and perturbations) under which the robot will successfully navigate to its goals. Some philosophical accounts propose, more boldly, that the mind should be understood in terms of properties of dynamical systems in general. (Gelder & Port, 1995;

Kelso, 1995; Kelso et al., 2013; M. D. Lewis, 2005; Spivey, 2008; Warren, 2006). There is also a cluster of recent work supporting the analysis of particular biological, neural and cognitive systems in dynamical systems terms (Braud et al., 2018; Brette, 2019; Christoff et al., 2016; Cowan et al., 2014; Shenoy et al., 2013)

In the previous chapter I outlined the key features of dynamical systems model that are relevant here, which had to do with the way lower-order components give rise to higher-order phenomena in dynamical systems. There, I argued that lower-order components can contribute to the state-changes of the system in a *global* way even when they give rise to higher-order phenomena that are relatively local; lower-order, dynamical components are not always contained in the phenomena they generate. This suggests a way to localize cognitive processes without insisting that the system of interactions that they arise from are similarly local, thereby painting a very different picture than that implied by Functional and Physical Internality. If we explain cognition in terms of interdependent processes of perception and action connected in complex feedback relations, we can better appreciate the significance of the body and the environment in the structure of cognition (Clark, 1998; Hurley, 2001). To pursue an embodied view of cognition then, we should frame our investigation of cognition according to this idea, which I define as Basis Loops:

> **Basis Loops**: Cognition arises from coordinated interactions among cyclical processes of perception and action.

This claim leaves room for substantial debate about exactly how involved the body outside of the skull is in cognition. It implies that cognition cannot be understood completely in isolation from perception and action, but it allows for weaker and stronger interpretations of how extensive the ties between thinking, perceiving, and acting are.

On a weak reading, much of cognition could be explainable in terms of symbols realized by neural activity in the head, just as long as the functional roles of these symbols are shaped by an organism's perceptual and motor capacities. One may accept Basis Loops and yet hold that the only cognitive abilities we should expect to subsume *ongoing* interactions between perception and action are very basic ones – especially those involving bodily movement that occur over short time-spans, under variable conditions. When it comes to cognition involving highly abstract or distal properties, one might think the significance of sensory and motor machinery is rather remote. Consider, for example, the ability to reason about justice, plan a large event, or solve logic puzzles. Clark and Toribio (1994) coined the term "representation-hungry" to describe such domains of cognition, suggesting that it is difficult to see how one could possess such abilities without relying on brain-based representations of some sort. Some have tried to account for our competence at such tasks in terms of representations in the brain that serve as *predictions* or *simulations* of potential effects on the sensory and motor parts of the nervous system (K. J. Friston & E. Stephan, 2007; Gallagher & Allen, 2018; Gentsch et al., 2016; Grush, 2004; Hohwy, 2013). Such a representational scheme depends on the cognizer's capacities to perceive and act (albeit somewhat indirectly), and so it is compatible with Basis Loops and with the idea that simpler cases of cognition might not such involve brain-based representations.

On a stronger reading of Basis Loops, all cognitive activity occurs by way of bound-up processes of perception and action *as they occur*. On this view, the sensory and motor capacities of the body must play a role in our understanding of all cognition – brain-based predictions or simulations of what one might perceive or how one might move cannot make up a cognitive process on their own. Recall my earlier discussion of how the thinking

process in the performance of dance seems to manifest as in the structure of a back-and-forth between how one is moving and what one perceives. One could say, in broad strokes, that cognition in such cases involves a coupling between the sensory-motor capacities of the body and the relevant features of the environment – the spatial layout, the music, etc. Such a case can arguably be understood entirely in terms of interacting processes of perceptually-guided action and action-guided perception. As I have said, it is a matter of dispute whether the dynamic coupling of sensory-motor capacities to the world can explain how we cognitively engage with such abstract entities as elections and logical operators.

So, extant accounts suggest different ways that Basis Loops might be further specified, but it has broad support as a focal point for thinking about the sense in which cognition is embodied. My brief analysis of the thinking process involved in dance also supports Basis Loops, and to the extent that it captures the structure of a broad range of cognitive capacities, it supports a stronger reading of Basis Loops. Clark's (1998) discussion of people's methods of solving puzzles of various kinds has an analogous structure; one goes back and forth between attending to the puzzle board and attending to an individual piece, perhaps rotating it or setting it in a designated area for later attending. Again, movements and perceptions are intertwined in a way that the puzzle-solver depends on for the task. The analogy could be imperfectly extended, for example, to a large writing project. One's ability to complete the project involves the ability to write the first section without precisely knowing how the remainder will shape up, ad reviewing one's work and making changes to what was written or outlined along the way. This shows that feedback relationships between what we could loosely call information-gathering and information-

producing activities, functioning in the process of a larger information-producing activity, are arguably very general phenomena.

That said, in the example of a large writing project, the stages of information-gathering and -producing I identified are not perceptions and bodily motions, strictly speaking, but are themselves quite abstract cognitive achievements. If we hope to connect Basis Loops to such achievements, we might zoom in further, on the basis of the basis…of those abilities. After all, we might say, one does need to put fingers to keyboard and look at a screen in order to write even the first section. Except that one can also use pen and ink (though one shudders to think of it), complicating the question of what explanatory relevance one's hands and eyes have in the capacity to write a book. Such complicated matters demand complicated models and, as I mentioned above, there is not general agreement about the role of sensory-motor processes in tasks that demand sensitivity to distal and abstract properties. What I want to highlight here is that the investigation of embodiment thus far centers on issues to do with the content of representations, the relationship between perception and action, and a dynamic, environment-inclusive picture of the explanation of perception, action, and cognition. Focused on these issues, the discourse has overlooked the body except insofar as it is the bearer of sensory-motor machinery. I now propose to consider the body's explanatory relevance in a further sense.

3. Explanatory Role of the Body

3A What Body?

The idea that cognition must be understood in terms of interdependent processes of perception and action suggests that the body is explanatorily relevant, but it does so indirectly, via the fact that sensory-motor machinery is on the body, not just inside the head. Supposing one accepts Basis Loops, one might wonder whether this is the only sense in which the body is explanatorily relevant. Simply put, for the purposes of understanding cognition as embodied, by "the Body" do we mean anything other than the bearer of perceptual and motor capacities? I officially denote the question as follows:

> **Body's Relevance**: Is the body explanatorily relevant to cognition aside from the way it shapes sensory-motor engagement with the environment?

I am unaware of any explicit arguments for an affirmative answer to Body's Relevance, but to my knowledge this matter has not been taken up in earnest in discourse about the embodiment of cognition. Little theoretical attention has been given to a conception of the body that stands apart from perceptual-motor activity. This strikes me as unfortunate, both because I think it is an interesting topic and because I think the answer to Body's Relevance is "yes." I therefore aim to productively complicate the discussion of embodiment by bringing the question to the fore, and aim to develop a view that understands the Body's relevance in a deeper sense. The proposal here is to treat the Body as an explicitly defined, theoretical entity that supports a larger view of cognition, so I capitalize the term (and when I mean to refer to bodies in an intuitive, pretheoretical sense, I will not capitalize). As I am delving into unexplored territory, my exposition will be somewhat condensed. With luck, future work will examine this matter more rigorously.

Informally, one may recognize a body as a contiguous, skin-bound entity that includes all the matter that makes up an animate being at a time. Of course, cells within the Body are constantly being shed and replaced from without. Even sudden, major changes to limbs or organs do not seem to amount to the destruction of the Body or the acquisition of a new one. Thus, a specific or persistent physical composition is not a necessary feature of the Body. It is not obvious to me that even contiguity is essential, at least for the purposes of explaining cognition. Nothing seems to rule out the possibility that the Body of an individual cognitive being could comprise unattached parts. Arguably a colony of ants makes for an intuitive case of a kind of scattered Body. It does seem indispensable that the relevant sort of Body is one that is *alive* or self-animating. A dead "body" cannot exhibit cognition, so it is no Body at all, in this context. Thompson (2007) and Johnson (2017), among others, foreground the connection between cognition and life. At times these authors seem to treat "living organism" and "body" as synonymous terms, but this is still too vague to cut cloth when it comes to the Body's Relevance. One wants an account of what it is for a body to be living, and how that is explanatorily relevant to cognition.

An important idea I propose to build on in accounting for the Body is that of *self-organization* in a complex system whose parts interact over time. Others rely on this notion in their understanding of embodiment (Bruineberg & Rietveld, 2014; A. Chemero, 2011; Jirsa & Kelso, 2004; Kelso, 1995; Varela et al., 1991) and the formal notion itself goes back much further (Ashby, 1953; Von Foerster & Zopf Jr, 1962). Relatively simple examples of self-organization can be found in fluid systems. Such systems have a fine-grained structure that is typically described in terms of interactions between particles (molecules), and they can exhibit macroscopic structure, that is, patterning or order in the

way many particles interact over time. A vortex ring, for instance, generated in the wake of a sweeping fin, is a self-organizing pattern of activity of fluid particles. A vortex ring is not identified in terms of the specific particle-interactions that physically realize it at a specific space and time. Although some such particle-interactions are necessary in order for a vortex ring to actually arise, a huge and diverse range of particle-interactions-over-time realize the pattern we identify as a vortex ring. One could say the "vortex-ring-ness" we observe is not so much a characterization of the particle-interactions as it is a characterization of the *order* of the system of fluid particles. Thus, the presence or absence of a vortex ring is a higher-order feature of fluid systems. More specifically, the vortex ring is higher-order with respect to lower-order features described at a smaller temporal and spatial scale – in this case, the positions, properties, and forces between the particles that make up the fluid. In the terms of the previous chapter, we would say the lower-order interactions figure in the generative basis of the vortex ring; that is the sense in which they are explanatorily relevant, lower-level processes. What might this have to do with the Body's Relevance?

I have articulated a view on which at least some cognitive processes are best understood as higher-order phenomena generated by system of dynamic relationships among processes of perception and processes of action. I want now to suggest certain other dynamic relationships are also necessary to cognition. I propose we understand cognition as a kind of self-organizing structure, where lower-order dynamics do not just involve processes of perception and action but processes associated with the Body in a more basic sense. To develop an intuition for this, some Bodily processes we might consider are those that turn ingested food into energy used by perceptual-motor systems (partly to obtain and

digest more food). We might also consider the bodily structures that protect internal organs from airborne contaminants and collisions, like skin, bones, immune system, and certain reflexive behaviors. Note that the maintenance of these protective structures is also dependent on the continued functioning of the internal organs they protect. Also, in complex bodies like ours, the proportions of various compounds must be kept within specific ranges by processes whose dynamics maintain homeostasis (Cannon, 1929; Cooper, 2008; Gershenson & Fernández, 2013). These examples suggest a rough notion of a self-organizing Body as a coordinated ensemble of processes. Also, the processes I just described do not involve perception and action. Such Bodily processes do not appear to represent or relate a perceiver-agent to external objection, which processes of perception and action do, I take it. These processes also continue, basically uninterrupted, when one falls asleep. In short, (merely) Bodily processes lack the basic, world-involving character of perception and action. If processes like these are relevant to understanding cognition, then the Body's relevance goes beyond its identity as a sensory-motor machine.

I will offer a basic model for understanding the Body as a kind of higher-order activity generated by the self-organizing dynamics of certain systems. First, I want to briefly clarify my methodology. I propose to start by considering a few, fairly undisputed generalizations that apply to the bodies involved in all of the cognitive systems we have come across. I will use these general observations to inform a basic, formal account of the Body in terms of properties of a dynamical system. Then I will look back at cases, exploring the implications of the formal account and checking for surprising results. After that I will be ready to consider what this account suggests about the Body's relevance for understanding cognition.

3B The Dynamically Generated Body

One basic thing to note about the bodies of cognitive systems is that they involve a *boundary* with reference to which we can distinguish "internal" from "external." A major thing this boundary seems to do is keep internal parts on the inside and external parts on the outside; a boundary that fails to do this will swiftly fail to be associated with a cognitive system. Another feature of such a Bodily boundary is that the external area is much larger than the internal area; bodies persist in an environment of which they are a small part. However, within that small part are a host of Bodily processes involved in the maintenance of the boundary, like the examples I mentioned above. The working order of these internal parts is of primary importance to the way a body manages to stay around. Some external processes are important to the body's staying around too. In particular, sources of energy used by the internal processes need to get from the external to the internal side of the boundary. These are an exception to the general rule of preventing boundary-crossings. This makes for a first-pass description of the kind of bodies cognitive systems seems to involve – now for a formal characterization of the Body.

First, a dynamical system that generates a Body must include a boundary between internal and external. Any measurable, spatially local discontinuity in the behavior of lower-order processes in the system can make for such a boundary. Self-organizing boundaries are easy enough to come by in nature – a difference in polarity, as in that between oil and water, does the trick. The vortex ring example I discussed earlier also involves an identifiable separation between internal and external parts of the system. To model the Body, I propose we use systems that exhibit such a boundary and meet three

further conditions; Size Asymmetry, Contribution Asymmetry, and the Transaction Condition.

Size Asymmetry says that the external part of the system is much larger than the internal part. It need not specify an exact ratio; requiring that the external part is several orders of magnitude larger than what the Bodily boundary is safely inclusive of all the relevant bodies we have encountered. An apparent counterexample or borderline case might be that of a fetus in the womb. Such a system involves a coordination of processes that maintain a skin-boundary, but the "internal" part seems to take up most of the space in the system. Looking more closely though, this is not an adequate way to identify the relevant, larger system. We should say that a model of a fetus could not accurately describe how it works over any significant period of time without appealing to processes external to the womb and even the mother. So, such a model would exhibit Size Asymmetry after all. However, this suggests a possible case of what I will call a "Megabody," which sustains itself in a universe not much larger than itself. If we adopt my proposal here about how understand cognition partly in terms of the Body, we will end up with a view of cognition that could not apply to a Megabody. It might seem surprising that there could not be a Megabody that exhibits cognition in anything like the sense we do. If this is an unintuitive implication though, it is not an overly concerning one. Until we have a physically plausible cause of a cognitively active Megabody to consider, the model of the Body I propose offers a good fit for our purposes. If the account of embodied cognition based on this Body is ultimately successful, it will help us make sense of the fact that cognition always is found to occur via small bodies acting in large environments even though cognition might not seem to have anything inherently to do with size.

Contribution Asymmetry says that processes on the inside of the boundary make a larger *contribution to the order* of the system. The relative contributions of various processes within the system can be compared in terms of difference-making relations of the generic sort the scientific experiments uncover, discussed in previous chapters (Woodward 2003, Strevens 2004). To illustrate, consider a traveling vortex ring, where particles are swirling with high angular velocity on the inside, and slowly drifting by on the outside of the boundary. Now imagine we can choose to inject a small force into the fluid on the inside, or at an equal distance from the boundary on the outside. If our intervention is forceful enough, it will cause the vortex ring to promptly dissipate regardless of whether the injection is located on the inside or outside of the boundary. However, there is a range of interventions that will destroy the vortex ring if they take place on the inside and not if on the outside. There are presumably also interventions with the opposite character – certain ways to change the flow of fluid immediately outside the vortex ring that destroy it even though the same perturbation would be tolerated on the inside. To the extent that a self-organizing boundary is more robust to external changes than internal changes a similar distance from the boundary, it exhibits Contribution Asymmetry.

I take it vortex rings do tend to exhibit Contribution Asymmetry in some minimal extend, but I have not run the relevant experiments and the precise robustness conditions of a vortex ring are not important here. If we compare a vortex ring to single-celled organism we will find the degree of Contribution Asymmetry far greater in the latter case – enough to make the vortex ring case look roughly symmetric. There is a huge array of minor modifications one could make to an internal part of the cell that would ensure that it

quickly ceases its self-maintaining activities, and comparably very few modifications that do this and must occur outside the cell. The array of ways to destroy a vortex ring is much less complex, and what complexity there is does not coalesce so predominantly in the internal part of the system. I propose we specify that, in the case of a Body, the contribution to order of internal processes is several orders of magnitude larger than that of external processes. This gives us a model that does not apply to vortex rings. Like the Megabody, a vortex ring falls nearby but outside of the class of systems characterized by the two Asymmetries just described. Putting things together, a self-organizing boundary that exhibits these two Asymmetries is a local phenomenon, generated by potentially global dynamic relations, where the ensemble of processes whose coordination generates the boundary mostly occur within the boundary. This already seems to be a description of living bodies and little else.

However, it is possible for a system to conform to the two conditions above and yet starkly differ from the sort of bodies typically found alongside cognitive beings. Specifically, we can imagine a system wherein the self-maintenance of a boundary is achieved without any processes crossing that boundary. The ensemble of internal processes could be mutually sustaining in a self-sufficient way, so that internal parts of the system only ever interact with other internal parts, and likewise for external parts. It seems universal that the bodies of cognitive beings are not perfectly independent in this way, but rather depend on intake processes to gather resources that internal, body-sustaining processes use up, and depend on processes that expel substances that build up. The **Transaction Condition** says that a system exhibits a Body only if some process regularly crossing the boundary is a component of the self-organizing dynamics of the system. This

condition entails that, if Transactional, boundary-crossing processes are prevented, there is some internal component of the system that precipitously increases or decreases in value until the Body is destroyed. A system that exhibited Size and Contribution Asymmetry but did not meet the Transaction condition would be what I call an Undying Body. Given the fact of increasing entropy in physical systems, we can straightforwardly see why the Transaction Condition would be met by any physical body that satisfies Size and Contribution Asymmetry – that is, why we find no truly Undying Bodies in nature. The complexity of the system's internal, boundary-sustaining processes entails that the internal part of the system must maintain a relatively orderly, low-entropy state. If this part of the system maintains this low-entropy state despite entropy continuously increasing everywhere, it must be because of an interaction crossing the boundary that lowers internal entropy at the cost of increasing external entropy. The Transaction Condition is not about entropy per se though; what it means is that the Body, although it generally involves keeping the insides inside and the outsides out, depends critically on regular transactions with its environment.

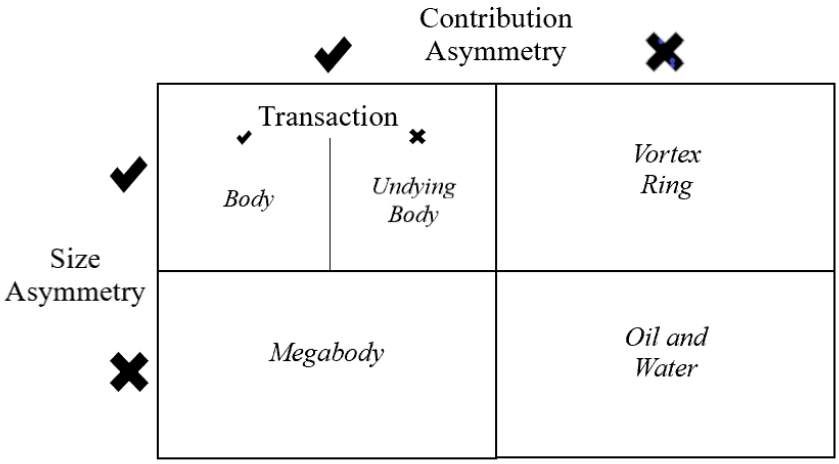

Figure 4
Classification of self-organizing boundaries according to whether they meet conditions of a Body.

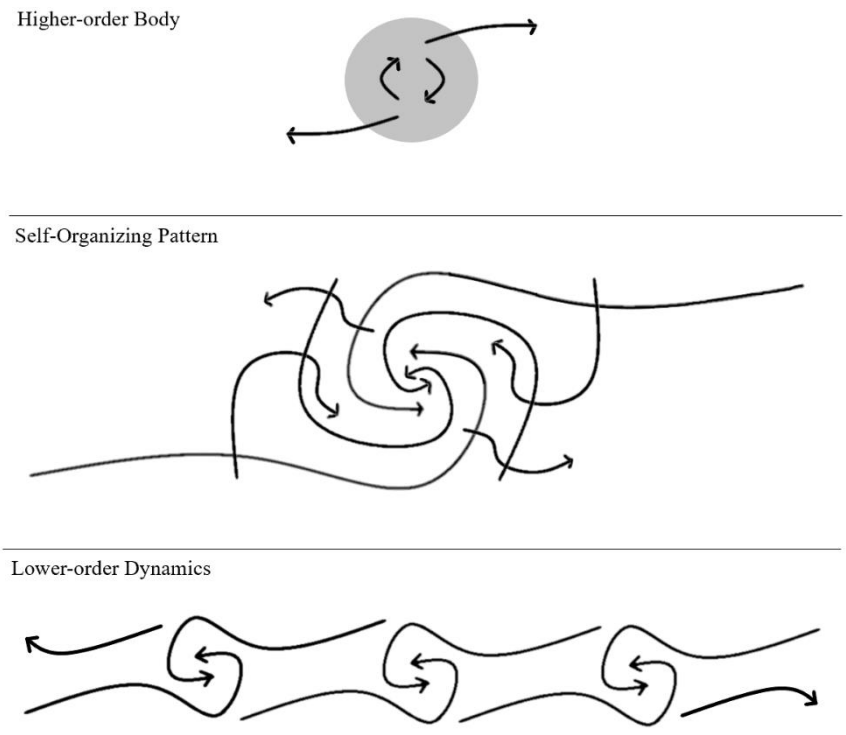

Figure 5
Diagram of self-organizing Body. Arrows in the bottom level represent how the system is changing at the smallest scale. The middle level portrays the way lower-order parts organize to form a self-maintaining structure. Note that some components extend into or from the space surrounding the local pattern, whose boundaries are vague at this level. The top level depicts the higher-order process generated by these lower-order dynamics, that is, a Body. Arrows in the top level represent internal and Transactional processes.

3C Fleshing Out the Body's Relevance

On this account, a Body is a higher-order feature of a particular kind of dynamical system; one featuring a self organizing boundary containing the internal parts of the system, which take up a much smaller area than external parts but do much more as a part of maintaining the system's Bodily order, and where some components of the self-organizing boundary cross that boundary. Note that such a notion of Body is not subject to ship-of-Theseus style challenges; it is precisely the manner in which the Body is ever-changing, ever rebuilding itself using parts of its environment, that is at the heart of this account. One way this notion of the Body departs from the way others have appealed to dynamical systems is that other accounts tend describe a cognitive system as a open dynamical system (Kelso, 1995; Thompson, 2007; Varela et al., 1991). The relevant dynamical system in this picture is the whole, Body-environment system, which is *closed*, and which generates the Body as a local, self-maintaining order of the system. Crucially for my purposes here, this Body is not identified in terms of its relation to processes of perception and action. My claim here is that a Body of this sort is relevant to understanding cognition, because we should understand cognition as arising in Bodies of a certain kind. On this view, the dynamic basis from which a cognitive process arises is also the dynamic basis from which the Body of the cognizer arises.

Basis Body: Cognitive processes are higher-order activities of a Body.

This, I contend, is an important sense in which the Body is explanatorily relevant to cognition. This depicts a basic sense in which cognition is plausibly embodied, which

has nothing directly to do with perception and action. Not everything apt to be modeled as a kind of Body is a cognitive being. Single-celled organisms, for example are Bodies that seem not to perceive or intentionally act, much less think. Basis Body is compatible with a diversity of views about perception, action, and cognition, because it says nothing about what conditions a system must exhibit, beyond those involve in the Body, in order to be an Embodied Perceiver-Agent or an Embodied Cognizer. To wrap things up, I will more precisely spell out a version of a view that incorporates Basis Body, adding detail to this account of the Body's Relevance. To do this requires making broad claims about the nature of perception, action, and cognition, and in less space than I have devoted to the Body. Therefore, my discussion here will be speculative, but I hope it helps to depict the thoroughgoing sense in which cognition can be understood in terms of the Body.

The complex, mutually sustaining processes involved in a Body make the system's order robust to sudden perturbations of certain of its parts, to a certain extent. This is true even of the vortex ring; a small enough injection of force to the internal parts will not destroy the ordered pattern of activity. However, a Body exhibits robustness along many more dimensions than a vortex ring given the greater complexity of the ensemble of processes involved. Further, one can define a *measure of stability* as a function of the relationships among internal processes that tolerate a range of perturbations. Stability is also varies over time, as the robustness conditions of these processes varies depending on the system's constantly changing state. As internal activity levels are caused to rise or fall, the system may not be able to withstand perturbations that it previously could – the Body becomes vulnerable to a broader range of order-breaking effects. Such a measure of stability is implicated in the Transaction Condition; stability decays in the absence of

necessary Transactional processes, which replenish it. Similar notions of stability and robustness figure prominently in applications of dynamical systems models to robotics, discussed in section 2C, above (Cowan et al., 2014; Reverdy & Koditschek, 2018).

The point of defining the sense in which a Body exhibits varying stability over time is that it provides a norm with respect to which internal processes can be classified as functional or malfunctional. For instance, a regular Transactional processes that stabilizes a Body that has somewhat decayed is functional for the Body, whereas a process that intakes material that is unneeded and just gets in the way of other internal processes, destabilizing the Body, is malfunctional. Note that this analysis does not assume that the Body's most common states – the dynamics that tend to prevail when nothing unusual is happening in the environment – are its most stable ones. The norm here is not historical or statistical. It could be that the Body's default or normal order, given the relevant initial conditions, does not involve the most stable arrangement of inner processes. In other words, it could be that some perturbations increase stability. The norm of stability is immanent; the Body needs to be stable in order to be at all, over time. In this sense, the Body as dynamically generated comes with its own definition of what its parts are *for*.

The notion of the immanent normativity of Bodily processes puts one in position to articulate a profound role for the Body in the structure of cognition. Again, there are multiple ways this could be spelled out and I will just offer one here. My basic suggestion is that there is a reasonably clear way to distinguish mere Bodies from Embodied Perceiver-Agents. Within the class of systems that generate a Body, one can further categorize systems according to the character of the Bodily motion involved. A Body might not have to move through its environment, as its stabilizing, Transactional processes might occur

with the Body just sitting or drifting along aimlessly in its environment. That is, external processes might regularly create the necessary conditions for Transaction. On the other hand, in order to maintain stability, a Body might rely on internal processes to propel the entire Bodily boundary through its surroundings, to particular locations. That is, the self-organization of a Body might or might not involve *directed locomotion*. Those Bodily systems that do involve directed locomotion can exhibit many more dimensions of stability-impacting behaviors than those that do not. This is because such a Body's position over short time-scales relative to distant parts of the environment is an important component of stability, dramatically complicating the space of perturbations to which the Bodily system will stabilize in response.

Insofar as the Body exhibits *directed* locomotion, internal locomotor processes must vary according to which direction of motion is stabilizing. This means that certain internal processes of a locomoting Body must reliably co-vary according to the Body's distance and direction from various external parts of the environment. In other words, for locomoting Bodies, certain components of the Bodily system must function to carry information transmitted from external processes to internal ones. This describes a special kind of Transactional process. Consider a more general Transactional process that replenishes energy resources that have decayed – in a word, ingestion. What such a process does to help sustain the Body can be identified in terms of relations to other internal processes, because their contribution is local to those processes. Ingestive processes involve external (or better, external-to-internal) objects, so they do carry information about the environment. However, one can understand the functional role of such processes just in terms of how they modify other internal processes involved in maintaining the Bodily

boundary; their function is not defined by the information they carry. Transactional processes specific to locomoting Bodies function to carry information about locations in the environment because that information must be used for motion that is directed toward or away from those locations. Bodies that engage in directed locomotion have, by definition motor components whose activity must be robust to the position-at-a-time of distant objects, so they must rely on sensory components to carry information about those objects, in addition to motor components that cause the Body to move. In other words, Bodies that maintain stability partly by moving around rely on interdependent processes sensation and locomotion. Such bodies seem to exhibit the normative, world-involving character of perception and action – they provide a plausible model of the Embodied Perceiver-Agent.

To get a better handle on this account, we can apply it to some examples. To consider clear cases first, we humans fit this model of an Embodied Perceiver-Agent, and a living cell that picks up nutrients drifting along in the surrounding medium, lacking any system of propulsion, is a mere Body. A borderline, quasi-locomotive case can be found in organisms that rely on small stinging cells called cnidae to catch prey, like coral and jellyfish. In short, such a creature sends out a small, sharp part of its body out into the (nearby) surrounding space in a way that is sensitive to the location of an external process (whether prey is nearby). To do this requires internal processes that function to carry information about and move with respect to something external. But cnidarians do not need to move the whole of their body – they do not need to move anything farther than a small fraction of their own body's size, and they do not need to carry information about anything more distal than the reach of these small stinging cells. We can contrast these creatures with the oft-discussed case of bacteria that rely on magnetosomes (Barsalou, 2008;

Dretske, 1986; Kelso et al., 2013; Ruth G. Millikan, 1989; Piccinini, 2018; Pietroski, 1992). These bacteria contain processes that carry information about the direction of the Earth's pole, which is the adaptive direction to move in, or what my account describes as the stabilizing direction. Supposing the movement patterns that characterize directed locomotion involve the whole body, this view of the Embodied Perceiver-Agent applies to magnetosomes but not to cnidarians.

One might ask, "Why the whole Body?" – why does the movement of small parts a small distance not count as directed locomotion? Recall the reasoning behind defining the model in a way that rules out Megabodies, vortex rings, and Undying Bodies. Given pretheoretical observations of what are commonly taken to be bodies involves in cognition, I looked for conditions to describe a class of dynamical systems that generate a higher-order phenomenon that behaves like bodies do. Similarly, in order to cleanly distinguish Embodied Perceiver-Agents from mere Bodies, I want to specify a feature that a subset of Bodies exhibit in an extreme degree, making them distinctive. This feature is a many-fold increase in robustness that comes from relying on stabilization processes that involve locomotion, versus a Body that passively receives all of its vital Transactional processes. The range of motion and sensation involved in Cnidarian predation is so small that it does not exhibit this feature. As compared to a strictly internal, ingestive process, Cnidarian predation does not, by its nature, have robustness conditions that are orders of magnitude more complex. Having said all this, the primary point here is not whether cnidarians perceive, but how the preceding account of the Body figures in a larger understanding of embodied perception and action.

There are several questions not answered by the kind of model I am offering. For instance, it is not clear, based on what I have said, how to individuate processes of perception and action, or how to determine the perceptual and motivational content we should ascribe a Body depending on how those processes work. It is beyond my scope to investigate those questions here. Now that I have sketched a distinction between mere Bodies and Embodied Perceiver-Agents, the final pertinent question is this: what makes the difference, if there is one, between an Embodied Perceiver-Agent and an Embodied Cognizer?

Just as the Perceiver-Agent, on this view, is dynamically based in Bodies of a special sort, the Cognizer can be understood in terms of a characteristic ordering of perceptual-agential activities over time. Recall that the Bodies of Perceiver-Agents involve stabilization processes that require directed locomotion, which involves being robust to different ways the Body is positioned vis-à-vis external processes. We can thus describe the Perceiver-Agent as engaged in a kind of meta-stabilization. The relevant stabilization processes might, for example, be one of moving to and consuming food. The intertwining of information-gathering and locomotor processes that the Body relies on make the food-acquiring process robust with respect to a wide array of perturbations, so perception and action provide ways of stabilizing the Body's (internal) stabilization process. Taking this approach one step further, my proposal is to understand cognition as involving an additional layer – as a meta-meta-stabilization that we roughly think of as *learning*.

To illustrate, I will build on the earlier example of the magnetically sensitive bacterium, supposing it to be as simple a case we might find that fits this Embodied Perceiver-Agent model. Imagine we can intervene in such a system by placing

magnetically charged objects in the environment, call them "distractors." Now let us imagine further that some magnetically sensitive process in the bacterium varies smoothly and monotonically as a function of *both* the direction of the magnetic pole and the direction of a nearby distractor, simultaneously. In other words, some internal process carries information about the direction of both external entities – the pole (movement toward which is functional), and the distractor (movement toward which is malfunctional). One could visualize this internal sensor as something like a small compass that, rather than just pointing at the nearest pole, wobbles in a complicated way that can depend on multiple magnetic fields.

Even if the bacterium contains a doubly-sensitive process like I just described, the distractor might nonetheless utterly baffle the bacterium, rendering its movements aimless or worse. This would be the case if this internal process were not connected to the bacterium's means of locomotion in a way that allows for guidance according to the information it carries about both of these sources. Further, the bacterium might be baffled by the distractors no matter how many times it encounters them (I assume that this is the case with actual magnetically sensitive bacteria). However, it might not be this way (stretching realism somewhat for the purposes of clarity); it could be that the bacterium is not guided by this information-carrying structure at first, but eventually, after it repeatedly is sent wandering by these encounters, its internal dynamics are somewhat affected according to whether it wanders into more or less nutrient-rich waters. If an internal process is affected by this feedback *and* by the yet-to-be-useful doubly-sensitive structure (within an appropriate timeframe), the bacterium could develop patterns of motor activity that are guided by this information. In short, by learning to use information about the direction of

the distractor vis-à-vis the planet's magnetism, the bacterium would learn to swim directly where it would if there were no distractor; it would learn to ignore the distractor.

When an Embodied Perceiever-Agent is able learn in above sense – to acquire new perceptual or locomotor abilities, or relations among them – their underlying stabilization process is robust across an exponentially larger array of conditions than if the Body were stuck with whatever perceptual-motor abilities it has at a time. If one learns from experiences with a certain kind of obstacle how to promptly locomote around that obstacle to an intended destination, then the set of movements one can make toward that goal becomes far more complex. In other words, the array of paths that stabilization might involve spans many more dimensions than if the Body cannot learn to stabilize in a variety of ways. I want to suggest that an Embodied Perceiver-Agent that learns – that explores the space of possible patterns of perception and action, acquiring new functional processes characteristic of its individual experience – is an Embodied Cognizer.

I have just described learning in terms of the acquisition of new perceptual-motor paths to a particular goal, but this is arguably not the only important sense of learning here. One can also think of learning as the acquisition of wholly new capacities or new goals. For instance, one might "learn to write" in a sense that goes beyond merely "acquiring a new means of verbally communicating," and pursuing the goal of writing something or of writing well might not be aptly described as pursuing any goal that an Embodied Agent has always possessed. This more dramatic kind of learning, where it is not an agent's means that expand, but their basic capacities and aims, might also be an essential part of what it is to be an Embodied Cognizer. If that is the case, the approach I have been outlining will need to support some way of accounting for this kind of learning in terms of the

characteristic behavior of certain kinds of Bodies. A more detailed examination of perceiving, locomoting, and learning by a Body will have to be pursued in subsequent work devoted to those topics – I have completed the inquiry I set out to here. I hope to have offered an understanding of the Body that can serve as a productive foundation for understanding cognition.

Conclusion

My main goal has been to advance the study of what role the body plays in our understanding of cognition. I discussed prevailing appreciation for the body's role in structuring capacities of perception and action, and then I sought to uncover a more general sense of the body's relevance. I proposed we understand the body in terms of a kind of self-organizing activity, characterized by an asymmetry in size, an asymmetry in the location of processes involved in the self-organization, and a transactional relation between internal and external parts of the system. My basic suggestion was that cognition is to be understood as embodied in the sense of arising in systems with a Body of a particular sort. To develop this suggestion further, I described the kind of stabilization achieved by a Body's internal processes, determining a normative standard with reference to the Body and thus providing for a kind of inward-looking, proto-intentionality. I described how this kind of stability can be meta-stable in a Body that can engage in directed locomotion, bringing informational relations to external objects into the space of what the Body's

stabilization processes are robust to. This offers a way to understand perception and intentional motion as a species of Bodily activity. Finally, I suggested that a further layer of meta-stability achieved over the course of experience lets us model a Body as learning, and that the characteristically flexible and individualized patterns of perception and action that arise in a learning Body are distinctively cognitive. This, I propose, captures a foundational sense in which the Body is explanatorily relevant to cognition.

Bibliography

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*(1), 43–64.

Aizawa, K., & Journal of Philosophy, Inc. (2007). Understanding The Embodiment of Perception: *Journal of Philosophy*, *104*(1), 5–25.

Anscombe, G. E. M. (2000). *Intention*. Harvard University Press.

Ashby, W. R. (1953). Design for a Brain. *British Journal for the Philosophy of Science*, *4*(14), 169–173.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645.

Barwise, J., & Seligman, J. (1997). *Information Flow: The Logic of Distributed Systems* (Issue 3, pp. 397–401). Cambridge University Press.

Baumgartner, M., & Gebharter, A. (2016). Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. *British Journal for the Philosophy of Science*, *67*(3), 731–756.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421–441.

--- (2013). Thinking Dynamically About Biological Mechanisms: Networks of Coupled Oscillators. *Foundations of Science*, *18*(4), 707–723.

Beer, R. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, *72*, 173–215.

--- (2016). The Dynamics of Active Categorical Perception in an Evolved Model Agent: *Adaptive Behavior*.

Beer, R., & Williams, P. L. (2015). Information Processing and Dynamics in Minimally Cognitive Agents. *Cognitive Science*, *39*(1), 1–38.

Berry, E. D. J., Allen, R. J., Mon-Williams, M., & Waterman, A. H. (2019). Cognitive Offloading: Structuring the Environment to Improve Children's Working Memory Task Performance. *Cognitive Science*, *43*(8).

Block, N. (2005, May 1). *Action in Perception by Alva Noë*. The Journal of Philosophy.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765.

Braud, R., Pitti, A., & Gaussier, P. (2018). A Modular Dynamic Sensorimotor Model for Affordances Learning, Sequences Planning, and Tool-Use. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(1), 72–87.

Brentano, F. (1874). *Psychology From an Empirical Standpoint* (Issue 2, p. 241). Routledge.

Brette, R. (2019). Neural coding: The bureaucratic model of the brain. *Behavioral and Brain Sciences*, *42*.

Brooks, R. A. (1991). New approaches to robotics. *Science (New York, N.Y.)*, *253*(5025), 1227–1232.

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, *8*, 1–14.

Butler, J. (1736). *The analogy of religion, natural and revealed, to the constitution and course of nature. To which are added two brief dissertations: I. Of personal identity. II. Of the nature of virtue*. London, Printed for J., J., and P. Knapton.

Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological Reviews*, *9*(3), 399–431.

Chalmers, D. J. (2006). Strong and weak emergence. In P. Davies & P. Clayton (Eds.), *The Re-Emergence of Emergence: The Emergentist Hypothesis From Science to Religion*. Oxford University Press.

Chemero, T. (2001). What we perceive when we perceive affordances: Commentary on Michaels (2000), _Information, Perception and Action_. *Ecological Psychology*, *13*(2), 111–116.

--- (2011). *Radical Embodied Cognitive Science*. MIT Press.

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews. Neuroscience*, *17*(11), 718–731.

Churchland, P. S. (1989). *Neurophilosophy: Toward a Unified Science of the Mind-brain*. MIT Press.

Cichy, R. M., & Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends in Cognitive Sciences*, *23*(4), 305–317.

Clark, A. (1998). *Being There: Putting Brain, Body, and World Together Again*. MIT Press.

--- (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press.

--- (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press USA.

Cohen, J., & Meskin, A. (2006). An objective counterfactual theory of information. *Australasian Journal of Philosophy*, *84*(3), 333 – 352.

Cooper, S. J. (2008). From Claude Bernard to Walter Cannon. Emergence of the concept of homeostasis. *Appetite*, *51*(3), 419–427.

Costa, A. J., Silva, J. B. L., Pinheiro-Chagas, P., Krinzinger, H., Lonnemann, J., Willmes, K., Wood, G., & Haase, V. G. (2011). A Hand Full of Numbers: A Role for Offloading in Arithmetics Learning? *Frontiers in Psychology*, *2*.

Couch, M. B. (2011). Mechanisms and Constitutive Relevance. *Synthese*, *183*(3), 375–388.

Cowan, N. J., Ankarali, M. M., Dyhr, J. P., Madhav, M. S., Roth, E., Sefati, S., Sponberg, S., Stamper, S. A., Fortune, E. S., & Daniel, T. L. (2014). Feedback control as a framework for understanding tradeoffs in biology. *Integrative and Comparative Biology*, *54*(2), 223–237.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, *68*(1), 53–74.

--- (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press, Clarendon Press.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, *72*(November), 741–764.

Damasio, A. R., Damasio, P. of N. A. R., & Damasio. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. G.P. Putnam.

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science*, *44*(1), 43–64.

Dawson, M. R. W. (1998). *Understanding cognitive science /*. Blackwell,.

Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press.

Degenaar, J., & Myin, E. (2014). Representation-hunger reconsidered. *Synthese*, *191*(15), 3639–3648. JSTOR.

Dennett, D. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.), *Minds, Machines and Evolution*. Cambridge University Press.

Dretske, F. (1981). *Knowledge and the Flow of Information* (Vol. 92, Issue 3, pp. 452–454). MIT Press.

--- (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, Content, and Function* (pp. 17--36). Oxford University Press.

--- (1988). *Explaining Behavior: Reasons in a World of Causes* (Vol. 100, Issue 4, pp. 641–645). MIT Press.

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

--- (2002). Intelligence without representation – Merleau-Ponty's critique of mental representation The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences*, *1*(4), 367–383.

Eliasmith, C. (2005). A new perspective on representational problems. *Journal of Cognitive Science*, *6*, 97–123.

Eliasmith, C., & Anderson, C. (2003). Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems. In *Cambridge, MA*.

Fodor, Jerry A. (1975). *The Language of Thought*. Harvard University Press.

--- (1980). Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioral and Brain Sciences*, *3*(1), 63–73.

--- (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Issue 2, pp. 289–293). MIT Press.

--- (1995). *The Elm and the Expert: Mentalese and Its Semantics*. The MIT Press.

--- (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology* (Vol. 51, Issue 205, pp. 549–552). MIT Press.

Frank, T. D., Profeta, V. L. S., & Harrison, H. S. (2015). Interplay between order-parameter and system parameter dynamics: Considerations on perceptual-cognitive-behavioral mode-mode transitions exhibiting positive and negative hysteresis and on response times. *Journal of Biological Physics*, *41*(3), 257–292.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Friston, K. J., & E. Stephan, K. (2007). Free-Energy and the Brain. *Synthese*, *159*(3), 417–458.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Clarendon Press.

Gallagher, S., & Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, *195*(6), 2627–2648.

Gardner, M. (1970). Mathematical Games. *Scientific American*, *223*(4), 120–123.

Gauthier, J., Loula, J., Pollock, E., Wilson, T. B., & Wong, C. (2019). From mental representations to neural codes: A multilevel approach. *The Behavioral and Brain Sciences*, *42*, e228.

Gelder, T. van, & Port, R. (1995). *Mind As Motion*. MIT Press.

Gentsch, A., Weber, A., Synofzik, M., Vosgerau, G., & Schütz-Bosbach, S. (2016). Towards a common framework of grounded action cognition: Relating motor control, perception and cognition. *Cognition*, *146*, 81–89.

Gershenson, C., & Fernández, N. (2013). Complexity and information: Measuring emergence, self-organization, and homeostasis at multiple scales. *Complexity*, *18*(2), 29–44.

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, *193*(2), 559–582.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1–19.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, *43*(3), 379–401.

Godfrey-Smith, P. (1992). Indication and adaptation. *Synthese*, *92*(2), 283–312.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*(3), 377–396.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, *156*(3), 563--585.

Haken, H., Kelso, S., & Bunz, H. (1985). A Theoretical Model of Phase Transitions in Human Hand Movements. *Biological Cybernetics*, *51*, 347–356.

Harbecke, J. (2015). Regularity Constitution and the Location of Mechanistic Levels. *Foundations of Science*, *20*(3), 323–338.

Harman, G. (1983). Problems with Probabilistic Semantics. In A. Orenstein & R. Stern (Eds.), *Developments in Semantics* (pp. 243–237). Haven.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346.

Hatfield, G. (1991). Representation in perception and cognition: Connectionist affordances. In W. Ramsey, S. P. Stich, & D. Rumelhart (Eds.), *Philosophy and Connectionist Theory* (pp. 163--95). Lawrence Erlbaum.

Haugeland, J. (1978). The nature and plausibility of Cognitivism. *Behavioral and Brain Sciences*, *1*(2), 215–226.

Hintze, A., Edlund, J. A., Olson, R. S., Knoester, D. B., Schossau, J., Albantakis, L., Tehrani-Saleh, A., Kvam, P., Sheneman, L., Goldsby, H., Bohm, C., & Adami, C. (2017). Markov Brains: A Technical Introduction. *ArXiv:1709.05601 [Cs, q-Bio]*.

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press UK.

Hurley, S. (2001). Perception and action: Alternative views. *Synthese*, *129*(1), 3–40.

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content* (pp. xxv, 206). MIT Press.

Izhikevich, E. M., Conway, J. H., & Seth, A. (2015). Game of Life. *Scholarpedia*, *10*(6), 1816.

Jacob, P. (2019). Intentionality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University.

Jirsa, V. K., & Kelso, S. (Eds.). (2004). *Coordination Dynamics: Issues and Trends*. Springer-Verlag.

Johnson, M. (2017). *Embodied Mind, Meaning, and Reason: How Our Bodies Give Rise to Understanding*.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*(3), 339–373.

Kaplan, D. M., & Craver, C. F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science*, *78*(4), 601–627.

Kelly, S. D. (2010). The normative nature of perceptual experience. In B. Nanay (Ed.), *Perceiving the World* (p. 146). Oxford University Press.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior* (pp. xvii, 334). The MIT Press.

--- (2008). Haken-Kelso-Bunz model. *Scholarpedia*, *3*(10), 1612.

Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Networks*, *37*, 120–131.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, *73*(1--2), 31–68.

--- (2010). Thinking with the Body. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, *T*, 176–194.

Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Kraemer, D. M. (2015a). Against “soft” statistical information. *Philosophical Psychology*, *28*(1), 139–147.

--- (2015b). Natural probabilistic information. *Synthese*, *192*(9), 2901–2919.

Krueger, J. (forthcoming). Music as Affective Scaffolding. In D. Clarke, R. Herbert, & E. Clarke (Eds.), *Music and Consciousness II*. Oxford University Press.

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement* (pp. xxxi, 481). Lawrence Erlbaum Associates, Inc.

Lewis, D. K. (1969). *Convention: A Philosophical Study* (Vol. 20, Issue 80, p. 286). Wiley-Blackwell.

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences*, *28*(2), 169–194.

Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25.

Marr, D. (1982). *Vison* (Issue 3). W. H. Freeman.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463--502). Edinburgh University Press.

Millikan, Ruth G. (1984). *Language, Thought, and Other Biological Categories*, MIT Press.

--- (1995). Pushmi-Pullyu Representations. *Philosophical Perspectives; Atascadero, Calif.*, *9*, 185.

--- (2001). What has Natural Information to do with Intentional Representation? *Royal Institute of Philosophy Supplement*, *49*, 105–125.

--- (2004). *Varieties of Meaning: The 2002 Jean Nicod Lectures* (Issue 3, pp. 674–681). MIT Press.

--- (2007). An Input Condition for Teleosemantics? Reply to Shea (and Godfrey-Smith). *Philosophy and Phenomenological Research*, *75*(2), 436–455.

--- (2017). *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford University Press.

Miracchi, L. (2017). Generative explanation in cognitive science and the hard problem of consciousness. *Philosophical Perspectives*, *31*(1), 267–291.

--- (2019). A competence framework for artificial intelligence research. *Philosophical Psychology*, *32*(5), 588–633.

Montero, B. G. (2016). *Thought in Action: Expertise and the Conscious Mind*. Oxford University Press UK.

Müller, V. C. (2009). Symbol Grounding in Computational Systems: A Paradox of Intentions. *Minds and Machines*, *19*(4), 529–541.

Neander, K. (2017). *A Mark of the Mental: A Defence of Informational Teleosemantics*. MIT Press.

--- (2018). Teleological Theories of Mental Content. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018). Metaphysics Research Lab, Stanford University.

Noë, A. (2005). *Action in Perception* (Vol. 102, Issue 5, pp. 259–271). MIT Press.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences*, *24*(5), 939–973; discussion 973-1031.

Papineau, D. (1987). *Reality and Representation* (Vol. 100, Issue 1, pp. 109–111). Blackwell.

Piccinini, G. (2015). Physical Computation: A Mechanistic Account. In *Physical Computation*. Oxford University Press.

--- (2018). Computation and Representation in Cognitive Neuroscience. *Minds and Machines*, *28*(1), 1–6.

Pietroski, P. M. (1992). Intentionality and TELeological Error. *Pacific Philosophical Quarterly*, *73*(3), 267–282.

Pylyshyn, Z. W. (1984). *Computation and Cognition*. MIT Press.

Reverdy, P., & Koditschek, D. (2018). A Dynamical System for Prioritizing and Coordinating Motivations. *SIAM J. Appl. Dyn. Syst*, *17*(2), 1683–1715.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688.

Rockwell, T. (2005). Attractor Spaces as Modules: A Semi-Eliminative Reduction of Symbolic AI to Dynamic Systems Theory. *Minds and Machines*, *15*(1), 23–55.

Scarantino, A., & Piccinini, G. (2010). Information without truth. *Metaphilosophy*, *41*(3), 313–330.

Schöner, G. (2008). Dynamical systems approaches to cognition. In *The Cambridge handbook of computational psychology* (pp. 101–126). Cambridge University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science (New York, N.Y.), 210*(4471), 801–803.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication* (pp. vi, 117). University of Illinois Press.

Shapiro, L. (2019). *Embodied Cognition*. Routledge.

Shea, N. (2007). Consumers Need Information: Supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research, 75*(2), 404–435.

Shea, N., Godfrey-Smith, P., & Cao, R. (2018). Content in Simple Signalling Systems. *British Journal for the Philosophy of Science, 69*(4), 1009–1035.

Shenoy, K. V., Sahani, M., & Churchland, M. M. (2013). Cortical control of arm movements: A dynamical systems perspective. *Annual Review of Neuroscience, 36*, 337–359.

Siegel, S. (2014). Affordances and the Contents of Perception*. In B. Brogaard (Ed.), *Does Perception Have Content?* (pp. 51–75). Oxford University Press.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145–159.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*(1), 1–23.

Spivey, M. (2008). The Continuity of Mind. In *The Continuity of Mind*.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421–457.

Stich, S. P. (1983). *From folk psychology to cognitive science: The case against belief* (pp. xii, 266). The MIT Press.

Strevens, M. (2019). Explanation, Abstraction, and Difference-Making. *Philosophy and Phenomenological Research*, *99*(3), 726–731.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action* (pp. xxiii, 376). The MIT Press.

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

Turvey, M. t. (1992). Affordances and Prospective Control: An Outline of the Ontology. *Ecological Psychology*, *4*(3), 173–187.

Van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *The Journal of Philosophy*, *92*(7), 345–381. JSTOR.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience* (pp. xx, 308). The MIT Press.

Von Foerster, H., & Zopf Jr. (1962). *Principles of Self-Organization: Transactions of the University of Illinois Symposium*. Pergamon Press.

Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, *113*(2), 358–389.

Wheeler, M. (2005). *Reconstructing the Cognitive World: The Next Step*. MIT Press.

Wilson, J. (2015). Metaphysical emergence: Weak and Strong. In T. Bigaj & C. Wuthrich (Eds.), *Metaphysics in Contemporary Physics* (pp. 251–306). Poznan Studies in the Philosophy of the Sciences and the Humanities.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

Wilson, R. A. (2004). What Computations (Still, Still) Can't Do: Jerry Fodor on Computation and Modularity. *Canadian Journal of Philosophy*, *34*(sup1), 407–425.

--- (2014). Ten questions concerning extended cognition. *Philosophical Psychology*, *27*(1), 19–33.

Winning, J., & Bechtel, W. (2019). Being Emergence vs. Pattern Emergence: Complexity, Control, and Goal-Directedness in Biological Systems. In S. Gibb, R. Hendry, & T. Lancaster (Eds.), *The Routledge Handbook of Philosophy of Emergence* (pp. 134–144).

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation* (Issue 1, pp. 233–249). Oxford University Press.

--- (2017). Physical modality, laws, and counterfactuals. *Synthese*, 1–23.