January 2013

Embodying Social Practice: Dynamically Co-Constituting Social Agency

Brian W. Dunst
University of South Florida, bdunst@mail.usf.edu

Embodying Social Practice:

Dynamically Co-Constituting Social Agency

by

Brian W. Dunst

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Philosophy
College of Arts and Sciences
University of South Florida

Major Professor: Alexander Levine, Ph.D.
Co-Major Professor: Stephen P. Turner, Ph.D.
Charles Guignon, Ph.D.
Joanne Waugh, Ph.D.
Mark Bickhard, Ph.D.

Date of Approval:
April 4, 2013

Keywords:
Dynamic Systems Theory, Social Practices, Embodiment,
Agency, Extended Mind, Phenomenology, Enactivism

Dedication

I dedicate this dissertation to my wife, Amy, who has tolerated my antics and supported me both financially and emotionally; and has kept life interesting and exciting. I could not have accomplished this without you.

Acknowledgments

This project would have been doomed from its inception if it hadn't been for the support and encouragement of Alex Levine, I cannot discursively express the depths of my appreciation and gratitude. I am also especially thankful to Stephen Turner and Charles Guignon who have both explicitly and implicitly shaped the way I approach thinking philosophically, and have encouraged me throughout this process. Joanne Waugh has always been an indispensible advocate for me in any of my projects while Mark Bickhard has been a model of professionalism, and helped me out in a time of need.

I would also like to recognize all of my professors from the USF philosophy department, who have helped hone my philosophical chops by pushing and testing me in all the right ways.

Of particular importance have also been my friends and colleagues; especially everyone who has bothered to humor and tolerate my arguing, skepticism, and stubbornness. Some of you have disproportionately shouldered the brunt of this: Nate Draluck, Megan Altman, Casey Rentmeester, Jared Kinggard, Steve Burgess, Philip Scuderi, Anthony Fernandez, Jeff Hinzmann, Pete Olen, David Eck, Hannah Behrmann, Erin Seeba, Gordon Bearn, and embarrassingly many others.

Finally, I must thank my family. Your love and support has allowed me to come this far. This goes especially for my wife Amy, and my parents Bob & Chris Dunst. Last but not least, I must thank Big Mama and Batman who have done their best to distract me in all the cutest ways.

Table of Contents

List of Figures

Abstract

In this dissertation I examine the relationships between metaphysical and epistemological notions of agency and social practices. I argue that the methodological framework of Dynamic Systems Theory (DST) is a fruitful way to discuss or explain these relationships, and through this framework I develop an account of Dynamic Embodied Agents (DEAs) in which agents can be understood both as dynamic systems interacting with their environments as well as social persons engaging in social practices and institutions. Through the conceptual lens of DST, DEA bridges the explanatory gap between individual agency and social practices left by the received views.

In the philosophy of mind, the received view of agency stems largely from treating cognition "methodologically solipsistically" which historically inherits its legacy from largely Cartesian attitudes isolating the inner life of the mind from the outer existence of the external world, and of other minds. By treating the mind as something that can be examined separate from the "external" environment, I argue that the received view fundamentally misconstrues the character of mental life. Instead, agents are constitutively and inextricably tied-up in their worldly activities. It makes no sense to try to understand cognition separate from the active, embodied, engagement of agents with and in their worlds. Drawing from more recent approaches to cognition including embodied, extended, embedded, enactive, and ecological accounts that actively attempt to escape traditional Cartesianism; I develop an account that understands agents as dynamic

systems sensitive and respondent to various aspects of their concrete situations. Such aspects can include physical, social, cultural, perceptual, and conceptual constraints and affordances; which together construct a situation's meaningfulness for the agent.

However, these approaches to cognition – embodiment, extended and embedded cognition, enactivism and ecological cognition – often fail to emphasize all of the relevant ways that mental life is shaped or constituted. Ecological and enactive approaches to cognition often miss the social and cultural dimensions of phenomenal experience (what Shaun Gallagher has, perhaps maladroitly called "philosophical autism"). The Embedded and Extended approaches to cognition often fail to adequately transcend the Cartesianism that motivated them in the first place (e.g. they fell subject to criticisms of "cognitive bloat" and charges of not satisfying the "mark of the cognitive"). And some accounts of embodiment play-up the importance of one or some aspects of mental life at the detriment of others (e.g. emphasizing the social nature of embodied experience while downplaying physical or conceptual constraints). I develop Dynamic Embodied Agency as a way of understanding agents as differentially sensitive to *all* of these relevant aspects of mental life.

In the field of social theory there are two primary approaches to understanding social activity. Either social activity is accomplished by independent autonomous individuals and is reducible to the activities of each of the individuals involved (often assumed by economic theorists), or social activity operates at an ontological level irreducible to the independent and autonomous activities of individuals. Many social theorists acknowledge the first view as an oversimplification, but see the second view as

positing a ghostly collective entity for which there isn't enough empirical evidence.

Some theorists however – communitarians – see the order or ontological priority as

reversed. Social existence does not emerge out of the complex interworkings of

individual autonomous agents; rather understanding one's self as an individual

autonomous agent is a genealogical or historical product or construction of a more

primordial social existence.

In either case, there is a difficulty in relating the activities of individuals to social

activities; in relating personal significance to social significance. There seem to be two

distinct and incompatible levels of analysis or explanation: the individual and society.

Here what is needed is some theoretical or explanatory construct that can bridge the

explanatory gap between individual agency and social agency. The most obvious locus

to look for a connection between the two is in the notion of social practices. Social

practices are activities in which individuals participate such that they align or attune

themselves to the way(s) that "one" – a member of the social community – does things.

So, for example, one uses a bicycle for transportation, or exercise, or recreation, or

relaxation. One rides a bicycle in a certain manner (e.g. one conforms one's body to sit

on the seat, to move one's legs in a certain fashion to pedal, to hold the handle bars in a

certain manner to help balance and steer the bicycle, and so on). Riding a bicycle is then

a straightforward example of engaging in a fairly well-recognized and structured social

practice. One could attempt to use a bicycle in relatively unrecognized or less-structured

ways (like using a bicycle as a musical instrument, or as a hammer), and this would make

such bicycle-related activities less identifiable as *social practices* but rather as *deviant*

behavior – but behavior that nonetheless has social ramifications. *Riding* a bicycle, at any rate, is a social practice.

So the question, or the issue, of social practices is in how to understand them in terms of individual and social activities. Are social practices reducible to the activities of (relatively) autonomous individuals; or are they social forces that impose themselves on individuals? The DEA approach that I develop rejects this dichotomy. Dynamic Embodied Agents are not independent, autonomous individuals – they are open, far-from-equilibrium systems; which means they are highly dependent on dynamically interacting with their environment. The environment, for a DEA, is anything to which the system is differentially sensitive or responsive, and can include physical, social, cultural, perceptual, and conceptual influences. To oversimplify matters, we can think of physical influences as the material stuff one finds in one's environment—one's physical environment. Perceptual influences include one's bodily shape and sensitivities—the ways in which one's body can engage with the surrounding world. Social, cultural and conceptual influences shape how an agent "understands" her situation; but this "understanding" is not necessarily a conceptual or reflective understanding (though it can be); it could be more of a pre-reflective sense of how the world directly "shows up" to the agent. For example, a situation – perhaps a Catholic's first communion – could be fundamentally or pre-reflectively imbued with the phenomenal feeling of being engaged in a sacred ritual, and this feeling "colors" the meaningfulness and active possibilities of the situation for the agent.

So far, the sense of Dynamic Embodied Agency that I've given seems to privilege

the pre-theoretical sense of what an agent is – an agent is a *person*, like you or me; and

we tend to pre-theoretically think of persons as autonomous individuals. I'm a person,

you're a person, perhaps even my cat is a person, maybe Cylons are persons too; but *we*

aren't *a* person. A nation isn't a person, it is comprised of persons. A corporation isn't a

person, it is comprised of persons; and perhaps that is one reason why the Citizens United

decision that affirms legal "corporate personhood" gives some pause. This pre-

theoretical intuition about what counts and doesn't count as a person is in many ways

based on an historically established Cartesian folk-psychology that carves the world up

such that everything is either a mental (*res cogitans*) or a physical (*res extensa*)

substance.

However, DEA is neither methodologically nor metaphysically wed to this dualist

picture. Such things as national or corporate persons are, at least in principle, *possible*

from the perspective of Dynamic Embodied Agency. So long as the constraints of

personhood (or criteria for counting as a person) are satisfied (and as long as such

constraints don't circularly presuppose that persons must fit our pre-theoretical

intuitions), then any dynamic embodied system can be a person; and anything that

constitutively factors into the composition of that system also gets included in the

system's personhood. This, for example, accommodates some of the more radical results

of the extended mind thesis; which argues for the possible inclusion of entities "beyond

the bounds of skin and skull" into an agent's cognitive architecture. The Extended Mind

thesis argues, for example, that an Alzheimer's patient's notebook – used in lieu of a

brain-bound biological memory – constitutes the memory of that agent, and thus counts

as part of the agent's cognitive apparatus. The Dynamic Embodied Agency thesis argues

further that if the notebook constitutively factors into the agent's personhood (e.g.

without it, she wouldn't count as a person) then it is part of the agential system – it is

included in who she is as a person.

Dynamic Embodied Agency, is, importantly, *dynamic* – so perhaps sometimes the

notebook is constitutive the Alzheimer's patient's personhood, and sometimes it isn't,

depending crucially on the specific conditions of particular situations. DST provides a

wealth of conceptual tools through which one can fruitfully analyze the qualitative ways

in which systems change as a result of changes to the functions, parameters, and variables

that define (or constitute) a system. It can help explain how seemingly slight changes can

result in dramatically new system behavior – which helps explain the possibility of

emergent phenomena. With the possibility of emergence comes the possibility of

qualities or properties genuinely attributable at "higher" levels of analysis, including

social levels. This allows for the possibility that social practices may not reduce to the

activities of individual autonomous agents; but are instead attributable to the particular

complex relations between individuals and other contributing situational factors.

However, DST is merely a *methodological* framework, not a metaphysical doctrine. It

doesn't make claims about or justify the existence of "higher-order" *entities*. It doesn't

claim that social-level entities exist; rather it furnishes explanations for, and between

social and other levels of phenomena. This explanation may allude to emergent

properties or phenomena without reifying them metaphysically. The ontology utilized in

DST then is not a *metaphysical* ontology; but rather an *epistemological or*

phenomenological ontology.

Further, while DEAs co-constitute and integrate with- and within their dynamic (social, cultural, physical and interpersonal) environments in high-bandwidth, differentially sensitive, non-linear, and open-textured relationships; Agents cannot *share* or *transmit* meaningful information between each other. However, DEAs are not *isolated* from each other—they *interact* meaningfully with those around them. While information is not shared or transmitted, DEAs can and do engage each other in meaningful, dynamic, high-bandwidth, differentially sensitive, non-linear, open-textured interplay. Just as DEAs constitutively couple with their physical, social, and cultural environments, so to can they constitutively (and meaningfully) engage each other "interpersonally" – this is precisely in what *interpersonal communication* consists; on the account of DEA proffered.

Dynamic Embodied Agency is then an account that draws on the resources of Dynamic Systems Theory to explain how agents are social actors capable of differentially engaging together, socially, in practices and institutions.

Chapter 1: Cognition, Perception, and Agency

> *"One's own body [le corps propre] is in the world just as the heart is in the organism: it continuously breathes life into the visible spectacle, animates it and nourishes it from within, and forms a system with it."[1]*
>
> — Maurice Merleau-Ponty

1.0   Introduction

Since the Modern era, theories of cognition have distinguished between active and passive modes of experiencing. Locke's empiricist theory was by-and-large passive, requiring a good deal of training up from a *tabla rasa*, or blank slate, before cognitive processes could capable of critical thinking or to synthesize sense experiences (perceptions) into categories of understanding[2]. To Locke, since all knowledge is *a posteriori,* there is nothing innate or *a priori* about the categories humans tend to use, rather they are conventions and artifacts of the sense modalities and the socially inherited information available. Locke distinguished between the passive 'simple' ideas presented through sense perception (e.g. colors, sounds, tastes, smells) and more 'complex' ideas (e.g. numbers, causal inferences, abstract ideas, relations, etc.) that are built up through the active work of cognition[3]. Locke's 'active' cognition was essentially a highly complex associative machine capable of taking simple ideas and combining them[4].

---

[1]   Merleau-Ponty, M., & Landes, D. A. (2012). *Phenomenology of perception*. Abingdon, Oxon: Routledge. p. 209 [245].

[2]   Cf. e.g. *An Essay Concerning Human Understanding* Bk. II Ch's 1-3, 9.

[3]   Ibid. Bk. I, Ch's 1-4.

[4]   Ibid. Bk IV, Ch's 1-5

Later, Kant introduced a theory of cognition that relied on the spontaneous organizational or patterning work performed on incoming sense impressions or raw perceptions, "schematizing" them into universal categories of understanding[5]. This schematization enlisted the help of the imagination – a cognitive faculty separate from the understanding[6]. Many "post-Kantian" cognitive theories stem from one or both of these models of cognition[7], and paint the cognitive agent as a passive recipient of sense impressions. That is, these cognitive theories focus on cognitive work as it is performed upon internalized sense stimulus rather than the interactions with their local environment in which, and through which, cognitive agents actively think. Kant's introduction of spontaneity was radical in that it made space for an agent; but this spontaneity only had a chance to function within a limited scope, and only after sense impressions had been internalized from the external world. In a sense, what lacked in Kant's and post-Kantian theories of cognition was *agency*: information flowed unidirectionally from the sensory world to the understanding; never in reverse.

Such theories of cognition privileged, mostly tacitly, the inside/outside division implicit in Cartesian Dualism[8]. This distinction between the inner realm of beliefs, desires, thoughts, and other experiences, and the outer realm of raw physical goings-on available to sensation—the perceptual information that human sensory systems "take in"—is still largely assumed in contemporary theories. For example, Jerry Fodor's

---

[5] Cf. e.g. *Critique of Pure Reason* A139-A142
[6] Cf. e.g. *Critique of the Power of Judgment*, 20:211-20:216.
[7] To be sure, Descartes and Hume also play a large role here. To a certain extent, however, both Descartes' Rationalism and Hume's Empiricism are subsumed in Kant's and Locke's cognitive theories as critical responses to their predecessors.
[8] Cf. Descartes' *Meditations* I & II.

"methodological solipsism" depends explicitly on the inside/outside division, and proceeds by automatically (or "methodologically") assuming the fundamental soundness of the boundary conditions supporting this division.[9] To Fodor, if we want to make any progress on understanding how the mind works, we must treat the mind as a more or less isolated and self-contained system.

Following Kant, Fodor and others treat cognitive processes as essentially representational. If we—following Kant—understand the bombardment of sense organs with perceptual information from the external world as the sense impressions' being "presented" to the agent as "intuitions", then such *mere* presentation is, as Kant shows, not enough for understanding. To achieve understanding, the cognitive agent must synthesize these data into the natural categories to which they belong; and this requires spontaneous and active work. The intuitions upon which this spontaneous cognitive work has been properly performed are now "*represented*"; they are presented to the isolated cognitive agent in a way that is suitable for understanding. Fodor understands such representation as a structure articulated in a sort of "Language of Thought", a language in which the cognitive subject (in this case, a functioning brain) can shuttle information to various compartments or modules of the brain specially trained up or adapted for certain cognitive tasks[10].

Accordingly, sense-based experiences are the result of a series of successive

---

[9]   Fodor, Jerry (1980), "Methodological Solipsism Considered as a Research Strategy in Cognitive Science," *Behavioral and Brain Sciences*, 3: 63-73.

[10] Here I have combined two theses central to Fodor's theory of cognition—the modularity hypothesis and the language of thought (LOT) hypothesis. These two theses are necessarily related, as Fodor explains in *Modularity of Mind: An Essay on Faculty Psychology* (1983). His theorizing about LOT first appears in his 1975 The Language of Thought wherein he expresses the hypothesis that, if cognition is essentially representational, then cognitive relations must be essentially rule-governed in precisely the same ways that formal languages are syntactically rule-governed.

layers of brain operations that eventually produce familiar subjective phenomenal experiences. It seems then that there are still just two viable pictures of cognition for the representationalist: first, of a cognitive subject well-insulated by layers of pre-phenomenal cognitive operations but still maintaining an inside/outside boundedness. Second, if the first option doesn't suit, the representational picture can afford a non-dualistic, (often cashed out as materialistic) image. In this physicalist sense, cognition is "matter in motion" all-the-way-down, reducing mental processes to brain processes such as the physiochemical workings of firing neurons. The first in some respects recalls a Cartesian substance dualism, the second—often called "Type-Identity Theory"[11]—offers a monistic reduction of the mental to the physical.

For most of the history of philosophy of mind and cognitive science, these two were the only two proposals widely taken seriously—and of them, dualism was generally thought to be extremely problematic at best, and completely untenable at worst. Because of such basic difficulties, somewhat more recently cognitive models that focus on active agent-world interaction have been introduced. These models situate agents and their concrete local environment as active co-participants in the constitution of an agent's cognitive world[12]. For our purposes, it suffices to begin with the efforts of Francisco Varela, Evan Thomson, and Eleanor Rosch, who developed a novel, embodied approach to cognition in the early 1990s. This new embodied approach to cognition acknowledged that sense-experience (or perception) is shaped not only passively by whatever

---

[11] Cf. Place, U.T. (1956). "Is Consciousness a Brain Process?", *British Journal of Psychology*, 47, pp. 44–50

[12] Depending on which specifics one is interested in, one might begin in the 1910s with Husserl, the 20s-30s with Heidegger, or for a more explicitly embodied approach, the 40s and 50s with Merleau-Ponty. More on this later.

phenomena impresses itself upon a cognitive subject, but also actively in the sense that

agents (rather than mere *subjects*) can and do dynamically engage with and effect change

upon the world around them. On this view it was no longer tenable to maintain a strict

boundary between inside and outside; between presentation and representation; or

between sense perception and phenomenal experience. The processes of cognition were

re-envisioned as a complex interaction of body interfacing with its local environment. In

this way, embodied cognition clearly opposes both Cartesian dualism—in particular the

strict division of separate substances (*res cogitans*, and *res extensa*), and purely

reductionist and representational accounts of cognition—in particular the idea that *only*

internalized representational content is cognitively meaningful.

Beginning with Brentano's[13] and Husserl's accounts of intentionality[14], and later

developed implicitly by Heidegger[15] and more explicitly by Merleau-Ponty[16], a

movement within the phenomenological tradition has emphasized the relevance of the

body in cognitive subjectivity. Rather than separating inner and outer realms, the

phenomenological tradition approaches the appearance of separate substances from an

interactive perspective. Mind and Body are not two separate (or separate kinds of)

entities; instead, the workings of one fundamentally implicate and depend upon the

workings of the other. To think otherwise is to fail to adequately account for the

fundamental evidence of actual lived phenomena[17]. So on the embodied approach to

---

13 Cf. Franz Brentano, *Psychology from an Empirical Standpoint*, Bk. II
14 Cf. Edmund Husserl, *Ideas: General Introduction to Pure Phenomenology*.
15 Cf. e.g. Heidegger's discussion of readiness-to-hand in *Being and Time* (§§67-72).
16 The entirety of Merleau-Ponty's *Phenomenology of Perception* marks the first sustained
    phenomenological meditation on embodied subjectivity.
17 The phenomenological intuition here is that the world dynamically presents itself as simultaneously the

cognition an agent navigates her world utilizing all available perceptual tools—regardless of 'where' they are located or of what they are made. Some of these tools include various types of intentions (conceptual, motor, emotive, etc.), others include the transparent and skillful know-how incorporated in the deployment of the agent's body, while still others differentially adapt to and utilize the resources, constraints, and other perceptual aspects afforded by the perceptual agent's dynamic environment.

No longer constrained by the strictures of methodological solipsism, phenomenological embodiment allows cognitive theorists to model the mind as a complex inter-action between the functional workings of the brain, the transparent skillful know-how of the body, and various physical availabilities or "affordances"[18] of the local environment. In this way, embodiment theorists took themselves to be giving a more authentic account of what is phenomenally given—the actual kinds of perceptual encounters that human agents experience.

1.1   *Embodied Agency and Free Will*

Primitive to this way of understanding cognition is the notion of agency. With the assumption of agency comes a new collection of problems. When we think of ourselves as agents we understand ourselves as in some ways *free* and in other ways constrained. No agent can be completely constrained, and to the extent that the physical world obeys deterministic laws, there must be some accounting for the degrees and dimensions of

---

kind of thing that can affect and be affected by the perceptive subject. This is the functional difference between a mere perceiving cognitive subject, and an embodied agent. What is important to the phenomenological approach is that we in fact experience bidirectionality in causal efficaciousness.

[18]   This is a term of art introduced by J.J. Gibson. His "Ecological" theory will be discussed in further detail shortly.

freedom required by an embodied account of cognitive agency. That is, how does an embodied account of cognition pare with a thoroughgoing naturalism? How is it possible to account both for phenomenal experience and physical laws in a consistent manner? Surely a theory of embodied cognition—one that privileges *the body*—is on better physical footing than a dualistic conception of the mind; but an embodied agent's body is not the same kind of thing as either an unused, inanimate object or a tool. An embodied agent's body is a *lived* body—the kind of thing that reliably, continually, consistently, and transparently factors into the agent's perceptual cognition—that is, the embodied agent's lived body *constitutes* that agent.[19]

Further, free agents are free in some ways, and not in others. For example, I am free to move my left arm in certain ways because my current environment constrains it in such a way that allows it to be moved in a specific fashion. The amount of energy afforded by complex metabolic processes similarly constrains and affords certain types of movement, and can include, under such circumstances a class of possible movements. And, just as the amount of energy in an open system is a differential function of the activities and movements within that system, the local system containing my body can differentially acquire and lose various amounts of energy depending on the kinds of activities in which I engage[20].

---

[19] The distinction here is akin to Husserl's well-known discussion of the differences between *Leib* and *Körper* (cf. e.g. Husserl's Fifth Cartesian Mediation, §§42-62) . *Leib* is the body as an active locus of lived experiencing, while *Körper* designates an objective physical object. Importantly, nobody holds the position that these two notions are mutually exclusive. A body can factor into situations significantly and simultaneously as both *Leib* and *Körper*. What is important here is that the body *qua Leib* is a necessary condition for agency.

[20] For an elaboration on the idea of metabolic constraints on action and the kinds of activities available to (or afforded by) the interfaces between embodied agents and their interactive environments, see Norman, Donald A. (1999). "Affordances, Conventions and Design." *Interactions* 6(3):38-43, May

But what about social factors? Consider the situation in which nobody is telling me not to move my left arm in the way I would like to, and in which there are no social strictures or taboos preventing my doing so—so in this way I am again in certain ways "free" to perform various possible actions, including moving my arm like so. Thus, the constraints on and motions afforded to my arm in this situation are importantly influenced by how I see the action in *social* terms. Additionally, when I do move my arm, its reach can only extended so far. If I wanted to reach further I could expand my reach by taking up some tool to extend beyond my arm span. Or, I could stand up and move my entire body closer to what I'm attempting to reach. My body is, as is the grasping tool, a useful *thing*—but it is not *merely* a thing—it is that through which experiencing occurs; it is itself the condition for any usefulness. If there are no tools available in my local environment, or if I am unable to successfully incorporate a tool, these shortcomings constitute real forms of constraint in my concrete situation. So too for social factors. If I am unable to conceive of certain forms of action, or I perceive there to be social norms prohibiting certain modes of activity—this affects the kind, and quality of action I can undertake. Further, if social taboos are well-established, certain actions may not show up at all as available or possible ways to act. Thus there are myriad ways of constraining and affording embodied agency, and in many cases what counts as a constraint or affordance depends on the particular intentions of the particular embodied agent in her particular social, cultural, historical, and physical situation. Put simply: affordances and constraints are functions of an agent's embodied 'grasping' of her situation[21].

1999, ACM Press

[21] Here I shrink from using the word "understanding" rather than "grasping". The word "understanding"

8

In a more fundamental way however, freedom and in particular the idea of free will entail a certain subjectively felt sense of agency—that *I* am the one in control of my actions. An adequate cognitive theory must give some accounting for this qualitative, phenomenal sense of freedom. Embodied agents are intentional, and the basic issue with the project of naturalizing free will is of accounting for the basis of intentionality. Where does intention come from? Are we strictly physically determined? Socially determined? If not, why not? How not? In some sense the embodied account attempts to skirt this issue by taking agency as both metaphysically and epistemologically primitive. According to many construals of embodied cognition and other phenomenological theories, agency is—like electromagnetism and gravity—a thoroughly natural phenomenon[22]. It is something that is "in", and "of" the natural world. It is, like other natural phenomena, a dependable and law-like way for things to go. The fully naturalistic view of embodied agency sees agency as an emergent property of the parts of certain kinds of systems (perceptual systems) in the same way that a hurricane is an emergent property of more micro-scale physical processes.[23]

---

carries with it considerable conceptual baggage. If "understanding" is understood cognitively in the way that Kant does, then one can interpret what I mean here subjunctively—it is what the body would do if it were the kind of thing that could understand. Of course, much of the argument for embodied cognition goes against this way of understanding "understanding". Instead, we should think of "understanding" as the way in which an agent is able to make meaningful her situation. In this sense when we think of agents as essentially embodied it should be clear that bodies are precisely what can "understand" in a non-subjunctive sense. To avoid these pitfalls, I opt for "grasping" here as a term of art. It is not meant to evoke Frege's usage of the term, but rather a new usage—whose embodied entailments (actively reaching out to grab a hold of) are fully intended.

[22] Cf. e.g.: Franck Grammont and Dorotheé Legrand (2005, 2010), Pierre Livet (2002, 2006), Jean-Luc Petit (1999a, 1999b).

[23] For example, something like this is suggested by David Chalmers' "Type-F" neutral monism (cf. eg. "Consciousness and its Place in Nature" in Stitch & Warfield (ed.'s) Blackwell Guide to the Philosophy of Mind. 2002). Chalmers suggests that there is perhaps some more primoridal substance (which he calls "protophenomenal") from which all of consciousness, qualia, physical substance, and physical, psychological, and psycho-physical laws emerge.

1.2   Embodiment, Dualism, and Reductionism

Much of the motivation for adopting an embodied account of agency is to avoid

the problems associated with dualism and reductionism. Concerning the issue of agency,

there are two basic problems with dualism: First is the problem of separate substances.

Dualism attempts to straddle both sides of the issue when it comes to our intuitions about

on the one hand our inner, personally-felt experiences—our phenomenal sense of being

intentional agents with free will; and on the other, intuitions about the mechanistic

lawfulness of physicalist explanations of the world around us that overwhelmingly seem

to accurately describe and explain the way things work. For the dualist, there are simply

two separate kinds of substance: the inner, experiential substance of cognition (*res

cogitans*) and the outer, physical substance of the world (*res extensa*) that obeys physical

laws. The problem of separate substances is then the problem of how to make two

entirely unrelated kinds of substance interact—and to do so in a way consistent with the

known modes in which agents *qua res cogitantia* affect their environs, as well as the

ways in which environs *qua res extensa* affect agents. No dualist has ever adequately

solved this problem.[24]

The second problem is the classic "problem of other minds". Even if dualism

succeeds at coherently explaining the seemingly conflicting intuitions of inner

---

[24] That is, the Dualist faces the double-problem of explaining the causal efficaciousness of the physical on
the mental, and (more problematically) vice versa. Without explaining this, mental causation of
agentive action is a miracle.

phenomenalism and outer physicalism[25], it still faces the difficulty of determining what

kind of entities other people are[26]. We have a social sense of others. This means that

when one (who takes herself to be a *res cogitans*) encounters another, she encounters the

other *as another* person or *res cogitans*. This raises a couple of related issues. Is the other

res cogitans the same *kind* of substance as I? Or is he an entirely different kind of

substance (perhaps not *res cogitans* but a third *res altera*[27]?). How can I know whether

the other is the same kind of substance as I? There is a worry here that a condition for the

possibility of knowing whether two substances are of the same kind, the two (I and the

other) must at least be able to come into substantial contact—but if this can occur, the

new problem of being able to determine a basis for individuation between me and the

other arises. We run the risk of substantial indiscernibility, and therefore of the possibility

that we are [*sic*] identical[28]. Thus, the only way one would be able to know that I and the

other are of a kind is to *infer* it from experience[29]. Otherwise we are left only with the

possibility that the other is merely *res extensa*—the other is not a *subject* but a mere

physical *object*; and this collapses the dualist's position into one of cognitive solipsism.

---

[25] It doesn't.

[26] This is a more metaphysical formulation of the problem of other minds, which is usually understood as an epistemological problem – the question as to how one can know the existence and content of other minds. To be sure, one cannot separate the metaphysical from the epistemological concerns in addressing the problem of other minds because any explanation that addresses epistemological concerns necessarily commits one to metaphysical assumptions, and vice versa.

[27] "Other thing". But here too once faces myriad difficulties: In what way is this *res altera* an "other". To be an other is to be, in some way, similar to one's self; and if the other must be similar, then it cannot be understood as being an entirely separate or distinct substance at all. Further, if one other is a distinct separate substance, then how about a second other? Are two others alike in kind (e.g. they are both res altera) or are they distinct in kind (e.g. that each case of *res alterae* is a separate substance from the other)? The effect of these issues is of expanding dualism into a triism or pluralism which only compounds the problems outlined above. Kant faced a similar difficulty in attempting to characterize the role of the imagination in schematization of sense impressions data for organization by the understanding. The third critique is largely Kant's failed efforts to reconcile this problem.

[28] An absurdity.

[29] To add to the difficulties, at best this inductive inference is made from a single, decidedly biased case.

Thus, the problem of other minds is the problem of reconciling our sense that others are *subjects*, both "like me" and "other". Like the problem of separate substances, the problem of other minds has never adequately been solved by the dualist. Unlike the problem of separate substances, the problem of other minds poses a formidable difficulty for monist theories as well—especially in attempting to develop adequate theories of interpersonal communication.

Regarding reductionism, we once again face an impasse of intuitions. In our regular, unreflective, everyday use of language we tend to draw upon both physicalistic and mentalistic concepts—that is, as social linguistic agents who have inherited a dualistc folk-psychological tradition, our encultured attitudes assume a form of dualism.[30] Reductionism is the attempt to subsume all of either physicalistic or mentalistic language under the other. As a matter of historical fact most reductionists tend to be materialists, intending to subsume mentalistic language under physicalist descriptions. As such, for example, when one says "I desire Malbec and believe the distributor has some in inventory", the reductionist suggests that we should understand terms like "desire" and "believe" rather in terms of the underlying physical states or processes that accompany them. The mental is said to "supervene" on the physical. To compare: a physicalist *eliminativist* (contra the physicalist reductionist) holds that mentalistic terms like "desire" and "believe" are strictly nonsense—they do not express anything physically meaningful (and thus do not express anything about *reality*)[31,32]. They should not be subsumed under

---

[30] Edmund Husserl addresses precisely this in his discussion on the "natural standpoint". Cf. Husserl, Edmund; trans. W. R. Boyce Gibson. (1962). *Ideas: General Introduction to Pure Phenomenology*. New York: Macmillan. 96–103, 155–67.
[31] It is not that eliminativists believe that all mentalistic language has no relation to physical reality—

a better description, but should rather be supplanted whole-cloth. I will not here discuss

the eliminativist position because it refuses to take seriously precisely what is in need of

explanation[33]. The problem the reductionist faces is predominantly in providing a

functional mapping from the linguistic and conceptual domain that utilizes mentalistic

language and concepts to the physicalist domain. Try as she might, the reductionist's task

resembles that of playing Whack-a-mole: every time she appears to make an inroad in

bridging the gap between mentalistic and physicalistic language, the process of reduction

itself opens a new rift in need of bridging[34]. Further, some mentalistic terms seem to be

irreducible to the current stock of physical concepts[35]. One well-identified example is

that of the "hard problem" of consciousness. "Consciousness", it appears, is irreducible to

purely physicalist description. The "hard problem" of consciousness is in giving a

naturalized account of consciousness that adequately retains the richness of the concept.

If consciousness cannot be explained by using an exclusively physicalistic description[36],

it must then "emerge" from or "transcend" the purely physical—which then requires

some irreducible, nonphysical explanation. However, the possibility of a non-

---

rather that *qua* mental, mentalistic language can do nothing but distort or mislead about reality (which is monistically physical). Thus, for eliminativists, all mentalistic terms unnecessarily distort reality—even if they can be cashed out in terms of physical goings-on (and thus eliminated). According to the elimitivist, the only way to know which mentalistic terms can (or cannot) be cashed-out in physicalistic language is through scientific, empirical investigation.

[32] For a sustained argument on behalf of eliminative materialism see Paul *Churchland's Scientific Realism and the Plasticity of Mind* (1979). For a critique, Lynne Rudder Baker's *Saving Belief: A Critique of Physicalism* (1987)

[33] That is, it explicitly refuses to play by the rules of persuasive dialogue. The eliminativist claims that he does not need to acknowledge the plausibility of the opposing viewpoint. Essentially, the eliminativist suggests that the only way to explain the seeming existence of mental phenomena is to explain them away. This is precisely a refusal to acknowledge that there is something in need of explanation - and thus constitutes a mere refusal to participate in the dialogue.

[34] Cf. e.g. Hellman, Geoffrey, and Frank Wilson Thompson (1975), "Physicalism: Ontology, Determination, and Reduction", *Journal of Philosophy* 72: 551–564.

[35] Cf. e.g. Fodor, Jerry (1974). "Special Sciences," *Synthese*, 28: 97–115.

[36] That is, using only the stock of concepts and [causally-closed] laws that current physics admits.

physicalistic explanation is precisely what the reductionist methodologically prohibits. It should come as no surprise that despite concerted effort over the past sixty years or so reductionism has not yet succeeded, and that the "hard problem" has not yet been solved.[37]

If the embodied account can avoid these problems, it begins to look comparatively attractive, given the seeming intractability of the problems inherent in dualistic and reductionistic accounts of cognition. Both dualism and reductionism share the same Cartesian history, and as such have inherited the same paradigm for thinking about cognition—namely that *propositional attitudes* are the meaning-bearing objects that signify the [purported] differences between the mental and the physical[38]. This tradition sees cognition as a two-way flow of discrete packets of information (representations). Information flows from outside in through a process of representing sense-impressions. Information flows from the inside out through an intentional process of agential action. The former has its source in the extended world, the source for the latter is in the agent's internal phenomenal experience and will. The traditional approach has been to reveal the mechanisms by which these two, unidirectional, linear streams of information come

---

[37] It should be noted that while the "hard problem" of consciousness has provided difficulty for reductionism since the rise of Type-Identity theories of the mind (cf. e.g. Gilbert *Ryle's The Concept of Mind*, Hutchinson, 1949; U.T. Place's "Is Consciousness a Brain Process?", *British Journal of Psychology*, 47:44-50. 1956; and J.J.C. Smart's "Sensations and Brain Processes". *Philosophical Review*, 68:141-56. 1959) the term "the hard problem of consciousness" itself appears only with David Chalmers' "Facing up to the Problem of Consciousness" in *The Journal of Consciousness Studies* (2:3, 200-219) in 1995.

[38] Bertrand Russell first described propositional attitudes in his 1912 *The Problems of Philosophy*. Nonetheless, it is fairly clear that propositional attitudes are what Descartes had in mind in his formulation of res cogitans in the second meditation when he says "But what then am I? A thinking thing. And what is that? Something that doubts, understands, affirms, denies, wills, refuses, and also senses and has mental images." Later W.V.O. Quine took issue with the elusiveness in providing adequate translation of propositional attitudes in first-order predicate logic. Alternatively, Daniel Dennett uses propositional attitudes as the basis for his concept of "the intentional stance".

together to constitute one cognitive agent[39]. By contrast, the embodied account rejects this traditional picture of separate, linear channels of information flow and replaces it with a *nonlinear*, mutually causal and reciprocally constitutive interactive and dynamic system of body and local environment[40]. The "nonlinearity" of the embodied account is a significant divergence from the linear Cartesian tradition in that it puts the agent and world into direct, "high-bandwidth," [41] and continuous contact with one another—rather than requiring a series of translations, distillations, conversions, and other mechanisms that function as simultaneously separating and connecting agent and world.

This fundamental change in approach both complicates and simplifies the argumentative burden for the embodiment theorist. It complicates things in the sense that moving from a linear theory to a nonlinear theory usually increases the degree of complexity of the system the theory is capable of modeling. In a sense, this is a trade-off between simplicity and accuracy in theory description. The anomalies faced by the linear approximations of information flow in dualism and reductionism became too much to bear for some who identified the source of the problem as arising from an unwarranted, linearizing, simplification. Such simplification was seen as a distortion of what *really* goes on in cognition: thinking and perceiving are simply not "low-bandwidth," serial, exchanges of discrete packets of information. Alternatively, things are simplified for the embodiment theorist by shifting to a nonlinear approach because the old, seemingly

---

[39]   For the dualist this means explaining how mind and body can interact; for the reductionist this means explaining how the mental is expressible in purely physical terms. Both agree that there is a well-bounded 'something' pertaining to mentality that needs explaining.

[40]   To be clear, locality of environment is a matter of degree (of influence), and the notion of 'environment' I mean to use here is maximally inclusive—so contains things like social, cultural, historical, biographical, and physical influences. The idea of "degree of locality" is similar to the concept of "horizon" as it is deployed by many in the phenomenological tradition.

[41]   This phrase is taken from John Haugeland's *Mind Embodied and Embedded*, 1998.

intractable problems of separate substances, other minds, and the hard problem of

cognition are dissolved by this shift in paradigm. Thus, much of the motivation for

shifting from the traditional Cartesian approach to the embodied approach comes from

intuitions about the fruitfulness of further research, and misgivings about the traditional

approaches abilities to overcome these enduring faults.

1.3   *Embodied* or *Extended?*

In the mid 1990s, around the same time that the embodiment thesis began to take

hold in cognitive science, another non-traditional thesis about cognition began to surface.

This thesis—the extended mind thesis—grew out of two related research projects that

both enjoyed popularity in the 1970s and 1980s. The first was a thesis in philosophy of

language and mind called "semantic externalism," developed variously by Hilary

Putnam, Tyler Burge, and Saul Kripke (to name just a few). The second was a

constellation of approaches born out of attempts to conceive artificial intelligence (AI)

independently of the then dominant computationalist / representationalist picture of

cognition. These approaches went by an assortment of names, such as connectionism,

parallel distributed processing, subsumption architecture, and cybernetics[42]. For

simplicity I'll refer to these approaches more generally under the umbrella term

"dispersed cognition"[43].

---

[42]   Cf. e.g. Rumelhart, D.E., J.L. McClelland and the PDP Research Group (1986*). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volumes 1 & 2.* Cambridge, MA: MIT Press ; Churchland, P. M., 1989, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, Mass.: MIT Press. ; Pinker, S., and Mehler, J. (eds.), 1988, *Connections and Symbols*, Cambridge, Mass.: MIT Press.; Smolensky, P., 1988, "On the Proper Treatment of Connectionism," *Behavioral and Brain Sciences*, 11: 1–74.

[43]   I would like to be clear that for the purposes of this text the terms "dispersed cognition" and "distributed

The semantic externalism thesis is the view that linguistic terms obtain meaning by their succeeding at attaching to aspects of or objects in the world without relying on the intension of the speaker who invokes them[44,45]. Importantly this view is, as stated, recognizably Cartesian in the sense that it relies upon the coherence of an internal / external distinction. The embodiment thesis, to contrast, takes as one of its primary tenets the problematization of such a distinction. If this is an accurate characterization of semantic externalism, then we have already located one potential source of tension between these two positions (and by extension between the two positions of embodiment and extended cognition). The classic example used in expressing the meaning externalist position is the well-trodden "Twin Earth" thought experiment. Because this is such a well-known bit of contemporary philosophy I will here only briefly discuss its basic contours.

Suppose there are two distant worlds that are identical in every way except one:

---

agency" (or what sometimes gets called "distributed cognition") should not be thought of as synonymous. Dispersed cognition is a thesis about how cognitive processes work and are spread out at various levels of cognitive activity. Distributed agency is a thesis about actions as enacted or shared across the activities of many agents. In both, the idea of 'distribution' suggests that the term being qualified (cognition, agency) is 'spread out' over or amongst multiple objects (or, in the case of agency, subjects). In a [fully integrated, feed-forward] connectionist network, for example, processing is distributed over a network of nodes and weighted connections. One may endorse either, both, or neither thesis and remain internally consistent.

[44] The term 'intension' refers to the linguistic semantic element of an agent's intention. For an extended discussion on the differences and relations between "intention" and "intension" see John Searle's defining 1983 work *Intentionality*. Here (as with Putnam's original formulation) the term 'intension' is used much in the same way that Frege understood the term "sense" (Sinn)—the psychologistic feeling of what one means when one uses a word.

[45] In the philosophy of language, "intension" is often defined by substitution failure in opaque contexts—the inability to discern a clear reference merely from grammatical structure. Cf. Gamut, L.T.F. (1991). *Logic, language, and meaning*, University of Chicago Press, Chicago.
Using this definition of intension, the semantic externalism thesis can be expressed as either (i) the claim that the external, intersubjectively accessible situation always provides enough information to fix context such that they cannot be opaque; (ii) reference is determined externally by context, so to the extent that something refers, it does so non-opaquely—even if language-users are not privy to that reference. Cf, Kripke, Saul. (1980). Naming and Necessity. Cambridge: Harvard University Press.

where one world, 'Earth', has H_2O, the other world, which we may call 'Twin Earth' has instead XYZ. In both universes, all agents refer to their own planet as 'Earth' and as the most prevalent substance on the surface of their respective planets as "water". For the Earthling, says the meaning externalist, uttering 'water' picks out substances that are predominantly composed of H_2O—the substance on Earth. For the Twin-Earthling, then, uttering "water" picks out substances that are predominantly composed of XYZ—the substance on Twin Earth. Thus, Earthlings who say 'water' *mean* H_2O, and Twin-Earthlings who say 'water' *mean* XYZ. As would need to be the case in order to say that Earth and Twin-Earth are identical in every way except the piecewise replacement of each molecules of H2O for molecules of XYZ, the physical characteristics of H_2O and XYZ would also have to be identical—or else the two worlds could not propagate identically, and would immediately begin diverging[46]. According to meaning externalists, since it would be absurd to think that Twin-Earthlings could mean H_2O when uttering 'water', and since Twin-Earthlings are brain-state identical with Earthlings [and, presumably, mental states supervene on brain states so Twin-Earthlings and Earthlings would have identical intensions] the ability to refer cannot be intensional—leaving by exhaustion only the possibility of extensional reference. This is what purportedly motivates Putnam's famous claim that "meanings just ain't in the head"[47].

---

[46] The claim here is that differences the micro-level qualitative characteristics of H_2O and XYZ would correlate to differences in macro-level qualitative characteristics. As with most thought experiments, the "Twin Earth" thought experiment faces some challenging conceptual hurdles. Since what makes H_2O what it is are its qualitative physical characteristics, if there are no qualitative physical differences between H_2O and XYZ as stipulated by the thought experiment, then there are simply no grounds for claiming differences between the physical substances H_2O and XYZ, and there is, in principle, no possible method by which one could individuate the substances (identity of indiscernibles)—which is sufficient for claiming their identity, e.g. that H_2O just is XYZ and vice versa.

[47] Cf. e.g. Putnam, H. (1973). "Meaning and Reference," *Journal of Philosophy* 70, 699-711.and Putnam,

Proponents for both embodiment and extended cognition have reason to take issue with many aspects of the twin-earth thought experiment; but the extended cognition thesis owes much of its motivation to the results fostered by semantic externalism. One major idea motivated in large part by the twin earth thought experiment is known as 'multiple realizability'[48]. Multiple realizability is the thesis that one mental state, process, or property can be variously instantiated by different supervenience bases in physical states, processes, or properties. For instance, earthling Oscar and twin-earthling Twin-Oscar are composed of different matter (Oscar's brain is composed mostly of H_2O, while Twin-Oscar's brain is composed mostly of XYZ, for example)—yet [if defined functionally] their mental states, processes, and properties are identical—thus [functional] mental states are (at least in principle, according to the meaning externalist) multiply realizable. This result was thought to have direct relevance to AI research. If human mental states are multiply realizable, the argument went, then they (or their AI equivalents) could be implemented in AI systems.

Thus, if mental states, processes, or properties are multiply realizable, then what individuates them cannot be their particular or unique relations with particular physical supervenience bases. Instead, mental states, processes, or properties are thought to consist in *functional* relations between physical sense-impressions and physical behavioral activity. This 'functionalist' approach understands mental states, processes, and

H. (1975) "The meaning of 'meaning'". In *Philosophical Papers, Vol. 2: Mind, Language and Reality*. Cambridge University Press.

[48] While Putnam's thought experiment did a good deal to provide motivation for the functionalist approach to the mind because it showed how semantics could be consistent with the notion of multiple realizability, it was not the first instance of the idea of multiple realizability. Earlier incarnations appear in Alan Turing's 1950 "Computing Machinery and Intelligence"; Marvin Minsky's 1968 *Semantic Information Processing*, and 1974 "A Framework for Representing Knowledge"; and Herbert Simon's 1957 *Models of Man*.

properties as causally efficacious relational states, processes, or properties relating some

physical input to some physical output. We can understand the relation between mental

and physical on this functionalist account as analogous to the relation between computer

hardware and software: the mind is like software which can be variously implemented on

different hardware architectures. What makes the software *software* is that it produces

certain kinds of output when presented with certain kinds of input—regardless of the

hardware platform on which it is implemented. As such, and in line with meaning

externalist and functionalist interpretations, calling something 'cognitive' or saying that it

'has a mind' is a matter of satisfying functional criteria.

From these assumptions Andy Clark and David Chalmers set forth their extended

mind thesis[49]. They ask, in essence, "if minds, like software, can be multiply realized in

different physical hardware, then what motivates our tacit identifying of minds with

brains?" At least in principle, the functional software of the mind can extend past the

boundaries of the brain; and since the mind is defined functionally, there should be

(again, at least in principle) multiple physically instantiable ways of realizing the criteria

for mind—of taking sense-impressions as input as producing behavioral activity as

output. As such, the extended mind thesis posits that cognition occurs in such a way that

the brain-world boundary is often if not always irrelevant to the functioning of a

cognitive system. From the perspective of the extended mind theorist, the embodiment

thesis amounts to the assertion of partial extension, and in many ways less radical than

the extended mind thesis. The extended mind theorist thinks of the embodiment theorist

---

[49]   Clark, A., and D. Chalmers, 1998, "The Extended Mind," Analysis, 58: 10–23.

as taking the extended mind thesis and stopping the functional extension at, instead of the boundary of the brain, the boundary of the body[50]. There are many reasons why this characterization of embodiment is inadequate, and we will discuss them shortly.

The second area of research from which the extended mind thesis draws is dispersed cognition. In the early 1980s many researchers in AI began to wonder whether computationalist models could adequately handle the task of modeling cognition of human-like intelligence. The computationalist had approached modeling cognition in terms of manipulation of discrete symbol tokens[51]. Since computation works on representational symbolic tokens internal to the cognitive system and not directly on what is presented to the cognitive agent in its external environment, the computationalist approach can be thought of as firmly entrenched within the Cartesian paradigm. It should also be noted that some versions of extended cognition understand the extended mind thesis as merely expanding upon the computationalist paradigm[52]. Instead of drawing the boundaries of a cognitive agent around the brain, the extended computationalist understands the fundamental activities of cognition to involve discrete symbol

---

[50]   For example, Andy Clark—the paradigmatic "Extended Cognitivst" discusses the boundaries of body as a matter of "interface" that can be altered, or augmented in various ways. So, he writes: "Sensing and moving are the spots where the rubber of embodied agency meets the road of the wider world—the world outside the agent's organismic boundaries." (Clark, 2008). And later says that agents that "constantly to negotiate and renegotiate the agent-world boundary itself" are "profoundly embodied agents". Clearly, Clark's vision of embodiment focuses more on this "cyborgic" aspect of re-forming or re-assigning the boundaries of what gets counted as the body, while downplaying the character and quality of the body's sensitivities to specific affordances in the local environment.

[51]   The paradigmatic example of such a computationalist approach is the "physical symbol systems" proposed in Newell, A., and H.A. Simon, (1972), *Human problem solving*, Englewood Cliffs, NJ: Prentice-Hall.

[52]   In his 2010 *The New Science Of Mind: From Extended Mind to Embodied Phenomenology*, Mark Rowlands argues that Andy Clark's 1998 and 2008 formulations of the extended mind thesis are both computationalist in this sense.

manipulation—though this can occur external to the brain, as well as body[53].

In contrast, dispersed cognition divides general functional cognitive tasks into smaller, modularized tasks that can be accomplished in parallel rather than serially. These smaller tasks needn't relate symbolically nor representationally *to the external world—* rather, their significance (symbolic or not) comes irreducibly from their functional relations with lower and higher level functional processes. In the connectionist approach to modeling cognition, mental functions are thought of as extremely complex interconnected networks of simple uniform units[54] (often meant to be, to some degree, analogous to neurons)[55]. The kinds and relations of the connections between units determine the character of the overall function executed. One common and important aspect of most connectionist models is that such interconnected complex neural networks must be dynamic. Somewhat abstractly, at a given time a connection between units in the network can be activated to varying degrees represented by a numerical value called a 'weighting'. For example, a weighting of a connection might represent the probability that the neuron will generate an action potential spike. But one weighting alone does not

---

[53] Hilary Putnam's "division of linguistic labor" (Putnam, H., 1975, "The Meaning of Meaning", in *Mind, Language and Reality*, Cambridge: Cambridge University Press: 215–271.) should be understood in this way. Semantic meaning of, for example, natural kind terms is determined through an extended and external process of epistemic expertise shared socially though linguistic constructs that allow competent (though non-expert) language users to refer correctly to, say, an Ash tree through some properly-constructed causal history that allows particular loci of reference to become meaning-bearing. That is, I—as a non-expert on kinds of trees—am able to correctly (or incorrectly) identify the tree outside my window as an Ash because there is somebody (an expert) who could, if asked, individuate this kind of tree from all other kinds of tree; as long as there is a causal social epistemic link between me and this expert, as well as a causal history of reference linking the expert to the initial 'baptism' of Ash trees.

[54] This is in line with the UNIX programming ethos prevalent in the 1980s popularized by the inventor of the UNIX Pipeline, Doug McIlroy's slogan "Write programs that do one thing and do it well. Write programs to work together."

[55] cf. e.g. David Rumelhart's 1989 "The Architecture of Mind: A Connectionist Approach", Paul Smolensky's 1989 "Connectionist Modeling: Neural Computation / Mental Connections", and Paul Churchland's 1989 "On the Nature of Theories: A Neurocomputational Perspective"; all of which can be found in Haugeland's *Mind Design II* . MIT Press, 1997.

constitute the full significance of a function or action—it is only within the whole of a particular complex neural network that the weighting of this one neuron could take on such a significance; and only if activation of other related neuronal units allows for the spread of the action potential signal in a way characteristic of the function or action for which the activation is said to be significant.

The more general point is that the connectionist or dispersed model of cognition can but does not need to fit with traditional computationalist and representationalist pictures of cognition[56]. If one's goal is to avoid the problems associated with traditional Cartesian and reductionist accounts of the mind, one *could* do so with a properly nuanced dispersed model. Such a model would have to take care to avoid computationalist / representationalist assumptions however. But this at least appears in principle to be possible for the dispersed cognitive model to achieve.

We have already seen one interpretation of the embodied mind thesis that puts it in agreement with the extended mind thesis in some ways, and in disagreement in others. As seen from the extended mind perspective, embodiment is a partial extension from the methodological solipsism of the mind-brain identity thesis. The extended cognitivist sees embodied cognition as the rejection of mind-brain identity, and the embrace of mind-body identity. As such, an extended cognition enthusiast would likely be apt to equally dismiss embodied cognition as a sort of warmed-over identity theory. This would be a mistake. Unlike the methodological solipsism of a mind-brain identity theory, embodied

---

[56] For instance, such networks can be thought of as either continuously dynamic (described by differential functions), or discretely dynamic (described by iterative step functions). The connectionist approaches that emerged in the 1980s were largely discretely dynamic networks—which are (at least in principle) consistent with the discrete representationalism of classical computationalist approaches. Continuously dynamic networks, on the other hand, are not.

cognition constitutively requires and causally depends upon close and direct interaction between body and environment. If we understand the methodological solipsism of "intracranialist" theories (like Fodor's representationalist computationalism) as holding in abeyance or methodologically bracketing as irrelevant to the study of cognition the character of both sensory input and behavioral output, then we should take the embodied cognition thesis to firmly deny the coherence of such a methodology. It makes no sense, according to the embodied cognitivist, to try to theorize about or understand the body as isolated from its local environment, or apart from its social, cultural, and historical situatedness. In fact, the embodied cognitive agent constitutively incorporates precisely this multifaceted situatedness in its complex and dynamically integrated system of bodily affordances and constraints. The very idea of 'body' (*Leib*) for the embodiment theorist must of necessity incorporate all of these elements traditionally thought of as 'external' (in the methodologically solipsist sense). Such affordances and constraints not only causally factor into changing and shaping the embodied system, but must be understood as constitutive of embodied agency—making the embodied system what it is.

Seen from this 'embodied' perspective, the 'extended' criticism now appears off-mark. In fact, the embodiment thesis is now the one that appears more radical. The extended cognition thesis claims that cognition may possibly involve processes that occur outside skin and skull. This thesis is comparably weaker than the embodiment thesis's claim that cognition both causally and constitutively is afforded and constrained by the body's complex interrelation with its social, cultural, historical, and physical environment. The former's position asserts merely the *possibility* of incorporation of

entities external to the brain while the latter's makes such incorporation *necessary*. Far from warmed-over intracranialism, the embodied mind thesis radically departs from the Cartesian and reductionist traditions; and in some ways is more strongly committed to 'extending' cognition than the extended mind thesis is.

One last word on the conceptual relations between semantic externalism, functionalism, and the embodied mind thesis. We've already seen how the extended mind thesis depends conceptually on a commitment to functionalism, and that functionalism stems from the hard-won multiple-realizability of semantic externalism. Thus, to some extent the extended mind thesis is married to semantic externalism[57]. But what is the sense of externalism once cognition is understood as extended? The initial distinction for semantic externalism was between intension and extension. The meaning externalist's argument showed that linguistic intension cannot determine reference, so the only remaining option for reference is linguistic *extension*. Extension is not fixed internally ("intracranially"), so meaning must be external ("extracranial"). But with the extended cognitivist's denial of intracraniality, it is no longer *the head* that bounds inside from outside. The head is no longer the locus of the distinction between intension and extension. But this reveals a more fundamental problem. Where *is* the locus of the distinction between intension and extension? In fact, "intension" and "extension" needn't be conceived of spatially at all. Only under the very specific historical and conceptual circumstances of a received computationalist / representationalist view could such a spatial metaphor for meaning make any sense.

---

[57] But the reverse relation doesn't necessarily hold. One can be a meaning externalist without affirming the extended mind thesis—in fact most contemporary functionalists do.

Fortunately (or, perhaps unfortunately) functionalism provides an escape from the meaning externalist's spatial incoherence problem. Functionalism redraws the relevant boundaries around whatever operations or properties are to count as cognitive. According to the functionalist cognition takes in inputs, operates on them in characteristically cognitive ways (which allow for multiple-realizability), and produces outputs imbued with the hallmarks of having been produced by functions of cognitive agency. Importantly functionalism doesn't dissolve the boundary problems caused by Cartesian dualism and inherited by computationalist / representationalist accounts of cognition—rather it maintains them by de-spatializing them. But it does so at the price of multiplying conceptual boundary problems. Now instead of the problem of spatially locating the 'inside' and 'outside' of cognition, the extended cognitivist is forced to take seriously the need to give the *functional* boundary conditions of cognition. This is a problem with what Adams and Aizawa have dubbed "the mark of the cognitive"[58]. Because the extended mind is married to semantic externalism, it must address the problem of the mark of the cognitive. Instead of answering what counts as inside or outside the head, the extended mind theorist must now explain what counts as functionally cognitive or non-cognitive.

Alternatively, the embodied mind thesis *isn't* married to functionalism or semantic externalism, and so does not *necessarily* inherit the problem of the mark of the cognitive. Of course, an embodied theorist's commitment to either of these doctrines equally implicates their position in dealing with this problem, so mere adoption of the embodied mind thesis is not enough to avoid dealing with the mark of the cognitive.

---

[58] Cf. Adams & Aizawa, (2001) "The Bounds of Cognition". Philosophical Psychology. 14, pp. 43-64.

Instead, one must develop an account of embodiment that does not essentially rely on functionalism or semantic externalism. To my knowledge, few who affirm the embodied mind thesis explicitly address this issue[59]. I think that a coherent and strong account of embodied cognition can be developed without relying on either functionalism or semantic externalism as they have been outlined here[60], and further that if one *can* avoid the problem of the mark of the cognitive, one *should*[61]. So, to briefly recap: in order to avoid the traditional pitfalls associated with dualism and reductionism as well as the more contemporary problems relating to the mark of the cognitive one should adopt some version of the embodied mind thesis that disavows commitment to functionalism and semantic externalism. Further, by interpreting the embodied mind thesis properly, we can understand it as a more radical break with traditional dualism and reductionism, as well as a stronger departure from computationalist and representationalist accounts of cognition than what is availed by the extended mind thesis.

1.4   *Multiple Realizability& Type-Token Identity*

Without semantic externalism and functionalism, we lose previous grounds for asserting multiple realizability. Many find multiple realizability to be a desirable result, so from their perspective, abandonment of semantic externalism and functionalism might

---

[59]   One notable exception is Mark Rowlands (cf. e.g. Rowlands 2006 and 2010). He does a good job navigating these difficult issues, but his account of cognition ultimately relies on yet one more spurious assumption about agential ownership. More discussion on this matter in chapter 4.

[60]   Though what I have in mind does rely on a concept that bears a close resemblance to a notion from philosophy of biology called "biological function". This will also be discussed at greater length in chapter 4.

[61]   That is, rather than take it seriously. Even if one takes seriously the problem of the mark of the cognitive and is able to adequately defeat it, one's position is still seriously undermined.

be seen as 'throwing out the baby with the bathwater'. For without good grounds for multiple realizability what reason is there to think that the projects of AI can be, at their most basic levels, possibly successful? To put the issue in terms more closely related to the issues pertaining to this project, the concern surfaces as an issue of communicative expression: without good grounds for *sharing* semantic meanings (*a la* semantic externalism) what reason is there to think that social practices, at their most basic levels, can possibly be transmitted? If we understand the gains of adopting an embodied or extended account of cognition as finally giving a new basis for overcoming the old Cartesian problem of separate substances, our new worry is that we haven't made *any* progress on the front of Descartes' second problem: the problem of other minds. On the extended and embodied approaches, mind and world are placed in direct causal and constitutive contact with each other—so much so that there is a new worry that they eliminate the meaningfulness of the subject of inquiry (the mind) altogether. But this alone isn't enough to have settled the issue of boundary between one and another mind. Here I would like to highlight some intuitions about multiple realizability, and suggest that perhaps we shouldn't value it as much as the semantic externalism that motivated it; and further that without multiple realizability the hopes of AI researchers are not dashed; rather they are made more realistic.

Multiple realizability, again, is the idea that the same function can be implemented variously on different platforms. A function, it is often said, is the software that takes inputs, operates on them, and produces outputs. Since software doesn't depend

on a particular computational architecture[62] then, so the analogy goes, cognitive functions do not depend on for their successful implementation on particular cognitive architectures. To play off some intuitions about multiple realizability, let's look at how the idea can be seen working in an interpretation of two agents who utter the same phrase. If agents A and B are each told to utter the phrase "Kree Jaffa!", and each then does, there is some sense of the word 'function' in which we would like to say that they underwent the *same* process or function. They each took the input of the instruction to utter the phrase (at some level of abstraction), and by doing some cognitive work, were able to produce output (at the relevant level of abstraction) judged to be—in the relevant aspects—the same. Thus, on this interpretation the cognitive function required to produce agent A's utterance is functionally equivalent to the cognitive function required to produce agent B's utterance. That is, there is one function variously realized on the variant physical architectures (or substrata) of agent A's and B's cognitive resources.

There are a couple ways of understanding this, however. Are A's and B's cognitive resources really different? Or are they two instances (or tokens) of a common type? Using the software/hardware analogy, we can imagine swapping the software (the functional cognitive makeup) of one agent to the other—downloading either agent's mind to the other's body. If such a scenario is possible, what makes it possible? A likeness in bodies? A likeness in minds? Both? What would happen if we attempted to download

---

[62]   In fact many programs are written for a specific computer architecture. For example many Apple programs in the 1990s were designed to take advantage of the proprietary Power PC CPU architecture designed exclusively for use in Apple products. Such programs could not be "ported" to other architecture intentionally as a form of 'vendor lock-in'. However, even software that cannot be ported can be run in "non-native", or "emulated" environments. This possibility is assured by the Church-Turing thesis.

agent A's mind to a lion's body? What about the lion's mind to agent A's body? These

questions hone our intuitions about in what exactly multiple realizability consists.

Wittgenstein famously wrote "if a lion could talk, we could not understand

him"[63]. Though Wittgenstein would not assent to the use of psychologistic language, for

the purposes of our current exposition we can understand this aphorism as Wittgenstein's

taking a stand on the compatibility between a lion's embodied existence (*Leib*) and that

of humans. There are two interesting elements in this aphorism: first, that Wittgenstein

apparently believes that a lion's embodied experiencing is sufficiently different from that

of a human—enough to account for the impossibility of linguistic understanding[64]. This

point is important because it designates a distinction in *type*. Humans and lions are

linguistically incompatible because their embodied biographies are divergently

dissimilar. The second important point is that it is not merely *Wittgenstein* or *I* that cannot

understand the lion, it is *we* who are unable to understand. This suggests that

understanding—a function of linguistic meaning—is not something attained (for *us*

anyway; whoever *we* are) *privately*, rather understanding or linguistic meaning is

---

[63] Philosophical Investigations, §223.

[64] Another possible interpretation is that Wittgenstein believes that the lion is mentally incapable of
linguistic expression in general, or of human linguistic expression in particular. I do not think this is an
adequate interpretation however. In his *Remarks on the Philosophy of Psychology* (volume II)
Wittgenstein writes: "...In general I do not surmise fear in [another person]--I see it. I do not feel that I
am deducing the probable existence of something inside from something outside; rather it is as if the
human face were in a way translucent and that I were seeing it not in reflected light but rather in its
own." (RPP II, § 170) and ""We see emotion."--As opposed to what?--We do not see facial contortions
and make the inference that he is feeling joy, grief, boredom. We describe a face immediately as sad,
radiant, bored, even when we are unable to give any other description of the features.--Grief, one would
like to say, is personified in the face..." (RPP II, § 570). It seems clear from these passages that
Wittgenstein wishes to move past cognitive processing in his account of linguistic expressibility—rather
expression is *perceived* (at a pre-cognitive level). For an excellent discussion of this see Søren
Overgaard & Dan Zahavi "Understanding (Other) Minds: Wittgenstein's Phenomenological
Contribution" In E. Zamuner and D. K. Levy (eds.), *Wittgenstein's Enduring Arguments*. London:
Routledge, 2009, pp. 60-86.

distributed socially. For Wittgenstein meaning is necessarily relational, interpersonal. It cannot be located or isolated privately 'within' one's subjective experience. Action is made meaningful precisely by its performance within a social context. Without social context, there can be no meaning. So, because a lion is a different *type* of social entity— because the lion does not share *our* "form[s] of life" (*Lebensform[en]*)—this is precisely why we couldn't possibly understand it.

The question then arises as to how different social or linguistic types can be differentiated—in accordance with what mechanisms or principles? Additionally, what exactly makes, for example, a lion a different *type* of sociolinguistic entity but another human merely another *token* of the same type? Wittgenstein's position seems to limit the possibility of communicative meaning to the socially significant activities of *token* agents belonging to a social type—or "form of life" as it were. So, if one were able to prove that the differences between one human and another human versus the differences between one human and a lion were, instead of differences in kind, rather differences in degree, Wittgenstein would appear to have a problem on his hands—namely either that we should be capable of understanding lions, or that we should not be capable of understanding other humans[65]. The issue is of how to define the belonging relation for a sociolinguistic type. Constrain the criteria too much and all agents become type-token identical[66], which—if Wittgenstein is correct—rules out the possibility of meaningful language use. Loosen the criteria too much and lions, ants, trees, and rocks become

---

[65] The conspicuous alternative is that Wittgenstein is wrong about the bivalent nature of understanding— that we either [fully] understand, or we [fully] fail to understand. Perhaps understanding comes in degrees (Cf. e.g. Susan Carey's 2009 *The Origin of Concepts*)

[66] That is, each individual agent is its own type—no two agents can be judged to be 'the same' in any relevant sense.

candidates for social and linguistic intelligibility[67].

Another way of approaching the issue of multiple realizability is from the functionalist perspective. For the functionalist, the issue of how to define function carries with it the problem of fineness of grain. Another of Wittgenstein's favored examples in his worries with what has been called the "private language argument" is in how to make sense of personal claims of pain. On the one hand pain is a phenomenal feeling—it is something that is primitively felt by whoever is the recipient of the sensation of pain. In this sense, nobody but I can feel my pain, say, of a toothache. That is, this pain is fundamentally indexed to me, here and now in this instant of feeling it, and it is of a character so singular as to be impossible to accurately express it. As such, this pain is wholly singular. Wittgenstein argues directly against this sort of interpretation of 'private' sensations such as pain. 'Pain', after all is, at its most basic, a *linguistic* expression; and linguistic expressions attain meaning only though the myriad of uses and relations with other linguistic expressions and within a social, language-using community. Thus, to claim that pain expressions are utterly singular can be nothing but nonsense because pain in all cases is what is expressed or expressible—and the utterly singular is, indeed, inexpressible. To Wittgenstein then to be in pain is to bear the right kind of relation to how one behaves and expresses being in a state of pain—there can be nothing more to it.

Here we can recognize two extremes: the too-fine-grained Scylla of pain as being utterly singular (and thus 'private'), and the too-coarse-grained Charybdis of pain being

---

[67] In fact, this notion bears a striking resemblance to Anaxagoras' conception of noesis (νόησης).

something in which all can share (and thus eliminating the concrete possibility of intersubjective misunderstanding). Wittgenstein clearly wishes to avoid both extremes. If 'pain' is to be meaningful at all as a linguistic expression it must be *socially* understood. This rules out the possibility of the [private] fine-grained extreme. However, the meaningfulness of 'pain' is only retained if it does not apply to any and everything. If pain can apply to any and every experience of any and all things ever to have existed, being in pain becomes completely vacuous. But now again we are faced with the dilemma of how to go about deciding what is too-fine-grained and what is too-coarse-grained. To put this problem a little differently, we could ask whether pain is a concept that is multiply realizable. If so, *how* multiply realizable? The Wittgensteinian is forced to understand the notion of multiple realizability as an intensive relational property[68]. Once again, we find that the degree to which a function is multiply realizable depends on what is meant (in context) by the function being discussed. Wittgenstein's response is to look to our actual usages of concepts: how do we use the word "pain"? To Wittgenstein, how we respond to this question is precisely what it means to be in pain, and also reveals the degree to which being in pain is multiply realizable by others[69].

In his discussing of these issues it becomes clear that Wittgenstein does not to believe that giving the severally necessary and jointly sufficient conditions for belonging—for all and only the proper usage of language—is possible. Instead, he employs the metaphorical device of "family resemblances" to express a novel conception of inclusion without reliance on rules giving the conditions of inclusion. Thus,

---

[68] That is, a relational property expressible as a difference of degree. A function is not merely multiply realizable or not—rather functions must be understood as more-or-less multiply realizable.
[69] *Philosophical Investigations* §244 – 253.

Wittgenstein's 'family resemblance' approach attempts to give criterionless—yet still meaningful—conditions of appropriateness to distinguish between types and tokens[70].

I think there is much in Wittgenstein's approach that warrants thoughtful consideration, but I will beg-off prolonged discussion of this until chapter 2. Until then it will have to suffice for us to acknowledge that the problem of type-inclusion is not one that is programmatically solvable, and that multiple realizability is intimately tied to its solution. There is additionally the trouble with what to do about degrees of similarities and differences in body plans, as well as how to approach degrees of similarity and difference in the cultural, social, and biographical histories of individual agents. Recall that the question with which we are here preoccupied is in addressing how social practices are possible for embodied agents. If these difficulties prove insuperable the question must be answered in the negative (i.e. social practices are *not* possible). If they are not insuperable, I bear the burden of showing how this is [metaphysically and epistemically] so.

1.5   *Enactive and Ecological Agency*

We have already briefly examined some of the relationships between representational or computational approaches to cognition, the extended mind thesis, and the embodied mind thesis. Both the extended mind and embodied mind theses attempt to set themselves apart from the computational and methodologically solipsistic approaches to studying the mind or cognition. Both claim to make radical departures from this tradition while downplaying the radicality of the other's departure. While the extended

---

[70]   Ibid. §66 – 71.

mind thesis makes great strides in conceptually freeing itself from the dualistic strictures that have provided the basis for centuries of traditional thinking, it is at the same time incapable of completely overcoming them entirely. Part the problem for the extended mind thesis is its negative formulation, which depends on and derives from taking the Cartesian way of thinking to be coherent. To claim that the mind is extended is ultimately to simply redraw the spatial bounds of the mind—not to dissolve or 'unthink' those boundaries. This is evidenced by the need for extended cognition to take seriously the problem of the mark of the cognitive[71]. Alternatively, the embodied approach to mind fundamentally relies upon precisely the indiscernibility of clear boundaries between body and world. To this end, it is crucial to understand what *kind* of entity we are discussing when we affirm cognition as embodied.

The claim that the mind is embodied is not simply the claim that the mind is located inside a body. As a thesis that attempts to question the very foundations of centuries of study of the mind this would not be very interesting. In fact, depending on how it is understood, this formulation of an "embodiment thesis" is a rare instance of a claim that may enjoy universal assent. On the contrary, the embodied mind thesis is not just that the mind is covered by body, so to speak—that the mind *has* a body—but rather that body is essential to or *constitutive of* the mind, or cognition; that without the body there is *no* sense in which one could be discussing minds or cognition. All concepts involved in discussions of the mind are precisely concepts about *lived bodies* (*Leib*); that is, the mind *is* the body; the body's actions are the actions of an *agent*. But also as

---

[71]   This is precisely what Andy Clark does in a sustained way in parts II and III of his 2008 *Supersizing the Mind.*

discussed in contradistinction to the extended mind theorist's characterization of embodiment, the embodied mind thesis asserts that in order to understand and explain cognition we may not stop at the putative borders between flesh and world—we may not become methodological solipsists. According to the embodiment theorist, lived bodies (*Leib*) are special kinds of entities that are sensitive and open to the various constraints and affordances of their local environments. One condition for the possibility of being open to the various constraints and affordances of their local environments is that bodies move and can be moved. Bodies are in close, "high bandwidth" coupling or interaction with that in the local environment to which they are differentially sensitive.

There are two distinct ways of understanding this close, "high bandwidth" interaction. The first is the embeddedness thesis, the second is the ecological mind thesis. Both of these theses depend on a more general claim about movement—what can be called the enactive thesis. I'll first discuss what it means for an embodied agent to be *enactive*, then show the differences between *embeddedness* and *ecological cognition* respectively.

The idea of *enactivism* emphasizes the motile aspects of cognitive agency. It is the claim that in order for something to have or be a 'mind' it must have or be a mobile 'body'. In this way cognition is fundamentally tied to movement[72]. There are at least two ways for this relation between cognition and movement to be 'fundamental'. The first is in that the two are *causally* related—that either cognition causes movement, or that movement causes cognition. Neither of these is a particularly controversial claim.

---

[72] Because movement is trivially tied to bodies (e.g. nothing but spatially-related, extended bodies can move—movement is a spatial relation).

Without any stimulus, without being affected by anything, the character of cognition would be radically affected—so much so that it would no longer be clear that we are even discussing cognition at all[73]. On the other hand, if cognition could not cause movement, it is hard to imagine what cognition *does* or is *for*. To the extent that cognition involves some temporal process or function, there must be some output or result of having cognized—a cognitive *product*. So the idea that cognition is causally related to movement or activity[74] should be in no way contentious[75].

The second way in which the relation between cognition and movement can be thought to be fundamental is in that either, or both are constitutive of the other. To say that A is constitutive of B is to claim that B cannot be understood to be what it is without A's being what it is, in relation to B. For example, when I say that "copying others' work is constitutive of plagiarism", I am claiming that the concept of plagiarism includes in its definition what it means to copy others' work. Without the inclusion of copying others' work in this definition, we simply would not be talking about plagiarism—we would be discussing either something other than the concept of plagiarism, or perhaps nothing at all. So there are three potential "constitutive" relations between cognition and movement. Either

    (i)   cognition constitutes movement,

---

[73] That is, so long as cognition is fundamentally *intentionally* structured. Cognition must be *about* something, and nothing can be about anything without causality. George Kampis argues this persuasively in his 2002, "The Natural History of Agents", in: Gulyás, L., Tatai, G., Váncza, J. (ed.): *Agents Everywhere*, Springer, Budapest, pp. 24-48.

[74] Cf. e.g. R. F. Port, & T. van Gelder (Eds.), (1996). *Mind as motion: Explorations in the dynamics of cognition* (pp. 1-43). Cambridge, MA: MIT.

[75] In a certain sense, it is precisely the intuitiveness of the relation between minds and causation that grounds the problem of separate substances (occasionally called the problem of mental causation). Without such a tight (at least intuitive) connection, there couldn't be a 'problem of mental causation'.

(ii)   movement constitutes cognition, or

(iii)   cognition and movement are co-constitutive[76].

The easiest case to consider is (ii). If movement constitutes cognition then there can be no cognition without movement—what it means to cognize involves movement. While not entirely vacuous (viz. it is in fact a substantive claim[77]) it seems fairly innocuous to say that (ii) is true. Whatever we may mean by "cognition" it must implicate some kind of movement; whether the motive result is functional or physical. Cognition necessarily has a motile product. The first case, (i), must be considered with care. We need to think carefully about what is meant by 'movement'. For this we should return to the distinction between lived (*Leib*) and merely extended (*Körper*) bodies. If one's intuition is to say that merely extended bodies can move (for example in the sense that planets move) then it seems an abuse of language to claim that cognition constitutes movement—to say that what it means for something to move implicates the exercise of some kind of cognition[78]. If instead we define movement as *volitional*—dependent upon some agent's will, decision, or intention to act in a certain way—then the claim that cognition constitutes movement becomes the claim that an agent's movement implies that

---

[76] Outside the claim that there is some constitutive relationship between cognition and movement, there is a fourth option: (iv) that cognition and movement bear no constitutive relation to each other. Even when no constitutive relation is borne, there may still be some causal relation between them—and likewise, even if there is no discernible causal relation, it is still possible that cognition and movement bear some constitutive relation to each other.

[77] The dictum of methodological solipsism, for instance potentially denies the relevance of movement external to the brain barrier. If this is so, then it is possible claim that such methodological solipsists could deny (ii). Since methodological solipsism has already been disqualified, I shall not dwell further on this issue.

[78] Barring claims about the "mind of God". If one insists that movement necessarily implicates the existence of the mind of God (as Descartes seems to argue as the basis for our inference to the existence of *res extensa* in MeditationVI), then we may ask such a person if they aren't equivocating on the term "mind". Of course, these discussions, while interesting, take us too far from our current inquiries.

the agent utilizes cognition[79]. If volitional action originates in cognition, then we can comfortably affirm (i) as well[80]. Finally, to say (iii) that cognition and movement are co-constitutive is to claim that the two are defined in ways that rely on the other—that neither could be what it is without the other being what it is. Without cognition could there be movement? Without movement could there be cognition? If the answer to *both* is "no" then we affirm (iii). I am inclined to affirm (iii), which means I assent to both (i) and (ii) as well. I assent to (ii) in particular because I am committed to the notion that merely extended bodies (*Körper*) derivatively owe their conceptual existence to a more "primordial" lived embodiment (*Leib*). This commitment will be defended in chapter 4[81].

To say that the mind is embodied and *embedded* is to say that the physical interaction between a cognitive agent's body and the rest of the world strongly causally constrains the possible behaviors of that agent. This in turn causally influences whatever cognitive processes can occur for the embodied and embedded agent[82]. Defined in this way, the embedded mind thesis is a comparably weak claim about the *causal* relations between the embodied mind and the external world. Given the conceptual tools

---

[79] Of course, the definition of movement as volitional smuggles in a potential vicious circularity: if we understand terms like "will", "decision", and "intention" in cognitive terms, then all the addition of volition does is tautologously define cognition in terms of cognition. This is unacceptable. What is needed is a non-cognitive basis for volition. In chapters 3 and 4 I discuss one way to express the notion of [normative] decision making in a way that doesn't question-beg cognition.

[80] I have up until this point been discussing cognition as a particular aspect of mental life. Many understand cognition—thinking—to consist exclusively in the manipulation of non-derived, propositional content. Here I will need to deviate substantially from this "received" view. Instead I wish to hold in abeyance the precise character and details of what comprises cognition. Because I am not wed to this "traditional" view of cognition as trafficking solely in propositional content, it makes it easier to avoid entanglements in 'mark of the cognitive' arguments.

[81] But to anticipate the point, roughly: what makes "mere" objects (*Körper*) *objects* is that they are fundamentally objects *for some subject*—and subjects are embodied (*Leib*). Without subjects, there can be no objects.

[82] For an early explication of the idea of embeddedness see John Haugeland's "The Mind: Embodied and Embedded" in *Having Thought*. 1998. For criticism of embeddedness see Mark Rowlands' *The New Science of Mind: From Extended Mind to Embodied Phenomenology*, MIT Press. 2010.

developed earlier in this chapter, it should now be easy to see that the claim of embeddedness, while "high bandwidth" still fits squarely within the linear, Cartesian, 'dual-stream' view of cognition. It maintains and assumes an inside / outside distinction (though here shifted to the boundaries of skin and world) that the extended and embodied approaches have attempted to mitigate. As such it inherits all the same 'old' Cartesian problems.

The enactive approach to cognition owes much of its conceptual underpinnings to the groundbreaking work of J.J. Gibson in the 1960's and 1970's. As compared with the embeddedness thesis, Gibson's vision of perception emphasized the *constitutive* relations between the embodied perceiver and the 'ecology' in which a perceiving agent is always enmeshed. In the 1970's Gibson developed his theories of direct visual perception and ecological affordances[83]. His interests focused on the visual modality of perception instead of other ways in which agents participate with and within their ecological milieus, but he did not intend to privilege vision over these modalities. Vision just happened to be his psychological specialty, and the area in which he did his work.

Gibson's ecological approach to perception marked what he took to be a drastic departure from the way perception was being studied from the 1950s to the 1970s. This "received view" from which Gibson set himself apart followed in the Cartesian, Lockean, and Kantian traditions of treating perception as unidirectional. The agent, in her capacity as a perceiver plays a *passive* role taking in, for example, visual stimuli as it

---

[83]   Cf. (1972) *A Theory of Direct Visual Perception*, (1977) *The Theory of Affordances*, and most significantly (1979) *The Ecological Approach to Visual Perception*.

spontaneously bombards the retinal disk[84]. Gibson's approach put the perceiver in a more active role, according to which perception and action are closely interwoven and mutually constraining. It is not that visual stimuli cause perception, as the "received view" had it, but that the perceiving agent in many ways causes the visual stimulus though her active movement through her surroundings[85]. This is striking in experiments with the *Ganzfeld* ("whole field" [of vision]) in which subjects were deprived of visual stimulus[86]. Through saccadic eye movement, subjects reported visual perception where they knew none was possible. Gibson's explanation was that there is "invariant" information embedded in what he called the "ambient optic array". The "ambient optic array" is the particular matrix of light as it is from each and every possible perspective within a given environment. As an agent moves about her environment, the ambient optic array is differentially affected by the agent's movement. Through this movement some aspects of the ambient optic array take on a second-order invariance—their differential alterations occur in characteristic and predictable ways[87]. Gibson argued that this second-order invariance is what allows for us to perceive our surroundings in terms of objects with surfaces and textures, and helps determine which objects in our environment are moving and which aren't. What is important, however is that the invariant information embedded in the ambient optic array can only be obtained through agential movement. For Gibson, in a very concrete sense, movement is both causally necessary for and

---

[84] To be clear, Kant's theory of mind is active, but not environmentally or ecologically active – it doesn't act in the world, it acts on itself. The charge of unidirectionality still sticks.

[85] Ibid, pp. 170-188. This, however is different from the Kantian view that an intellect completely constitutes her world by exercising the spontaneous synthesizing powers of reason (schematism, categorization, etc.).

[86] Ibid. p. 151.

[87] Ibid. pp. 73-88.

constitutive of perception.

According to the received computationalist / representationalist view of perception, perception is *mediated* by algorithmic inferences or manipulations of internal symbols or representations. What is perceived depends equally on the input supplied by sensory stimulation of light, as well as the representational symbolic manipulations and computations employed in processing that information. Gibson's view of perception is, on the other hand, not mediated by computational algorithms, inferences, or representations—rather he saw perception as occurring *directly* through the *unmediated* differential experiencing of invariants in the ambient optic array as it is differentially affected by agential movement. These invariants are what reveal to a perceiver all objects, motions, and activities in the agent's local environment. The result of perceiving invariance in the ambient optic array is that no mental representations, processes, or inferences are required for visual perception. The activity of perceiving is not a unidirectional process resulting in an internal representation of the external environment; it is rather the direct experiencing of environmental invariants.

One significant result of Gibson's ecological theory of affordances is that many of the cognitive activities that were previously thought to necessarily occur in the brain were shown to possibly occur external to organismal boundaries. For the visual modality the ambient optic array serves as a case in point. Additionally, Gibson believed that the normative values and meanings or significances of situations, events, objects, and persons are not internal to a perceiving agent. Thus, it would be accurate to say that Gibson too thought that "meanings just ain't in the head". Instead, he thought that values and

meanings acquire their valuation and meaningfulness through situations that directly

reveal environmental affordances and constraints, or potentials for action, to agents. For

example, imagine a friend choking on chicken bones. Without the knowledge of how to

properly perform abdominal thrusts or CPR, we find ourselves constrained—unable to

provide help—by this very lack of knowledge. This situational constraining (or

affordance in the case of one who *can* perform abdominal thrusts or CPR) is what,

according to Gibson, normatively imbues situations with value and meaning. That the

situation *shows up as* affording or constraining particular agentive actions is precisely in

what value and meaning consists.

The radicality of Gibson's position cannot be understated. By locating cognitive

information in the ambient optic array itself, and not within some computational process

occurring within a brain, Gibson's theory of visual perception already depended on the

truth of the extended mind thesis—predating its Clark & Chalmers formulation by more

than two decades. By recognizing that this information only becomes significant through

organismal movement Gibson realized that the activity of perception (and by extension,

cognition) fundamentally relies on the close, nonlinear, dynamic coupling between active

embodied agents and their environments[88]. Further, he was sensitive to the fact that not

only is the agent differentially affected by the constraints and affordances provided by the

local environment, but that the agent herself differentially affects those same constraints

and affordances through her differential activity within those environs—that is, Gibson

recognized the *ecological* nature of agency. With this in place, Gibson was able to

---

[88] Ibid. pp. 133-143.

explain how agents come to perceive *themselves* as the sources affordances and

constraints, as well as perceiving their environment as constitutive parts of themselves[89].

1.6   Cognitive bloat and other colors of herring.

The move to reconceptualize cognition in a way that is inclusive of factors and

elements external to the brain, or on many interpretations external to the body, opens up

the door to a worry first articulated by Robert Rupert and popularized by Mark

Rowlands' reference to it as the "cognitive bloat" problem[90]. The cognitive bloat problem

is the worry that when we are willing to accept as constitutive of cognition elements and

factors that occur external to the brain and/or body, then we had better be able to clearly

demarcate in all cases where the boundaries of cognition are located. Without being able

to give such a demarcation, cognition can, as it were, "spill out" into the world,

encompassing and including any and everything that exists. The entire universe, so the

cognitive bloat argument goes, could constitute the cognitive apparatus of a single

cognitive agent—and this is preposterous. A related problem also falling under the

heading of 'cognitive bloat' is in not being able to define the limits of the subject under

cognitive investigation. The worry here is that when we allow for the extended mind

thesis we begin failing to be able to articulate *what exactly* cognition is; so we begin to

fail to identify *what exactly* we are studying. Without being able to define the subject

area, so the worry goes, we cannot possibly make progress in solving its [exactly

---

[89]   Ibid. pp. 182-208.
[90]   cf. Rupert, Robert.2004. "Some Problems for the Thesis of Extended Cognition." *Journal of Philosophy*
101:389 – 428. and criticism, Mark Rowlands' *The New Science of Mind: From Extended Mind to
Embodied Phenomenology*, MIT Press. 2010.

what's?] problems; instead condemned to fruitless cognitive research programmes—and this is preposterous and self-defeating. The problem of cognitive bloat can be put by way of analogous example: consider the task of answering a friend's query about tomorrow's weather forecast:

Beers, Bars, and Barometers [BBB] example:

You and your friend are seated in a pub in the center of town when your friend asks you "what does the local weather authority forecast for tomorrow's weather?" Not wanting to disappoint, you assess your current situation with a mind toward adequately and correctly responding to your friend's query. After much consideration, you remember that another friend both owns an iPad and is currently working at the café down the street. You surmise that this friend is likely to both have his iPad and to allow you to borrow it for these purposes. The café, if you recall properly, has wi-fi internet access, and you trust that you know how to search for the local weather authority's forecast using your friend's iPad.

The question is whether you are cognitively capable of adequately and correctly responding to your pub friend's query. Many would intuitively say "yes, you are capable of this cognitive task". But with worries of cognitive bloat in mind, perhaps we should pause and consider wherein lies such an affirmative response. We should focus on what is required to successfully complete the task at hand, as well as what aspects of the completion of this task should be considered genuinely cognitive. The enactivist intuition

is precisely that implementation is integral to (or constitutive of) cognition—nothing can be considered cognitive if it cannot be *enacted*.[91] But first, let us recognize a bit of a puzzle. The worry about cognitive bloat only arises when we suspect that there aren't any, or aren't enough, necessary and sufficient conditions defining the boundaries of what is to count as 'cognitive'. So to address the issue of what aspects of this example task should be considered genuinely cognitive is to presume that we already have sufficiently established such boundary conditions (viz. the "mark of the cognitive"). And if such boundary conditions are already established, then this example should prove wholly uninteresting.

First, in order to be capable of completing the task of telling my friend the weather, I must have some idea of what the task is that needs completing. I must interpret my friend's utterance in a way that sets up a problem or task for completion; and which gives at least rough criteria for success or failure. That is, I must understand my current situation as a problem, a task—I must recognize that something must be done. My actions should constitute a sufficient basis for both me and my friend to judge whether I've met those criteria or not[92]. Interestingly, if I begin to act by hailing the waiter to settle my tab, my pub friend may judge me not to have properly understood his query—even though this might be the first action in a series of planned actions which are intended to result in

---

[91] This central tenet of enactivism is shared by a majority of its proponents (e.g. Alva Noë, Mark Rowlands, Evan Thompson, De Jaeger & Di Paolo, Varela et. al., and perhaps Clark). For further discussion see Clark 2008, De Jaeger & Di Paolo 2007, Varela 1997, Thompson 2007, Noe 2004, Rowlands 2010.

[92] Of course, my friend and I can disagree as to whether particular criteria are met, or about which criteria are significant. We may have a conversation (or argument) about this very issue, or [perhaps incorrectly] assume that the other operates with the same criteria as our own. Difficulties can and do often arise in communicative and interpretive tasks such as these; and so a theory's ability to adequately represent the character of these difference should be viewed as a merit rather than as a demerit.

correctly responding to his query. Judgment of satisfaction of criteria such as these is often executed in the impoverishment of sufficient evidence.

After interpreting my friend's question, cognitively speaking, I engage in a sort of playful heuristic search for viable ways in which an appropriate response might avail itself. Included in this search might be considerations such as whether I want to take the query seriously in the first place, or if I can quickly respond with a pun. I attempt to remember if I had previously seen a forecast or not, and—regardless of the actual facts—I am constrained by a self-imposed (though implicit and undefined) time limit in such considerations—I decide that I must not have seen a forecast previously, at least not that I can recall under such constraints. I move on to consider what actions I could take in order to find out. Hail the waiter and ask him? Maybe, but the waiter's response may be unreliable. Scan the room for a television. None to be found. Give up entirely? Perhaps—but wait! Aha! I've recalled I have a friend who works nearby; and even better, this friend is likely to have an iPad with internet connectivity. I'll settle my tab, head over to the cafe, and ask to borrow my friend's iPad. The plan is now set.

Was this plan hatched in my head alone? Did the delicious beer have any effect on it? Perhaps. If it did, does this effect count as *constitutive* of cognition? Well, counterfactually speaking, we can ask whether the cognitive task would have been executed differently were it not for the effect of the beer. If the qualitative character of the task is made sufficiently different by the consumption of beer[93] then the beer is minimally a *causal* influence in the decision. But what would it take for the beer to be

---

[93]   Though here we encounter the problem of how to cash out the notion of "sufficiently different". How could one possibly compare what actually occurs with what didn't? How could one possibly make sense of such a comparison? Such problems are inherent to counterfactual thinking.

constitutive of cognition? Of *this particular instance* of cognition[94]? Intuitively we might want to say that this particular beer may be constitutive of this particular instance of cognition, but that beer (in general) is not constitutive of cognition (in general). But cognition is never actualized *in general*. It is always actualized in specific circumstances, and if particular instances of drinking beer each constitutively factor into particular cognitive activities, it seems reasonable to say that (at least in each of those circumstances) beer factors constitutively in cognition.

Here I have stumbled into what Fred Adams and Kenneth Aizawa have called the "coupling-constitution fallacy"[95]—the idea that proponents of the extended mind thesis fallaciously conflate the notions of causal coupling and constitution. It may be argued that what I have said in the previous passage amounts to such a conflation. The upshot of the previous paragraph is to show that (C) there can be no severally necessary and jointly sufficient general criteria for counting as cognitive (viz. bearing the so-called 'mark of the cognitive') because (P1) all cognition occurs in particular instances, and (P2) all particular instances bear purely causal relations (therefore, all criteria are necessarily particular, and not general). Another way of expressing this is to say that—as a committed physicalist—I claim that in a concrete sense, *everything* bears purely causal relationships—that *everything* is constituted through its causal relations. This supposition

---

[94]   The difference between these two questions is interesting. Is it possible that something be constitutive of cognition in one particular circumstance, but not be allowed to count as constitutive of cognition generally? I tend toward giving an affirmative response here—but only because I think a general concept of cognition is either incoherent or empty. On the contrary, I suspect that many would say that to "count as cognitive" is to exhibit the appropriate general characteristics—so nothing could count as cognitive in particular circumstances and not generally—because to "count as cognitive" is just to possess the necessary and sufficient [general] conditions. This is precisely the position against which I here argue.

[95]   cf. e.g. Adams F, Aizawa K (2008) *The bounds of cognition*. Blackwell, Malden, MA

does conflate cause and constitution in the sense that it conflates the more general notion of identity with the causal structure of the world. But does it conflate the causal "coupling" of cognitive and non-cognitive systems with the notion that causally coupled systems constitute a greater system bearing the so-called 'mark of cognition'? To this I wish to respond that it does not because I do not purport to give *any* such general criteria for counting as cognitive; which are what is asked-for by the requirement of the "mark of the cognitive". Another way of seeing my position is to understand all cognitive systems as unique, and provisional—defined by their concrete, particular, and dynamic causal relations. Because of this uniqueness, to ask for *general* inclusion criteria for what gets to "count as cognitive" [bear the "mark of cognition"] is to ask for an explanation that must necessarily distort any and all actual moments of cognition. Whatever it could mean to "count as cognitive" [generally] is precisely the extent to which its application to a particular system is distortive rather than explanatory.

The question as to whether I would be able to have cognized the task of figuring out the weather *in this way* is different than the question as to whether I would have been able to cognize the task [*at all*]. The former question understands instances of cognition as type-token identical. Each act of cognition is essentially unique. The second question understands instances of cognition as tokens of a type—the type in this circumstances being viable cognitive solutions to the particular query. So regarding the BBB example, we may ask whether the beer I consumed is constitutive of the general cognitive task "*one*" might employ in order to adequately respond to my pub friend's query. That is, is the beer *necessary* for *any* adequate solution? And to this our intuitions likely suggest a

response of "no", the beer happens to be constitutive of this particular instance of cognition, but is not constitutive of cognition pertaining to any solution to the query. But again, we may ask, why should we be concerned with "any" [generic] solution rather than *the specific* solution availed in *these particular* circumstances—especially when all cognitive situations are instances of the latter, and not the former? To put the issue here more abstractly, and to make the point more pertinent toward our discussion, it is not at all obvious that we should be attempting to describe how cognition is *generally*, because doing so appears to cover over and ignore salient cognitive aspects that are inextricably tied to functioning of actual and particular cognitive situations.[96] The worry here is that by focusing on general criteria for "counting as cognitive", we run the risk of ruling out specific, particular—but genuinely cognitively constitutive—factors indexed exclusively to the particularities of a situation.

What about the physical and sensorimotor activities involved in my particular solution to my pub friend's query? My solution requires that I relocate myself to the café down the street. In different circumstances where I was immobilized for example— perhaps by another pub-goer attempting to tranquilize a monkey with a blowgun who for some reason accidentally missed and instead hit me—I would not be able to rely on, or assume I could readily rely on the relative ease at which I am able to move myself down

---

[96]   In conversation, Megan Altman drew my attention to the similarity between seeking general "mark of the cognitive" criteria and the misapplicability of Kant's Categorical Imperative. For Kant, one's duty is determined solely by the purely general formal laws of reason—which in their generality cannot account for the particularity of any situational context. It is precisely because the categorical imperative is general, and abstract, that it applies; but it is equally precisely because it is general and abstract that is necessarily excluded from the particular context of any situation. The categorical always applies, but can never be applied. Likewise, generalized definitions of cognitions may apply (for example in attempting to pick out all and only those things capable of cognition) but can never adequately characterize the particular processes [of cognition in action].

the street to the café. That is, my solution depends on the effective use and availability of

sensorimotor affordances. Without such affordances obtaining, my solution is not just a

bad solution—it ceases to be a solution at all. If I cannot get myself down the street to the

café, I cannot complete the task, or answer my friend's query. I would have to either

reassess my situation, or give up. Thus, since my solution necessarily relies upon the

sensorimotor affordance of successfully walking down the street, part of the cognitive

task involves walking—the activity of walking too is constitutive of cognition[97].

Further, let us consider the role(s) of social practices as they may factor into this

issue of cognitive bloat. As with the effects of beer and the sensorimotor affordance

profile of walking down the street, I here claim that social practices are also constitutive

of cognition. In myriad ways our personal understandings of what it means to act within

social settings function to constrain and afford various particular and possible actions. In

the example above, my first action was to hail the waiter and ask for my check. My claim

here is that this action is *also* constitutive of the cognitive task of telling my pub friend

tomorrow's weather—that without engaging in the act of hailing and the subsequent

social practices associated with bill-paying etiquette, I would not be engaged in the same

sorts of activity as I would were I not to have hailed him. Leaving the bar having not paid

---

[97]   Again, I will be charged with falling prey to the "coupling-constitution fallacy". To be clear, the
conflation of causality and constitution is a blurring of a distinction with which I am comfortable. I do
not believe that there is any "fallacy" in it. One only sees a fallacy when one [unjustly] assumes a
commitment to the applicability of general "mark of the cognitive" criteria to particular instances of
cognition. When we question the very meaningfulness of "cognition" as a general term, it is no longer
obvious that there is any fallacy at all. Others (especially Menary, Clark, and Sutton in Menary, 2010)
have taken issue with Adams &Aizawa's "coupling-constitution fallacy", generally emphasizing the
idea of integration in defining cognitive systems. Defining cognitive systems as constituted by relations
of integrations rather than of coupling effectively works to skirt such criticisms. Others such as
Thompson & Stapleton (2009) have made similar arguments. I am sympathetic to these approaches, as
well as the idea that integration is significant in defining cognitive systems—but my approach needn't
avoid the charge—rather it faces it head-on.

the tab gives an entirely different social (and I argue, *cognitive*) character to my actions. It is the difference between adequately responding to my pub friend's query in a socially acceptable way, and doing so in a socially unacceptable way—and in many cases the social constraints are so stringent as to make this difference in acceptability be effectively prohibitive. Under such circumstances, the socially unacceptable "options" often fail to arise as options at all. I may not even imaginatively entertain them because of their social unacceptability, or the fact of their unacceptability could be enough for immediate dismissal as viable options. As such, social considerations and activities can and do factor constitutively into cognition.

For current purposes we can conclude that the problem of cognitive bloat only arises as a problem for someone who wishes to nail-down what cognition is *in general*. I have shown that any such general definition of cognition necessarily ignores or represses an aspect of all *actual* cognitive activities—namely that they, in each and every case, occur within some situation under some particular set of cognitively relevant circumstances. Each situation uniquely and differentially exhibits its own particular constraints and affordances. These constraints and affordances include salient aspects of the local environment, as well as salient aspects of the agent for whom they are constraining or affording—after all, if what I say is true, then there is no steadfast ontological boundary between the two. If one would like to ask the question of what cognition is—one must look to particular occurrences thereof. One should not generalize or abstract away from them. The demand for a "mark of the cognitive" is thus not only unmotivated, but detrimental to cognitive research.

1.7   **Dynamic Embodied Agency**

So far I have given critical arguments against viewing cognition and agency in certain ways. I have suggested problems with Cartesian, reductionist, [purely] computational / representational, embedded, and even some versions of extended, embodied, and enactive theories of cognition. In this section I would like to sketch a positive construction of a position that I think is capable of navigating the rough seas of agency scholarship and its criticisms. This picture will depend for its full effect on a vocabulary to be explicated further in chapters 3 and 4. As such, this should serve as a warning that my comments here should be understood as merely preliminary and provisional. More serious explication will appear shortly, once all the conceptual tools have been put in place.

As may have been evident in previous sections of this chapter, I posses a great respect and affinity for J.J. Gibson's approach to visual perception and the 'enactive' approach to cognition that followed. However, my purpose here is not to give an account of vision or perception. Instead, one should understand this as a work in the metaphysics (and related epistemology) of agency—addressing the question "what are agents such that our phenomenal experiences can arise and be meaningful as they apparently do?" In order to ask this question, we must already take seriously that our phenomenal experiences are in some ways primitive—we take as our starting point the veracity of experiencing. This has the effect of tying together metaphysical and epistemological claims. Whenever we attempt to make claims about what there is and how, we necessarily imply that the answers we give can and should be meaningful *to us*. Thus, any

metaphysics implicitly carries with it the epistemological baggage of bearing meaning. A metaphysics that is not, or has no possibility of meaningfulness (*to us*) is one that is ignored[98] (blind). It is literally and effectively insignificant. Similarly any epistemology that has no bearing on how things are (*in the world*), or on what there *actually* is, is one that is toothless (empty). Therefore, we must take seriously the phenomenality of experiences because the alternative (metaphysics which does *not* account for phenomenality) is incoherent. If anything, this is what licenses our taking of phenomenal experiences as primitive.

There are a great many varieties of phenomenal characters of experience and ways of seeing the world or one's situation. Again, depending on fineness of grain, one might even say that there are as many varieties of phenomenal characters of experience as there are moments of experiencing. This is, in some sense, another way of saying that every moment of one's life is unique (diachronic uniqueness), or that each of us experiences our lives in a way that is qualitatively and phenomenally different from everyone else (synchronic uniqueness). Of course at coarser grains of analysis we encounter similarities[99]. To assert something like the embodiment thesis is to make a somewhat coarse grain level assertion about what is or can be significant for many of us—but it isn't to say that there is something special about this level of analysis, nor is it to say there is something special about who is included in the class of entities for which the embodiment thesis can be significant. In this way, much of what I have to say, and

---

[98] Just as an ethics that has no practical bearing on particular situations is useless (cf. Footnote 96).
[99] Finding similarities, detecting patterns—these are the benefits of abstraction; and the follies. As discussed earlier, it is precisely the extent to which similarities and patterns can be detected that they distort what is actual.

believe to be compelling is simply a matter of my being impressed with a certain treatment of a subject matter at a certain level of analysis. My being impressed, and my opting to write about it betrays the functioning of an additional normative element—I think that you too, if you are reading this, should be impressed by similar arguments. The point here is merely to acknowledge the contexts and degrees of commitment to which I feel obliged. Overall, this argument can be read as one big conditional: *IF* you think that this is a compelling level of analysis, and *IF* you think that we are licensed to take as primitive phenomenal experience *then* you should be willing to accept the kind of account proffered here.

As I mentioned, I have a particular affinity for J.J. Gibson's theories of visual perception and affordances. This affinity is based in an intuitive agreement between how I feel my phenomenal encounter with the world is, and how I understand Gibson's theories to fit with my understanding of my phenomenal experiences. Argumentatively there is an important point to be made here: I do not wish to claim that my phenomenal experience and my understanding of Gibson's theories of visual perception and affordances should extend or apply to anybody else. Thus, there is a crucial skeptical thread I wish to weave into my argumentative structure. I wish to grant that others' phenomenal experiences and understandings are—at least possibly—radically different than my own. I grant this possibility while recognizing that it comes with both metaphysical and epistemological strings attached.

My skeptical claim interpreted metaphysically amounts to granting the possibility that what another is, may be a fundamentally different *kind* of thing than what is made

significant by my phenomenal experiencing[100]. In a certain sense what I am saying is that

my phenomenal experiencing is [potentially] different in kind than anything else in the

universe. Epistemologically however, I am not in a position that allows me to make a

claim one way or another. I could not possibly know how others' phenomenal

experiences are [if there are any]. This is not equivalent to the doubt that other

phenomenal experiencing occurs, rather it is holding in abeyance all and only what

cannot in principle be or become meaningful *from here*. The facts about how others'

phenomenally experience could not possibly be meaningful *to me* because they are in

principle out of the bounds of my experiencability. As such, my claims may strike one as

wildly inaccurate, and phenomenally false. To this all I can say is that perhaps our

experiences *are* wildly different than each others; since neither of us could know, it

cannot matter.

This is what makes communication possible—that one is always in a position of

being incapable of knowing for certain whether the other understands or not—or what of

an expression is understood, and how. Instead if we are to arrive at such a judgment it

must necessarily come from underdetermined and inadequate evidence. The result is

often that when we take others to understand, we do so provisionally—their

understanding is "good enough for government work"; it is pragmatically, and

epistemically adequate (for now, until proven otherwise). Many times this is how one's

sense of understanding another arises phenomenally; but this is not always how it is. Just

---

[100] That is, I allow for radical otherness—even to the extent that others' agency may not be recognizable as agency at all from my perspective. Of course, this is a merely academic problem—because the epistemic nature of otherness is such that precisely these kinds of problems cannot possibly arise; from my perspective it would be impossible to experience an others' agency. This will be elaborated in chapter 4.

as often there appears to be no judgment at all—phenomenally speaking our understanding of the other shows up as *directly accessed*. For instance, when I look at the face of my perplexed pub friend when my response to his query is to hail the waiter, I don't *interpret* or *judge* him to be perplexed; rather I *see it directly* in his expression[101]. For all that it could possibly matter to me—that is, for all that I could know, or could be or become meaningful to me here and now—my pub friend simply *is perplexed.* The general point that I am making is that for my friend to be perplexed is necessarily a fact *from here*—it is indexed to and only as broadly scoping as the system for which it is significant or meaningful. This is what it means to *be* perplexed—to arise as being in or expressing a state of perplexity for some affection-recognizing system. There is no objective fact of the matter as to what it could mean to 'be perplexed' beyond this. To expect more is to require or reach beyond epistemic possibility; which means reaching beyond meaningfulness[102].

Given that the basic issue addressed here is one of the metaphysics of agency, the possibility of communication is of fundamental significance. Are agents the kinds of things that are capable of communicating? To this, it is of course crucial to understand what is meant by 'communication', and from the above sketch, it may be clear to some that what I have in mind is strikingly different than a theory that understands communication as the transmission of information from one subjective locus in objective

---

[101] Recognize the influence of Wittgenstein here (cf. Footnote 64). Following the work of Eleanor Rosch, and subsequent research in mirror neurons (e.g. Rizzolatti, Iacoboni, Gallese, et. al.) there is good reason to take the hypothesis of direct access to others' affective states seriously. This can be noted in the recent surge of philosophical and psychological work on what is being called "empathy" or "empathy studies". Shaun Gallagher & Dan Zahavi have even begun a sort of cottage industry in synthesizing all of these threads, as well as historical studies in phenomenology (especially in Husserl, Scheler, and Merleau-Ponty) into a coherent theory of embodiment sensitive to sociality.

[102] Which is precisely what many flavors of metaphysical realism do.

space to another. I believe that within such an understanding of communication then

agents are not the kinds of things that are capable of communication. However, on a

'weaker' version of what it means to communicate, agents *are* capable of

communication. I will construct a notion of communication on which nothing needs to be

shared between or amongst agents said to be in communication with each other. This is

fortuitous, because metaphysically speaking, I claim that sharing meaning between or

amongst agents is impossible. Most basically, agents cannot share or elude their

particular cultural, social, physical, and biographical histories—they are inextricably tied

to their own perspectives. But as *historical* entities, agents are *dynamically* co-constituted

by, and integrated with and within their dynamic (social, cultural, physical, etc.)

environments. Agents do not *share* but they are not isolated—they are open, dynamically

interact with, and are sensitive to the world around them. Further, as embodied and

extended entities, agency may span parts of bodies, one body, many bodies, tools,

cultures, and social practices. I call my theory of agency "Dynamic Embodied Agency"

or DEA.

DEA can be variously understood as a theory of metaphysics, epistemology, or

meaning. Taking phenomenal experience as primitive means that aspects appearing in

any one of each of these three areas always fundamentally impacts the other two[103]. As a

metaphysical theory DEA addresses what kinds of entities agents are. According to DEA,

In order to understand what an agent *is,* means understanding what agents *do*. Agents

understand and act in their local environments in accordance with the constraints and

---

[103] This is true of any theory of metaphysics, epistemology, or meaning. Any and all commitments made in one area constrain the possibilities in either of the other two.

affordances which function to disclose or occlude available possibilities for understanding and action. A situation is the disclosure or occlusion of available possibilities as they arises or "show up" meaningfully for an agent. Situations are essentially dynamic in that agents and their local environments are in close causal and constitutive coupling—they each differentially and continuously affect each other, and in so doing make each other what they are.

Taken as an epistemological theory, DEA addresses the ways in which agent-environment couplings come to constitute a dynamic situation *for* a cognitive agent. Epistemologically significant are then issues pertaining to how information is communicated to and for an agent, given the kinds of interactions that make the agent and local environment what they are. Specifically we will be interested in how agents communicate interpersonally, and the relations between interpersonal communication, an agent's understanding of her situation, and the development and effects of social practices as they become significant for a socially situated dynamically embodied agent.

As a theory of meaning, DEA focuses on the relations between bodies and meaning-bearing situations. According to DEA the body is the *sine qua non* through which environmental and agential affordances and constraints come to signify a meaningful situation for an agent. A dynamic agent's degree of bodily 'likenesses' and "attunement" to salient aspects of her situation are what disclose a sense of belonging and agreement with others, and feeling of being "at-home" or comfortable in her situation. When an agent's bodily encounter with her cultural, social, physical, or biographical environment arises as more disjointed, the agent comes to see her place in the world as

59

alienated and ill-at-ease. Of course, because situations are dynamic there are different

degrees and qualitative characters of experience that can change (with varying degrees of

robustness or plasticity depending on how "stable" the agent's understanding of her

situation is) depending most fundamentally on particular characteristics of the body-

environment interaction.

In the next chapter I will continue in more depth discussing issues of

communication, and social practices—specifically addressing concerns and criticism

relating to issues of transmission and inheritance of social practices, sharing social

practices, and the ontological status and significance of social practices for socially

embedded agents. Once this groundwork is laid, we will be in a position to construct

more fully a theory of agency that allows us to adequately navigate both these issues as

well as those encountered in the discussion about cognition in this chapter.

Chapter 2: Social Practices and Agency

2.1   Metaphysical and Epistemological Background

With the basic structure of the account of agency I've just formulated in place, it

is now important to address the pressing concern of the problem of other minds. My

intention here is to give an explanation of the possibility of dynamic embodied agents

communicating with each other. I do not propose to solve the problem of global

skepticism about other minds (the solipsistic doubt that there *are* other minds out there).

Instead, I take it as granted that there exist a plurality of agents acting in the world.

Whether this is metaphysically the case or not, I will argue, is phenomenally and

epistemically irrelevant—and taking phenomenality as our only metaphysical primitive

(as I've attempted to justify in the previous chapter) any metaphysical claims that make

no difference phenomenally cannot *matter* to the kinds of dynamic embodied agents I

have outlined[104]. So as long as the world shows up to an agent through her experiencing

as populated with other agents, this itself is enough to warrant the metaphysical

hypothesis of other agents; as no stronger criterion for warrant is possible.

There are many and variegated senses in which one might call him or herself a

---

[104] In some ways, my thesis here is Quinean in the sense of preferring ontological "desert landscapes"—I
use epistemic significance as a constraint for metaphysical significance; and use metaphysical
significance to constrain metaphysical possibility. I am aware that these constraints underdetermine a
final metaphysical resolution, and in fact this basic metaphysical indeterminacy is a natural result of the
fundamental differences in perspectives explicit in my account of dynamical embodied agency. The
sense in which this is "Quinean" is that epistemic meaningfulness can be used as a mode of paring back
the metaphysical landscape. Tersely, my account privileges epistemic difference over metaphysical
(realist) similarity.

metaphysical realist. Here is not the proper place to rehearse such metaphysical

taxonomies, however some brief remarks describing and locating my position are

warranted, because metaphysical commitments have epistemic consequences and *vice

versa*—such commitments are mutually constraining. Thus, the picture of dynamic

embodied agency outlined in the previous chapter is only coherent on or possible for a

definite set of metaphysical commitments. For example, the requirement that agents are

necessarily *embodied* excludes the possibility of a purely idealist metaphysics (or

conversely for an idealist, dynamical embodied agency should appear as incoherent on

the grounds of the materiality of the body—lived (*Leibt*) as it may be). The requirement

that agents can *act* excludes the possibility of a purely static universe[105] (e.g. Platonism,

Four-Dimensionalism, Russell's so-called "at-at" theory of time, etc.); instead it relies on

a nonlinear, intensive notion of temporality[106].

     None of this is to say that—while non-standard—there is anything *inconsistent*

between dynamic embodied agency and a fairly flat-footed metaphysical realism – the

belief that reality, or aspects of reality, is ontologically independent of our conceptual

schemes, linguistic practices, beliefs, etc.. Certainly the tone of my position, indexing

metaphysical possibility to epistemic possibility, is quite *anti-realist*. But nothing about

the metaphysics of agency given in DEA precludes a thoroughgoing scientific realism[107].

---

[105] This is because action depends on the passage of time and the possibility of change.

[106] For excellent discussion of the qualitative differences in temporality see Henri Bergson's *Time and Free Will*, Division II of Martin Heidegger's *Being and Time*, and Theodore Schatzki's *The Timespace of Human Activity*. While there are interesting differences in each of these accounts, all of them share an affinity for what Bergson calls aionic duration—an intensively felt and overlapping vectoral time that problematizes the sequential past-present-future structure assumed by many theories of metaphysics. More thorough treatment of the concept of nonlinear temporality is given in chapter 4.

[107] At its most basic, scientific realism is the position that the unobservable entities posited by current best scientific theories are metaphysically real. One popular scientific realist position is the "Structural

Prior to taking a stand as a realist we may ask what is to qualify as "real". DEA is a

position that takes a metaphysical stand by first taking an epistemological stance.

But the problem of other minds goes further than mere metaphysical

commitments to realism. It ties together the metaphysical and epistemic issues of the

possibility knowing certain kinds of entities—namely *other minds*. In order to know

anything about other minds, there must be something about which it is knowledge—there

must *be* other minds. But what is the character of these minds? What are these minds

like? And what is the character of the epistemic activity required to know them? Both of

these concerns depend directly on the quality of epistemic encounter brought to bear by

agents. In a sense, the perspectivalism implicit in dynamic embodied agency invites a sort

of relativism: different agents are going to understand the very character of agency—of

what it *means* to be an agent—differently. That is—from either of our perspectives, we

may be *taking* each other *as* agents, but *what* either of us *means* by "agent" may be very

different. Of course, this possibility of difference is simultaneously both an epistemic and

metaphysical possibility. Epistemically, as perspectival agents—agents who perceive

from somewhere—dynamic embodied agents lack the requisite access to be able to learn

of such a difference. Thus, *if* there is a difference between my and your understanding of

what an agent is, it would be *in principle* impossible for me as a perspectived agent to

---

Realism" pioneered by John Worrall, which suggests that while the content of scientific theories may
not strictly be true, our best scientific theories tend to get better and better at capturing the structure of
reality—science accurately describes the real structure of the world, and gets progressively more
accurate (without major crises or revolutions in the way that Kuhn, Lakatos, and Feyerabend describe).
Dynamical embodied agency is consistent with structural realism, and scientific realism more generally
because there is nothing in it that directly or indirectly disputes an adequate interpretation of our current
best scientific theories.

know it[108]. To know the difference requires that one dynamic embodied agent has *direct access* to another dynamic embodied agency *qua agency*; the agency with that particular dynamic embodied perspective. Essentially such an agent would have to *be* the other— but under such circumstances, there would simply be no difference anymore, and no 'other' to speak of[109].

The consequences of this picture for an adequate theory of communication are quite significant. As long as an agent can have *direct* or *immediate* epistemic access to another, the other is no longer other. In recent literature on the topic of empathy, some commentators[110] have focused on the question of what constitutes 'directness' as it pertains to epistemic access. I find that it is most useful to see directness and indirectness as *differences in degree* of accessibility. The easier it is for an agent to incorporate something into her dynamic embodied agency the more direct it is[111]. This approach has fairly clear consequences: On the one hand, the directness of access is directly proportional to the degree of otherness. On the other hand, this means that the greater degree of otherness, the less epistemic access a dynamic embodied agent has. These results point to the significant ways in which dynamic embodied agents are constrained in

---

[108] Because of the impossibility of knowing the content of another's mind from inside, I will not make much of the point that emphasizes difference here. We have reason to suspect different perspectives produce differences in experience and vice versa because experiences define and shape our perspectives (and vice versa). It seems a safe (though inductive) inference to conclude that from a difference in agents' perspectives there should be a corresponding difference in agents' understandings and meanings. This sets the stage for discussion of the problem of sharing social practices later in this chapter.

[109] Cf. e.g. Merleau-Ponty's discussion of the problem of other minds in the preface to Phenomenology of Perception.

[110] For example, in his address at the 2012 Society for Phenomenology and Existential Philosophy (SPEP) titled Levels of Empathy, Dan Zahavi argues that "one should acknowledge that there simply isn't any established view on what "direct" means."

[111] The specifics of the ontology and mechanisms involved in "incorporating" something into one's dynamic embodied agency will be discussed more exhaustively in chapters 3 and 4.

their abilities to relate to each other—which, in turn, directly impacts the possibility for success in interpersonal communication between them.

The complexity of this issue is compounded by increasing the number of 'moving parts' (in our case, increasing the number of dynamic embodied agents, as well as number and varieties of parameters and variables to which DEAs are sensitive) involved in the epistemic task. Phenomenally speaking, there is a qualitative difference between how one encounters one concrete other (a second-personal "you") and how one encounters aggregate, or indistinct other*s* (first and third personal plural "we", "they"[112]). From whence does this qualitative difference arise? Is it a *result* of the quantitative difference (viz. is it an *emergent* phenomenon?), or is it a difference in phenomenal *kind*? This chapter attempts to sort through these issues.

More centrally, however, this chapter focuses on the character of *social* encounters, experiences, and communication—what it means to *be social*, or to engage in *social practices*. Many theorists have, at various times, in various places, and in various ways, invoked or alluded to social practices as *explanations* for other apparent phenomena[113]. From the perspective of this investigation of agency, it is of interest to try to understand exactly how the kinds of agents described in the previous chapter can be thought of as *social* agents. What could it mean to be a social agent, or to engage in a social practice as a dynamic embodied agent? In order to ask or answer these questions we must first navigate the different senses of sociality; as well as develop an

---

[112] This difference was played out vocally in arguments between Heidegger and Sartre in the 1930s and '40s. See later discussion on this point.

[113] Stephen P. Turner emphasizes three areas where he sees this as especially true: attempts to explain the conduct scientific communities, attempts to reductively explain political theories, and attempts to naturalistically explain morality (c.f. Brains/Practices/Relativism, 2002).

understanding of what exactly social practices and institutions are, or are meant to explain.

2.2   *Issues of Selfhood and Otherness*

The problem of other minds has traditionally been discussed in terms of, on the one hand, Cognitivist accounts of mind that can be roughly traced through a tradition following the *representationalist* picture of cognition first expressed by Descartes and later elaborated by Kant in the First *Critique*; and on the other hand, predominantly Empiricist accounts of learning that find their beginnings in Locke's and Hume's *Inquiries*. Despite its seeming initial lack of an account of cognition, the empirical approach enjoyed a renaissance following the eclipse of Kantianism. The idea behind the empiricist approach was that through the tools afforded by the highly empirical scientific method, inquiry could be conducted in such a manner that the advancement of understanding was inevitable. Compared with the rather speculative transcendental reasoning championed by Kant and the neo-Kantians, 19th century empiricism garnered much wider appeal. It is somewhat ironic then that empiricists were motivated more than anything else by *faith* in this inevitability in epistemic advancement. If the problem of other minds was to be solved, it would be solved by as-yet undiscovered *empirical* investigation—not speculative transcendental inferences to conditions for the possibility of knowledge.

The Cognitivist approach re-emerged in the early decades of the 20th century, when advancements in Boolean logic and computing began to make headway in classical

problem solving[114]. Thus the "classical" *computational* or cognitivist approach to modeling the mind was born. This picture took the mind to *function* like a Turing machine—processing the input delivered from the raw sense-data acquired from the world through bodily sense-perception. This data filtered through the 'programs' of cognition as *physical symbols*, or tokens that could be operated on by the inferential system of the human mind[115].

The computationalist approach to modeling the human mind could rely on the representationalist assumption of different kinds of *information* being transmitted through the media of external environment, sensory apparatus, and as symbolic tokens within the brain. What needed to be discovered was the ways in which that information was manipulated, translated, and converted through these media in such a way as to produce veridical internal representations of the outside world. The more-or-less unquestioned epistemic assumption of computationalist representationalism was that the contents of thought, symbolic representations, veridically *corresponded* to the way at least some of the world actually is[116]. This is significant to the problem of other minds because it made it easy to see how communication was possible, and suggests a fairly straightforward explanation of how communication works. If human minds are organized such that they operate on and produce representations that bear direct correspondence with the objective world, then that world can function as a common medium through which representations

---

[114] Cf. e.g. Alan Turing's 1950 "Computing Machinery and Intelligence", *Mind* LIX (236): 433–460
[115] Cf. Newell, Allen; Simon, H. A. (1976), "Computer Science as Empirical Inquiry: Symbols and Search", *Communications of the ACM*, 19
[116] Here I qualify the statement with 'at least some of' because nobody claimed that human sensory apparatus was capable of picking up on, or receiving all of the information embedded in the world. The processes involved with sensation are, for the representationalist, most fundamentally filtering processes; separating out the 'wheat', the relevant symbolically manipulable content of the world, from the non-conceptual 'chaffe'.

can be expressed, transmitted, and received between subjects.

In essence, the process can be understood as a simple encoding and decoding. Imagine that I wish to express a certain idea I have to my friend Fred sitting across from me. Perhaps I wish to say "Kree Jaffa!". The computationalist says that my thinking of this idea, and what I would like to say to Fred are just the operation of an interpretive processor on representative or symbolic tokens. When it comes time to produce some output, the appropriate symbolic token is sent to a sensorimotor-translator which takes the representative token and translates it into a signal that is instructionally significant for the appropriate lower-level sensorimotor systems (larynx, diaphragm, tongue, lips, mouth). In executing the received instructions, these lower-level systems *encode* the representational content I wish to express in a new medium, sound waves traveling from my mouth and radiating through the air, presumably in the direction (among others) of Fred's ears. Fred's ears pick up the audio signal and send it, still encoded, rather circuitously to the inner ear where air vibrations are translated to aqueous vibrations in the Organ of Corti, which in turn stimulate the spiral ganglion, and finally transmits an electrochemical signal down the Vestibulocochlear nerve. Each translational step does some *decoding*[117] of the information—but not until it reaches the Pons, is the sensory information fully decoded into a symbolic representation for Fred. At this point, as long as the various processes of translation did not result in loss of the relevant *representational* information[118], Fred has mentally available the symbolic representation

---

[117] And re-encoding; essentially *translating* (or schematizing) information from one format into another. Certainly not *every* aspect of the previous signal can be translated; something is always lost in translation (e.g. the mechanical character of vibrating air is not retained in the electrochemical signal).

[118] Again, the claim is not that *every* aspect of the external stimulus is represented veridically; just that *at*

"Kree Jaffa!".

So for the computationalist, communication is the largely straightforward, linear process of transmitting readily encodable and decodable information through intermediate media. Since the symbols necessary for computationalist accounts of cognition must be *discrete* it means that communication between two cognitive agents is a digital, all-or-none, affair; either the signal is transmitted (and thus a thought is communicated from one subject to another, or it is distorted, and what is received is not the same as what was expressed. Sufficient noise or loss of information means entire loss of signal, or complete communication breakdown. But, since this is clearly not the way our usual, mundane interpersonal interactions work, there must be something wrong with this picture. Fodor, for one, has suggested that problems crop-up whenever one supposes that an agent's mental processes or states depend in any way on states external to the agent. In order to understand an agent's mental processes or states, all one needs to do is investigate the *internal* workings of the [functional] parts that make up that agent. Fodor grants that external information is causally efficacious, but cannot give semantic content to the agent's mental workings[119]. Presumably for Fodor, communication in the sense of transmitting information from one agent to another is an epistemic impossibility. This has the effect of explaining *away* the problem of communication, and of other minds; rather than addressing it. What is needed instead is an account of communication that explains how it can possibly work without encountering the pitfalls of the mediated and discrete transfer of symbolic information that we saw earlier.

---

least some of what is processed bears some veridical relation to the external stimulus.
[119] Cf. Fodor, Jerry (1980), "Methodological Solipsism Considered as a Research Strategy in Cognitive Science," *Behavioral and Brain Sciences*, 3: 63-73.

By the late 1980s, basic questions about the representational computationalist

approach had not yet been answered (especially regarding how such a system could have

evolved by natural selection[120]), and some researchers returned to a more empiricist

approach. Instead of attempting to simulate the functional aspects of a Turing machine

with a serial von Neumann architecture, the new "connectionist" approach treated the

neuronal structure of the human *brain* as the model's basis. While this approach initially

showed great promise, it quickly ran into some severe setbacks. One problem arose in

methods of simulating the *brain*—what level of abstraction is appropriate when

simulating a complex biological system such as the human brain comprised of over 100

billion neurons? How should or could physical constraints be simulated? Another

problem was in simulating the *mind*—is the non-representational Hebbian learning in a

connectionist network at all analogous to the seemingly representational appearance of

conscious thought[121]?

Further, while the connectionist approach to cognition made strides in conceiving

the mind as more akin to a massively parallel processor, it still seemed as though the

picture for communication looked much like the computationalist one just described. As

much as the connectionist paradigm increased the bandwidth or channels of information

flow for associative connections *inside* the head, it was still constrained for

communication to the comparatively *low* bandwidth sensorimotor channels. Moreover,

---

[120] E.g. Cosmides, L. &Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality*. Cambridge, MA: The MIT Press.
[121] Cf. e.g. Haugeland, John, *Artificial Intelligence: The Very Idea* (1985). Cambridge, Mass.: MIT Press; and Fodor and Pylyshyn. "Connectionism and Cognitive Architecture: A Critical Analysis" in S. Pinker and J. Mehler, eds., *Connections and Symbols* (1988) Cambridge, Massachusetts: MIT Press

connectionism still trafficked in discrete tokens which bore little resemblance to the

actual continuous differential processes with which a realistic cognitive system interacts.

In short, connectionism changed little if anything about the problematic picture of

communication.

　　While advancements were being made in neurochemistry, neurobiology, and

cognitive psychology; philosophical accounts seemed incapable of producing results—we

still lacked a plausible initial approximation of how the mind works so as to produce

something as rich and complex as *consciousness*. We still couldn't explain how humans

communicate. Then, in 1991, the first[122] Embodied accounts of cognition began to

precipitate into cognitive science discourses[123]. As discussed in the previous chapter, this

approach to cognition treated the living human body—not only the brain—as the locus of

perception and understanding. While both Computational and Connectionist theories

subscribed (minimally) to a methodological individualism and thereby limited their

investigations to the brain, the embodiment approach broadened the focus to include not

only the flesh and bone encapsulating the brain, but also the various physical, social,

cultural, and historical environments encountered by the living body as the appropriate

loci for investigating the human mind. Since some of these forces are external to the

internal workings of the brain, the methodological individualism of Computationalist and

Connectionist theories could no longer be maintained so comfortably. While this may

have posed epistemic boundary problems for computationalists and connectionists, when

---

[122] The outcropping of Embodied accounts in the nineties was by no means the first instance of such ideas—these thoughts date back to the very beginning of philosophy of mind, and can be seen in Aristotle, Descartes, Husserl, Merleau-Ponty, Heidegger, and Sartre just to cite a few.

[123] E.g Varela, Thompson, Rosch, The Embodied Mind: Cognitive Science and the Human Experience. MIT Press. 1991

we look at the significance of this change for the study of interpersonal communication we can see that new avenues of possibility have opened up.

Varela, Thompson, and Rosch set the agenda for the Embodiment program in 1991 with their now-renowned *The Embodied Mind: Cognitive Science and the Human Experience*. They wrote that embodied cognition "depends upon the kinds of experience that come from having a body with various sensorimotor capacities, [that] are themselves embedded in a more encompassing biological, psychological, and cultural context"[124] and cite as inspiration for this thesis the work of Hubert Dreyfus, George Lakoff, and Mark Johnson. They closely tie the notions of the embodiment and embeddedness of cognitive agents with the bodily activities that are brought about, calling this tight connection between body and world 'enaction'. Enaction, according to Varela, Thompson, and Rosch consists of two primary aspects: "(1) perception consists in perceptually guided action, and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided"[125]. Notice here that there is here no mention of the *constitutive parts* of enactive systems of agency. All that is discussed is the act of perception (not the equipment through which perception occurs), actions and the relationship between perception and action for such agents, and emergent patterns of perceptually guided action. As discussed earlier, in order to *act* one must have or be a body—but nowhere in this account do Varela, Thompson, and Rosch say how that body needs to be put together[126]. The enactive agent is one who dynamically reciprocally co-

---

[124] Varela, Thompson, Rosch. *The Embodied Mind: Cognitive Science and the Human Experience*. MIT. 1991. p. 173
[125] Ibid.
[126] This may be somewhat misleading. It is clear that the authors believe that the particulars about bodies

constitutes her world—her activities shape the world around her, and the world around

her shapes the activities she embodies.

When we add to enactivism the "extended mind" thesis, objects from the

environment may then be taken up by an enactive agent into her cognitive schema as a

part of its—and her—cognitive *constitution*. Shaun Gallagher's distinction between *body-

schema* and *body-image* proves useful in discussing these matters[127]. It is important to

recognize early on that Gallagher's argument employs a *formal* distinction between the

concepts of a "body-schema" and that of a "body-image", and uses this distinction to

clarify the problem of intersubjectivity. It will be important to track and maintain this

distinction as formal or conceptual, rather than one that is intended as a metaphysical or

ontological commitment. He writes:

> Distinctions that can be made clear conceptually may not remain so at the
> level of practical behavior...I will argue that if the clear a proper distinction is
> made, these concepts carve up the conceptual space in such a way that leads to
> a productive understanding...rather than trying to locate or construct an
> intermediary entity to bridge the [self-other] gap, the task, as I understand it, is
> to create a coherent and contextually rich background theory.[128]

This means that while Gallagher's distinction is meant to help clarify the ways we talk

about intersubjectivity—borrowing from Plato, to "carve nature at its joints"[129]—it is not

---

and their environments are constitutively significant in the perceptual activities that such agents are capable of performing. What is important here is the recognition that they're not talking specifically about human bodies or minds.

[127] It is worth noting that Gallagher's use of this distinction between "body image" and "body schema" derives from Merleau-Ponty's usage of the terms (*l'image de notre corps*, and *schéma corporel* respectively) in *Phenomenology of Perception*. To complicate matters, the canonical Colin Smith translation of *Phenomenology of Perception* does not distinguish between these two; referring to both as "body image". This mistake has been corrected in the recent Donald Landes translation. For his part, Merleau-Ponty saw himself as employing a then-common terminological distinction from contemporary French research in clinical psychology.

[128] Shaun Gallagher (2005) How the Body Shapes the Mind (pp. 5-6).

[129] Plato's Phaedrus 265e.

meant to endorse specific theories of mind or social cognition as metaphysically correct. Nevertheless, throughout his 2005 book *How the Body Shapes the Mind*, Gallagher explains why he thinks the most consistent, and fruitful theory of mind is an "embodied" and "enactive" one, and why theories of social cognition must adequately and co-constitutively align with embodied selves. For this reason, it can be easy to misinterpret this distinction as the basis for an argument for a theory of embodied cognition.

On Shaun Gallagher's construal, the difference between the enactive agent's body and the world is a matter of inclusion in her body-schema. If objects arise in one's body-schema they are pre-cognitively[130] (and literally) incorporated into her enactive body. What this means is that the difference between the enactive agent and her world is not founded on any principled *physical* distinction but rather is a matter of the ways in which her situation arises *phenomenally*. The enactive agent phenomenally encounters her body differently than other items in the world—and when that encounter includes (in the instance of a blind agent) her cane, then this too 'is' her body and a part of who she is[131]. This raises another issue—this time with the distinction between subject and object. If this account is correct, then there is a blurring of this distinction—an "object" taken up and incorporated into an embodied subject's body schema; which becomes part of that

---

[130] There are many phrases in the phenomenological literature that, for our purposes, can be thought of synonymous. Among them are, on the one hand, "pre-thematic", "pre-reflective", "pre-cognitive", "non-representational", "pre-noetic", "lower level thought", "perception" (excluding cognition), and a host of others, and on the other hand "thematic", "reflective", "cognitive", "representational", "noetic", and "higher order thought" among others.

[131] Implicit in this conception of enactive embodiment is a certain kind of individualist notion of ownership of one's body. Phenomenally speaking, the blind person's cane is encountered as part of her agency because it factors in making her who she is—in a sense, it is a part of her, and this is the sense in which she 'owns' it. This phenomenal sense of ownership is hugely significant in whether or not some element in the environment is, or is not, to count as a proper constitutive part of the agent. Mark Rowlands even goes so far as to identify this kind of 'ownership' as the fundamental criterion for what Adams and Aizawa have called "the mark of the cognitive".

subject, is in some important senses no longer *merely* an object but has become *subjective* in the sense that it incorporated into the constitution of the subject—it is included in embodied agent's subjectivity and it participates in constraining and affording the kinds of relations in which that agent can engage.

This sets the stage for one of the more difficult and puzzling aspects of the embodied and enactive account of cognition as conceived by Gallagher: the problem of other minds. The enactive agent perceives her body not merely as a physically extended bit of matter (*Körper, le corps objectif*), but as lived (*Leib, le corps vécu*)[132]. This reverses the traditional Cartesian distinction between mind and body. While Descartes distinguished between two substances (*res cogitans, res extensa*) the phenomenologically experienced, lived body of Husserl and Mearleau-Ponty, for example, is one that is perceived and experienced *already as lived.* Only later, through meditation or reflective thought can such a distinction between two kinds of substances be made. Thus the Cartesian division of mind and body is secondary to or derivative of a more fundamental division between a "lived body" and that which is "other"; though as we have already seen, even this differentiation must be thought of as not rigidly defined, but rather fluid and contingent upon the concrete dynamic interactions between an agent and her environment.

Because this distinction is not rigid, what "counts as" other in each situated

---

[132] The *Leib/Körper* distinction is Husserl's, *le corps vécu/le corps objectif*, Merleau-Ponty's. In *Phenomenology of Perception*, Merleau-Ponty employs the phrase "*le corps propre*" (one's own body) rather than *le corps vécu*, though in the context of his work it is fairly clear that Merleau-Ponty intends for the two phrases to mean the same thing: in order for a body to be "one's own" it must be an agent's living body. Once again Colin Smith takes some liberties in translating "*le corps proper*" as "lived body", but in this case the consequences are less pernicious.

experience of the embodied agent needn't be uniform in its otherness—what is other is experienced intensively[133] as other. For this reason, other agents can arise for the embodied subject as true others; that is, others *qua subjective agents* rather than as mere objects of experience[134]. An often ignored consequence of this maneuver is that agency too must be thought of intensively. As we will see, the degree of likeness in subjectivity is directly related to the phenomenal degree of perceived likeness in bodies[135]. The ability to pre-reflectively compare bodies is a condition for the possibility of intersubjectivity and empathy.

Embodied agents do not merely *have* bodies, but in a very literal sense they *are* their bodies. The skills, habits, and practices acquired through one's lived-out experiences are skills, habits, and practices that find their proper intelligibility in the situated context of one's embodied and embedded life. They are first and foremost skills, habits, and practices of a body engaged in the processes of living. This is not to deny that they, in some sense, originate from the "external" environment. While such skills, habits, and practices are learned and normatively honed by the embodied agent herself; embodied agents are always situated in physical, social, cultural, and historical contexts. Precisely *what* such skills, habits, and practices *are* is dynamically developed through the particular concrete "living out" of a dynamic embodied agent's life. At a certain point in an agent's life, such skills, habits, and practices may *mean* something quite different than

---

[133] That is, it is experienced differentially—as a qualitative difference of degree.

[134] Further, aspects or elements previously incorporated into an agent's embodied subjectivity can also become other; one can become alienated from what one takes to be one's self, one can misrecognize one's self. More on this in chapter 4.

[135] This perceived likeness needn't occur at the "noetic" level, or the level of conscious awareness. Using Gallagher's term of art, we should understand perceived likenesses at the "pre-noetic" or "body-schematic" level (*in addition to* the "noetic" or the level of "body-image").

at another point for the same agent[136]. Additionally, we should keep in mind that on the

enactive account there can be no principled, static, or rigid differentiation between self

and other—such differences are a matter of degree in modes of access and bodily

similarity. Thus lived bodies, like merely physical bodies, are able to be encountered as

situated or embedded in the world—they are able to be experienced by others *as subjects*.

In this way, to exist as a dynamic embodied agent means to exist *as perceptibly*

accessible by another, and to be able to perceptually access others[137]. Moreover, this

means that others can be directly perceived in their sensory, emotional, and even

cognitive expressive subjectivity. We will presently investigate how this is a radical

departure from the traditional view.

The first noteworthy aspect of the picture of embodiment outlined above is that

'self', 'other', and 'world' co-constitute each other. There can be no account of self

without making essential reference to and including others and the world. Likewise, any

account of the world must include the fact of its inhabitation by subjects—a world is a

world *for someone*. Further, to be a self means to be recognizable as a self by another

self; and to recognize that recognition. Without this double association between self and

other, there can be no sense of self. To think of others as somehow secondary to an

understanding of one's self, is to uphold and maintain the Cartesian distinction between

---

[136] Some may think that because of this difference the 'sameness' of the agent is thrown into question. Without continuity in the skills, habits, and practices of an agent, identity cannot be maintained. Perhaps not. But I am not claiming that there is no continuity—in fact I do claim there is continuity; but it is an intensive, temporally nonlinear, dynamic continuity—changes unfolding on variant, variable, overlapping timescales. The difficulty for those concerned with the seeming discontinuity presented by the differences in an agent's phenomenal perception of his or her skills, habits, and practices at one or another time in his or her life is caused by failing to understand properly the temporality of a life lived. Only when one mistakenly takes a Platonist, Four-Dimensionalist, or Russell's "at-at" theoretical stance on time, can such a problem arise.

[137] Cf. e.g. the preface of Merleau-Ponty's *Phenomenology of Perception*.

mind and body—it is to take the first-personally experienced realm of "mind" to be more intimately known, and more ontologically primordial than bodies, both of one's own, and of others. The troubles for the Cartesian dualist are, then, double: (1) first they must account for, or bridge, the substantial gap between *res cogitans* and *res extensa;* (2) second, they must bridge the gap from their own first-personally experienced bodies to the bodies of others such that there remains an *analogy* which allows for the *inference* from the other's body to the other's mind.

The foregoing point can be traced back (at least) to John Stewart Mill and has been traditionally called the "argument from analogy". It can be characterized as follows:

> The only mind I have direct access to is my own, My access to the mind of another is always mediated by his bodily behavior. But how can the perception of another person's body provide me with information about his mind? Starting from my own mind and linking it to the way in which my body is given to me, I then pass to the other's body and by noticing the analogy that exists between this body and my own body, I *infer* that the foreign body is probably also linked in a similar manner to a foreign mind.[138]

As with any argument by analogy, it should be taken as a defeasible argument by inference to the best explanation. The presumption is that in recognizing a fundamental substantial difference between our own mind and body, we must give an explanation for the existence of others' inaccessible minds[139]. This explanation runs by the analogy we can construct on the likenesses assumed between my body and the body of the other. Since my agency has a mind-body structure, and the other has a body like mine, she too should have a mind-body structure just like mine. Alternatively, phenomenologist Max

---

[138] Zahavi, Dan. "Beyond Empathy" Phenomenological Approaches to Intersubjectivity", *Journal of Consciousness Studies*, 8, No. 5-7, 2001, pg 151.

[139] Wittgenstein problematizes this way of thinking in his classic discussions in questioning the pain of others (*Philosophical Investigations* §300-304).

Scheler has formulated arguments to condemn the argument from analogy:

1. To assume that our belief in the existence of other minds is inferential in nature is to opt for a far too complex cognitive account. After all, both animals and infants seem to share this belief, but in their case it is hardly the result of a process of inference.
2. In order for the argument to work, there has to be a similarity between the way in which my own body is given to me and the way in which the body of the other is given to me. But my own body as it is felt proprioceptively for me does not at all resemble the other's body as it is perceived visually by me.
3. How can the argument from analogy explain that we can empathize with creatures whose bodies in no way resemble our own, say a suffering bird or fish?
4. Even if all of these problems could be overcome, the argument from analogy would still be formally invalid. Noticing the connection between my own mind and my bodily behavior, and the analogy between my own bodily behavior and the behavior of a foreign body, all that I am entitled to infer is that the foreign body is probably also linked with *my own* mind.[140]

We have already seen some responses, or solutions to some of the problems raised by Scheler. Since the argument from analogy relies on Cartesian mind-body dualism, a sufficient alternative to this Cartesian picture will likely relate differently to these concerns. Since, on the dynamic embodied account of agency I am my lived body, the fourth consideration does not even arise. In our acknowledgements that similarity among bodies is in *all* cases a matter of degree[141], and that lived bodies are alike in kind, the third consideration is also circumvented—or at least pushed back into a dependence on whatever cognitive account of pattern recognition we can give. Further, as we will see, there is evidence to suggest that the second consideration is hasty in its formulation.

---

[140] This is Scheler's argument as reformulated by Dan Zahavi in "Beyond Empathy" Phenomenological Approaches to Intersubjectivity", *Journal of Consciousness Studies*, 8, No. 5-7, 2001, pg 152.
[141] I hasten to add here that such bodily similarities are properties of actual, concrete, relations encountered in real time. Lived bodies and the contexts in which they express themselves are dynamic and unfold on multiple overlapping timescales. As a result, the similarities and differences in bodies are contingent upon the particular conditions of specific intersubjective situations.

Research in mirror neurons, and shared representations indicates that the actions of others *are* instantiated in one's cognitive processes in the same way as one's own actions. Further, there is reason to think that—though proprioception and vision are distinct modes of perception—there is cross- or inter-communication between modalities[142], allowing for the possibility that my body as it is felt proprioceptively for me *does* resemble the other's body as it is perceived visually by me. Similarly, there have been extensive advances in experimental psychology (spurred in large part by Meltzoff's and Moore's groundbreaking 1977 study on imitation in neonates) to suggest that we should give a more direct and innate perceptual account than one that requires the cognition-intensive process of inference.

But we must also be careful not to step beyond the empirical evidence when it comes to direct perception of others' (emotional, sensual, intellectual, etc.) bodily expressiveness. One should not claim that what agents have access to are the processes occurring inside another agent's brain. Nor should one claim that direct perceptual access entails that misrecognizing one's own or others' (emotional, sensual, intellectual, etc.) expressions is impossible. After all, one simple counterexample—the familiar phenomenon of a "successful poker face"—precludes the defensibility of such a claim[143]. So what then could be meant by this notion of an agent "directly perceiving" another? Here there is no clear agreement between embodiment theorists. On one interpretation (shared by Dan Zahavi and Theodore Schatzki) the emphasis of the claim is on the

---

[142] Cf. especially Hurley & Chater (2005), Goldstein (2001), and Iacoboni (2009)
[143] A "poker face" expresses one emotional, sensual, intellectual, etc. state, without expressing what that agent actually feels, believes, thinks, etc. What is important to track here is that expressively lying is a genuine set of phenomena that should be acknowledged and accounted-for.

directness of the perception. Both Zahavi and Schatzki believe that the contiguity

between self and other stems from a shared commonly lived-in world in which both

intersubjective parties interactively participate. Approvingly citing Heidegger, Zahavi

writes:

> Dasein does not initially exist alone, and does not first acquire its being-with [*Mitsein*] the moment another turns up. On the contrary, qua its engaged being-in-the-world, Dasein is *essentially social* from the very start...Dasein's being-with, its fundamental social nature, is the formal condition of possibility for any concrete experience of and encounter with others...Dasein cannot be understood except as inhabiting a world which it necessarily *shares* with others...Under normal circumstances we understand each other well enough through our *shared engagement* in the *common world*, and it is only if this understanding for some reason breaks down that something like empathy becomes relevant. (*BE* 154-155, emphasis added)

What is important to Zahavi is that individuality is a later development derived from a

more primordial social being-with, which occurs through Dasein's essential 'thrownness'

in a ready-made world not of her own but rather a social making.

Following a remark Wittgenstein makes in *Zettel* §225, Schatzki similarly writes

that

> The ability to *perceive* (primarily see and hear) that someone is in such and such condition...pertains mostly to mental conditions manifested in behavior. Wittgenstein states emphatically that *one does not infer* from, say, fear or pain behavior that another person is fearful or in pain; one *sees* it:
> 'we *see* emotion.' - As opposed to what? - We do not see facial contortions and make inference from them (like a doctor framing a diagnosis) to joy, grief, boredom. We describe a face *immediately* as sad radiant, bored, even when we are unable to give any other description of the features. - Grief, one would like to say, is personified in the face. (Z, 225)[144]

Elsewhere Schatzki asserts that "a 'realist' viewpoint maintains that reality is *directly*

---

[144] Schatzki, Theodore. Social Practices: A Wittgensteinian Approach to Human Activity and the Social. p. 76

('bodily' as Husserl put it) encountered in perception"[145] and that "individuality is a socially constructed and achieved status. Personhood is an *effect* of social practices, in that expressive bodies, life conditions, and ascriptions/comprehension of these conditions exist only within practices"[146]. He also follows Heidegger in distinguishing between "having" and "being" a body: "to be a body, is to be able both to perform bodily doings and sayings and to experience bodily sensations and feelings... having a body is made evident in situations of breakdown, malfunction, discomfort, and incompetence, where the fact that one is a body manifests itself explicitly."[147] From all this we can glean that Zahavi and Schatzki share the position that embodied agency is a result, or effect of the socially-constituted practices that define the world; and from which direct perception of others is made possible. A more fundamental social existence is the transcendental condition for and the distal cause of the possibility of the phenomenal feeling of individuality. Transcendentally, our social embeddedness is the *sine qua non* for any self-conception, and thus a necessary requirement for the phenomenal sense of individual selfhood. Causally, it is through the myriad overlapping social practices, institutes and mechanisms of our experiences that causes or produces the effect of one's self-conception.

On the other hand Shaun Gallagher, in recognizing the non-inferential directness of perception of others emotional, sensual, and cognitive conditions, emphasizes the *perceptual* aspect. That is, if the hallmark of the similarities between Zahavi's and Schatzki's theories of embodied situatednesss is of sharing a common social, cultural,

---

[145] Ibid. p. 28
[146] Ibid. p. 35
[147] Ibid. p. 43

and historical world; then what differentiates Gallagher from this position is that Gallagher never emphasizes the element of *sharing*. In his *How the Body Shapes the Mind*, he only ever uses the words "share", "shared", or "sharing" in three contexts, none of which suggest the idea of two or more subjects grasping or comprehending or even perceiving one and the same object or situation. The first circumstance is in dismissing an incorrect viewpoint held by Kita. As they are not relevant to our present concerns, we will leave aside these instances of "sharing". The second situation in which Gallagher makes use of the concept of sharing is in referencing a term of art coined by Baron-Cohen called "shared attention", or what Gallagher himself calls "joint attention". In situations of shared attention [infant] subjects are said to "enter into contexts or shared situations in which they learn what things mean and what they are for"[148]. Following Trevarthen's distinction between primary and secondary intentions, Gallagher utilizes the phenomena of "joint attention" to exemplify secondary intentions—whose "defining feature is that an object or event can become a focus *between* people"[149]. Here it may seem that this is exactly the kind of situation where two or more subjects are grasping, comprehending, or perceiving one and the same object or situation—that is, this seems precisely to be an instance of social sharing. However, this is not the way that Gallagher intends it. Instead, his discussion is on how the infant perceives *her own* situation or object of attention *as shared*. Perceiving something *as shared*, and *actually sharing* are two entirely different concepts. The former is a claim about the phenomenal character of an agent's encounter, while the latter is a metaphysical claim about the world. Elsewhere

---

[148] Gallagher, Shaun. How The Body Shapes the Mind. p. 228.
[149] Ibid.

Gallagher writes that infants "9-11 mos. Are able to see bodily movement as expressive of emotion, and as goal-directed intentional movement ... [they are] however, not taking an observational stance, they are *interacting* with others ... [under] a sense of shared experience or intersubjectivity."[150] What I would like to emphasize here (whether he is aware of it or not) is that Gallagher is careful *not* to say that the infants *actually share* the experience. What he says amounts to the claim that infants are able to perceive, recognize, or understand situations *as shared*. This is more of a fact of embodied *perception* than it is of embodied and communal being-in-the-world.

Still, others take the claim of having direct perceptual access to the bodily expressiveness of others as demonstrating naïve metaphysical realism. There are ways in which we can understand this interpretation to fit in with either of the positions articulated above, or neither, depending on how one defines naïve realism. Some characterizations hinge their definitions on awareness of the external world. Aside from assuming a problematic internal/external division (which may be the influence of a lingering Cartesian dualism) the notion of awareness would need to be elaborated so as to include pre-noetic aspects of experience. At any rate, prolonged discussion of this matter would take us too far afield of our present concerns. While I think Gallagher's position is more defensible than Zahavi's or Schatzki's, their differences are slight compared with the more pressing conceptual matter of an agent's phenomenal comportment or "attunement" of embodied agents toward others.

Zahavi puts the point rather astutely. "Husserl and Merleau-Ponty [contra

---

[150] Gallagher, Shaun. "Two Problems of Intersubjectivity", Journal of Consciousness Studies, 16, No. 6-8. 2009, p. 293.

Heidegger] argue for a place for intersubjectivity in the very intentional relation to the world...As they put it, the subject is intentionally directed toward objects whose horizontal givenness bears witness to their openness for other subjects"[151]. Counting as a subject requires the double recognition of other (explained above), as well as an attitude of "openness" which recognizes and allows one's self to be taken as a self by others, and which allows one also to recognize others as *other selves*[152]. To be clear, however, we should keep in mind that the "recognition" discussed here need not be a conscious or "noetic" recognition. The whole process of recognizing another as another *subject* occurs pre-noetically, or below the threshold of conscious awareness; and one's conscious or noetic recognition is an *effect* of this more primary pre-noetic recognition. In addition, this attitude of openness provides the resources allowing agents to change, and to account for change in themselves and their perceptions of others. Evan Thompson writes "Consciousness is not solipsistically closed in upon itself; rather it is structurally open to the other in advance of any actual encounter between self and other"[153]. For Thompson, the open comportment of intersubjective agents is a structural aspect of embodied cognition, and thus in no way depends on the contingent experiencing of an actual other. Instead the open-structure is an *a priori* condition for the possibility of such an encounter[154]. For example, what allows us to perceive a tomato in its wholeness (e.g. as

---

[151] Zahavi, Dan. "Beyond Empathy" Phenomenological Approaches to Intersubjectivity", *Journal of Consciousness Studies*, 8, No. 5-7, 2001, pg 155.

[152] While the content of this claim seems both accurate and correct, as a criticism I do not think it meets its mark in Heidegger. In his notion of a clearing (Lichtung) Heidegger seems to be pointing at exactly this kind of open comportment.

[153] Thompson, Evan. *Mind in Life*. Harvard University Press. 2007. p. 383.

[154] While I support Thompson's approach here I suspect the details are rather more complex than Thompson suggests. I think that one should understand this open comportment not as a transcendental condition for social encounters, but rather more as a dialectical or reciprocal process of unfolding: my

having a 3-dimensional shape even though we only visually perceive one 2-dimensional facet) is that we perceive it *already* as having the potential to be seen by others[155]. Thus even our comportment toward non-subjective, mundane, objects in the world carries with it this other-intentional structural-openness[156]. This should not be understood as in opposition to what Heidegger expresses in *Mitsein*, but rather as complimentary to it.

Sartre famously critiqued Heidegger on exactly these grounds. For Sartre, the most primordial experience of others is the concrete interaction between agents. This interaction is face-to-face, or "confrontational", and bears the phenomenal character of conflict[157]. He argues that human-made artifacts, or "equipment" (*das Zeug*) only appear to the experiencing agent *as* equipment on the presupposed background of actual, concrete experience with others in this confrontational sense. We learn to use tools or equipment in concrete situations with particular others (face-to-face). Thus according to Sartre, being-with (*Mitsein*) is a comparatively *derived* sense of other-oriented interaction. It emphasizes the similarities or lack of difference between the agent as she encounters equipment, and any arbitrary other as he ('one') would encounter it. Sartre, following Husserl, acknowledges that if similarity is the basis for other-oriented interaction, then the self-other relation is no longer expressed. Instead, as Husserl notes, "Had I the same access to the consciousness of the other as I have my own, the other

---

social encounter determines the character of my openness, and my open comportment determines the character of my social encounter.

[155] One way to think of this might be to understand such perceptual openness as an agent's perception being conditioned by social counterfactuals; "if there were a person nearby, the tomato would be experienceable to that person like so".

[156] Compare this account of the perceptibility of the tomato to, for instance, Alva Noë's sensorimotor contingencies approach—which sees the tomato 'as affording' various potential interactions through the enactive agent's movement. (cf. *Action in Perception* pp. 62-67)

[157] This Sartrean position bears a striking resemblance to the relations of recognition that occur between the lord and bondsman of Hegel's 'Master/Slave Dialectic' in *Phenomenology of Spirit*.

would cease being an other and instead become a part of myself"[158]. Thus, the

phenomenological picture of the self that both Husserl and Sartre (as well as Merleau-

Ponty, and debatably Heidegger) entails a movement from an acknowledgement of an

other, or *alter ego*, immediately to the recognition of the self as recognized by the alter

ego, and finally to the realization that the self, *qua* alter ego, is one among many, and in

no way privileged. All that makes the self unique is its position as bearing the

phenomenal character of *being given to* the self, as well as the recognition that others

cannot, in principle, experience this *particular* relation[159].

    Husserl and Merleau-Ponty take self-otherness, or '*alterity in ipseity*' as a

transcendental condition for the possibility of experience of the *alter ego* (an other self),

and an important basis for the ability to 'empathize' with an other (viz. to recognize

'*ipseity in alterity*'). The experience of self-otherness occurs when one experiences

herself as another would, or could. This arises in Merleau-Ponty's exposition of "double

sensations" or what he later calls the *chiasm*:

> When I press my two hands together, it is not a matter of two sensations felt
> together as one perceives two objects placed side-by-side, but an ambiguous
> setup in which both hands can alternate the roles of 'touching' and being
> 'touched'...I can identify the hand touched in the same one which will in a
> moment be touching. In other words, in this bundle of bones and muscles
> which my right hand presents to my left, I can anticipate for an instant the
> incarnation of that other right hand, alive and mobile, which I thrust towards
> things in order to explore them. The body catches itself from the outside
> engaged in a cognitive process; it tries to touch itself while being touched and

---

[158] This is Dan Zahavi's translation of Husserl. E. *Cartesianische Meditationen und Pariser Vorträge*, *Husserliana I* (Den Haag: M. Nijhoff). (1973) p. 139. Zahavi also likes to point out that Heidegger agrees, at least partially, with Sartre's assessment, citing a personal correspondence from Heidegger to Sartre from October 28th, 1945: "I am in agreement with your critique of 'being-with' and with your insistence on being-for-others, as well as in partial agreement with your critique of my explication of death" (Towarnicki, 1993, p. 84). Both can be found in *Zahavi's Subjectivity & Selfhood: Integrating the First-Person Perspective*, MIT Press, 2005. pp. 154, 237n9.

[159] cf. Zahavi, "Beyond Empathy" p. 160. This is precisely the phenomenal experience of *first-personality*.

initiates a kind of reflection.[160]

Thus for Merleau-Ponty and Husserl, contra Sartre, the experience of self as a self

amongst others can be encountered in absence of a concrete interaction with an actual

other. This experience is of the self as an other through the body "catching itself from the

outside;" which in turn relates the self in its otherness (*alterity*) to the self in its 'my-ness'

(*ipseity*). It is only through this more basic experience that self-consciousness can arise,

and also through this mode of self-consciousness that one is equipped with the skills

required to empathize, or experience the other as one does herself.

Evan Thompson gives a loose sense of what embodiment theorists mean when

they discuss empathy:

> Empathy is a unique form of intentionality in which we are directed toward
> the other's experience. Any intentional act that discloses or presents 'foreign
> experience' counts as empathy. Although empathy, thus understood, is based
> on perception (of the other's bodily presence) and can involve inference in
> difficult or problematic situations (when one has to work out how another
> person feels about something), it is not reducible to some additive
> combination of perception and inference...Rather in empathy we experience
> another human being directly as a person—that is, as an intentional being
> whose bodily gestures and actions are expressive of his or her experiences or
> states of mind.[161]

As previously mentioned, the other's bodily expressiveness is given directly, but not first-

personally. If the bodily expressiveness of the other's pain, for example, were to be

perceived first-personally, it would not be the pain of the other, but rather would be

phenomenally perceived as one's own pain. This is not how the pain of the other is

perceived, however. As Edith Stein (one of Husserl's students) explains, one can "never

---

[160] Merleau-Ponty, Maurice. *Phenomenology of Perception*. Routledge. 1962. p. 106-7.
[161] Thompson, Evan. *Mind in Life*. Harvard University Press. 2007. p. 386.

get an 'orientation' where the pain itself is primordially given"[162].

This problem of 'non-primordiality'—or how to account for a mechanism that allows for the direct, perceptive, empathic feeling of another—is addressed more recently in experimental psychology, cognitive science, and philosophy of mind. One suggestion is variously called the 'common-coding theory' (e.g. by Prinz, Thompson, Carruthers, etc.)[163], or the 'ideomotor framework of imitation' (e.g. by Iacoboni, Rizzolatti, Gallese, Di Pellegrino et. al.)[164]. On this theory planned actions, action execution, and perceived active events are all commonly expressed in a single neural format, or "mentalese". Evidence given in support of this theory comes from 'mirror neurons' which "are activated either by the subject's own motor behavior *or* by the subject's visual observation of someone else's motor behavior"[165]. More recently, Iacoboni et. al. have shown that "goals have higher priority than movements in imitation"[166]. Thus, not only do agents perceive what another is physically doing using their own sensorimotor systems; emotionally feeling by using their own affective framework; or meaningfully gesturing at the proprioceptive level of body-schemata; but rather they perceive and privilege the goal-directed intention of those actions. In this way, embodied agents can be

---

[162] Stein, Edith. *On the Problem of Empathy*, trans. Waltraut Stein. ICS Publications. (1989). p. 7.

[163] Cf. e.g. Prinz, W. (2005). Experimental approaches to action. In J. Roessler and N. Eilan (Eds.), *Agency and self-awareness* (pp. 165-187). New-York, Oxford University press., Carruthers P, Smith PK (1996) *Theories of theories of mind*. Cambridge University Press, Cambridge, UK

[164] Cf. Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653-670. Gallese, V. (2005). 'Being like me': Self-other identity, mirror neurons, and empathy. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science*, Vol 1 (pp. 101-118). Cambridge, MA, US: MIT Press.; Iacoboni, M., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, 3, e79.

[165] Gallagher, Shaun. *How The Body Shapes the Mind*. p. 9.

[166] Iancoboni, "Imitation, Empathy, and Mirror Neurons", *The Annual Review of Psychology*, 60: 2009. p. 655

said to empathize with "how the other is, and what she is doing" in a maximally rich sense. Hence, the historical phenomenological claim that we directly perceive others' actions and intentions seems generally correct. In empathetic emulation, our brain activities, bodily comportments, and sympathetic sensorimotor responses are triggered not by our own bodily situations, but through our perceptions of the situations of others. As Gallese writes "when we observe actions performed by other individuals, our motor system 'resonates' along with that of the observed agent"[167]. This "resonance" is of the same phenomenal kind as any other worldly encounter. When we observe the world around us, our sensorimotor system responds directly to what is bodily available. So too with our empathic, intersubjective encounters. Thus, the evidence provided by "mirror neurons" supports a theory of direct perception in line with Gallagher's rather than Zahavi's or Schatzki's.

Of course, what counts as evidence is often a matter of interpretation and subject to observer biases, so we should probably take the results of these "mirror neuron" studies with a grain of salt until more substantial empirical evidence becomes available. What the studies are good for, however, is in showing that there are some cases in which, in a very concrete and literal sense, we perceive the other's experiences and lived mental conditions; in her intentionality, actions, and bodily gestures; *directly* as belonging to a person like ourselves. This does not mean that we perceive them first-personally, but rather the perception itself contains the experiences *as* the experiences of an other.

---

[167] Gallese, Vittorio. "The 'shared manifold' hypothesis: from mirror neurons to empathy" Journal of Consciousness Studies, 8 (2001), p. 38.

2.3   *Social Practices & Institutions*

Recently there has been renewed discussion about the tensions between social theoretical accounts of 'social practices', and individualistic accounts of experience[168]. If anything about the debate is clear, it is that both kinds of phenomena (social practice, individual experience) deserve to be taken seriously, and require a thoroughgoing and consistent account by which they can together be understood. The purpose of this section is to clarify how the concept of a social practice must align with the ways in which social agents understand their own actions as well as the actions of others. In particular I will be explicating by way of representative example the conflicting notions of *practice* put forth by Theodore Schatzki and Stephen P. Turner (and using our previous discussion of Shaun Gallagher's theory as a touchstone) in order to reveal differences in intuitions about social practices.

In his *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*, Theodore Schatzki develops an intersubjective social theory of practice largely consonant with Shaun Gallagher's. While they share a large set of commitments (for instance, regarding the importance of expressiveness of the body, the direct perception of others in their bodily expressions, and de-privileging 1st and 3rd personal accounts of social learning in favor of 2nd personal interactive and co-constitutive accounts, just to name a few[169]), perhaps the most striking difference is in the direction of causality and explanation. On the one hand, Gallagher is mainly interested in giving an account of

---

[168] For example the Communitarian critique of individualism and the subsequent boon in virtue and care ethics literature spawning from Alisdair McIntyre's and Charles Taylor's respective 1989 volumes *After Virtue* and *Sources of the Self*.

[169] Schatzki, 1996.

perception that explains the ways in which agents can be understood to interact intersubjectively. On the other hand, Schatzki's intent is to give an account of social practices that explains how concepts such as subjectivity and individuality arise and become meaningful through common, shared, social practice. Initially then, one might see these two approaches as incommensurable—after all, even though they appear to be giving largely similar accounts of the actual details pertaining to intersubjectivity, they swap the respective roles of *explanans* and *explanandum*—Gallagher's is a *phenomenological-perceptual* explanation of what amounts to social practices, while Schatzki's is a *social practice* explanation of perception and phenomenology. However, both authors take care to emphasize the importance in recognizing a co-constitutive relationship between (individualistic) embodied perception and (social) practice: individual agents contribute to, and are shaped by, the content of their social milieus. We will return to these themes shortly. First, let us put Turner's position on the table for examination.

Stephen P. Turner's approaches the idea of social practice both more skeptically and more obliquely. In his *The Social Theory of Practices,* rather than engage in the relatively positive project of theory construction, Turner expresses grave concerns about the explanatory power any notion of "social practice" could possibly provide. Turner is skeptical that the notion of a "social practice" is capable of *explaining* anything—instead he claims that the idea functions merely as a "stop-gap" for incomplete explanations. What is particularly helpful for my purposes is his discussion of "theory theorists" and "simulation theorists" in *Brains/Practices/Relativism* which provides precisely the basis

or vocabulary needed to compare the relevant aspects of Gallagher's "body-schema"/"body-image" distinction with the explanatory requirements for an account of social practices consistent with dynamical embodied agency. To complicate matters a bit, Gallagher (and to a lesser extent, Schatzki) gives an argument for the *direct* perception of others' minds that might be thought to diffuse Turner's criticisms. However, in as far as Gallagher's argument succeeds, it only does so partially and at the expense of the impoverishment of resources to combat another of Turner's worries.

Gallagher begins his work in *How the Body Shapes the Mind* by giving a history of the use of various concepts akin to "body-schema" and "body-image". He shows that historically, there has been little care in distinguishing them, and that one result of this lack of care was previous theorists failing to recognize the strengths and merits of an embodied account of cognition. Here "embodied" means that the lived, space-encountering, corporeal bodies of persons factor constitutively in the functioning of cognition. It is not equivalent with the identity theorist's claim that "mind is brain", but rather includes the doings of bodily activities in the functioning of the cognitive system (*Leib*). In considering this body (as opposed to the generically extended "body" of "mind-body dualism", (*Körper*)) we must first recognize that it is distinct in that it is experienced subjectively. However, when we acknowledge that some cognitive functions are carried out "unconsciously" or "pre-reflectively" (or, using Gallagher's term, "pre-noetically") there are really two ways of considering *how* the body is experienced—either consciously (reflectively), or unconsciously (pre-reflectively). Thus, Gallagher initially characterizes the distinction between "body-image" and "body-schema" as follows:

> A *body-image* consists of a system of perceptions, attitudes, and beliefs
> pertaining to one's own body. In contrast, a *body-schema* is a system of
> sensory-motor capacities that function without awareness, or the necessity of
> perceptual monitoring.[170]

Stated colloquially, your body-image is "how you see yourself", while your body-schema

is "how your body 'sees' its place in the world". Thus, the notion of "body-image" is

concerned with *reflective* and *discursive* cognitive modes of perceiving, believing,

understanding, representing, emotionally feeling, etc., while "body-schema" pertains to

the *pre-reflective, non-discursive* body's *ability* or *capacity* to move or constrain

movement, act or maintain posture, etc.[171].

Gallagher emphasizes the concept of *proprioception* as the generalized mode of

body-schematic encounter with the world. Proprioception is "the bodily sense that allows

us to know how our body and limbs are positioned. If a person with normal

proprioception is asked to sit, close his eyes, and point to his knee, it is proprioception

that allows him to successfully guide his hand and find his knee".[172] From this definition

we find that proprioception relates not only to the body-schema but to body-image as

well. In as far as we consider the "know how" of proprioception in its representational or

discursive content (e.g. my knowledge that I am able to successfully run this blind knee

experiment), this knowledge pertains to my body-image, and not my body-schema. On

the other hand, the motor, spatial, and relational ability of my body to succeed at the task

occurs "before" consciousness, perception, thought, or beliefs come or into

consideration—this is the "pre-reflective" character of body-schema. Thus proprioceptive

---

[170] Gallgher, Shaun. (2005) *How the Body Shapes the Mind*. Oxford. p. 24.
[171] Ibid. p. 24.
[172] Ibid. p. 45.

bodily awareness refers to *the lived body's* sense of itself and its surroundings (as considered in contradistinction to *my inner* sense of myself and my surroundings). This characterization of proprioception would be a bit loose according to Gallagher because the sense of "ownership" associated with body-image need not take the 3rd personal, observational stance—one needn't understand their body as an *object* to recognize it as "owned"[173], and likewise, the "inner" sense of self expressed here would strike Gallagher as overly 1st personal.

Importantly, Gallagher sees body-schema not in terms of a transcendental condition for the possibility of perception, representation, belief, emotion, etc.; but, following J. J. Gibson, as an open system of affordance and constraint. The body in its schematic role pre-noetically "organizes" an agent's immediate environment into an environment *for interaction* with the body in ways coherent with the body's capabilities. The environment, before *conceptual* or *reflective* organization, is manifested proprioceptively or body-schematically as allowing (affording) or disallowing (constraining) certain modes of bodily encounter. A good example to elucidate this point is standing in a cramped, unlit mop closet. It doesn't take discursive reasoning for one's body to "know" that this situation does not afford the possibilities of sprinting or seeing, though it may afford grasping the doorknob in a certain way and applying an amount of torque to free itself. It may be argued that it is not the *body* that "knows" these things— instead, this knowledge solely and exclusively resides in the representational, discursive, processes of the brain. But consider the door-opening action. Over the course of one's

---

[173] Ibid. p. 29.

life, one becomes habituated in door opening. It is a common activity, variously enacted

upon many different styles and mechanisms of door latches[174]. Discursively representing

the "high bandwidth" interactive activity of turning a door knob and applying pressure to

the door *cannot be done* on the timescales required to represent that action—there are

simply too many relevant aspects for a serial representational processor to process[175].

To be clear, I am not claiming that there is *no* representing going on in

cognition—rather it is *possible* that there are some cognitive situations in which

representation doesn't or cannot occur. In fact, I don't think there must *necessarily be*

such a situation *in actuality.* In principle, I am comfortable with representation *always*

accompanying (or even playing a *necessary* and integral role) in any or every cognitive

activity. My argument is simply that representation is not *sufficient* for cognition. We

should not think of representation as the entire explanatory basis for all cognitive goings-

on. It is a significant infelicity that such a position is often called "anti-

representationalism"[176]. Justifying this position is rhetorically difficult given the inherited

hegemony of the received view. In order to argue that cognition isn't *all and only*

representation all of the time, the standard of evidence requires finding an example in

---

[174] Opening doors is also a paradigmatic example of a social practice. What makes someone's activity an example of opening doors fundamentally involves a social background involving certain kinds of enclosures (rooms, buildings), hinges, and various etiquettes.

[175] The problem raised here is akin to the philosophical version of "The Frame Problem"—the problem of effectively symbolically representing the relevant aspects of an artificially intelligent robot's environment so as to successfully formulate and solve practical problems in real time. The problem of determining what is relevant from scratch turns out to be computationally "hard" (complexity theory classifies the philosophical version of the Frame Problem as NP-Hard and likely NP-incomplete). Even if the serial symbol processing of classical computational models were exchanged for massively parallel symbol processors, the consequences entailed by complexity theory suggests that adequate solutions to the frame problem are not likely forthcoming.

[176] Calling someone who claims that cognition is not all and only representational processing an "anti-representationalist" is akin to calling someone who is "pro-choice" "anti-life"—it gives the impression that the theorist makes a much stronger than she actual does.

which *no* representation occurs—but there may not be such a thing *even if it isn't all and only representation all of the time.* The thesis I wish to put forward seems to be *underdetermined in principle* by any possible evidence—which is unfortunate. However, suppose the counterfactual hegemonic circumstance where my weaker claim happens to be the received view. The stronger claim of representationalism would *also* depend on a standard of evidence that requires finding an example in which *only* representation occurs—but there may not be such an example *even if it is all and only representation all of the time.* The issue of theory decision is here rhetorical, not genuinely philosophical (or empirical). It is more so a function of which theory occupies the privileged position afforded to a received view than it is a function of which theory has the superior empirical support—neither does (from the other's perspective).

Now that we have sorted out the distinction between "body-image" and "body-schema" and their relations to proprioception, we are in a position to see how they factor into considerations about intersubjectivity and social practices. There are, of course, two maneuvers necessary to clarify the relevance of body-schema and body-image to social practices. First one must give an account that shows how body-schema and body-image differentially participate in social *interaction*. Then one must give reciprocally constitutive accounts of social interaction and the functioning of practices, customs, and other social institutions that shape body-image and schema. The general argument is that only dynamic, embodied, extended agents of the kind I have described are capable of enacting the kinds of social practices, customs and institutions that we actually do. A failed attempt at adequately motivating this reciprocal relation results in one of the

97

following outcomes:

> (1) An account with dynamic embodied agency as *explanans* to the *explanandum* of social practices,
> (2) An account with social practices as *explanans* to the *explanandum* of dynamic embodied agency, or
> (3) An account that is unable to 'bridge the gap' between dynamic embodied agency and social practices.

If the outcome we arrive at (3), there are three further possible explanations:

> (A) We must abandon a dynamic embodied account of agency in order to salvage a workable theory of social practice (deny the gap – no embodiment),
> (B) We must abandon the coherence of the notion of social practice in order to salvage a workable theory of embodied cognition (deny the gap – no social practices), or
> (C) Adopt a dualistic or pluralistic account which recognizes an unbridgeable explanatory gap between the type of account of agency we can give, and the type of account we can give about social practices (embrace the gap).

I will argue that both Gallagher and Schatzki fail to maintain the reciprocity to which they both pay lip-service, and that instead, Gallagher's argument leaves him with an account resembling (1), Schatzki's argument leaves him with an account resembling (2), and Turner's argument affirms something like (3-B). An argument resulting in (3-C) is unsatisfactory *as an explanation*[177], though *if* it turns out that it is the only defensible position, we may accurately call this the "Wittgensteinian Approach", as Wittgenstein never sought to explain, but rather to merely clarify by describing what is the case in a way that coheres with our concrete and ordinary phenomenological and social encounters in the world. Interestingly, contrary to what Schatzki claims, I do not think the approach he gives is "Wittgensteinian" enough to merit the title of his book (though our interests

---

[177] That is, it is unsatisfactory as an explanation of the relation between social practices and an account of agency because it denies there is such a relation. It amounts to explaining away the problem rather than addressing it.

here do not turn on this point)[178]. Again, the purpose of this chapter is to avoid all of (1) -

(3) (and thus 3A-C), by giving accounts of agency and social practices that adequately

shows how they dynamically co-constitute each other (viz. dynamic embodied agency

constitutes social practice(s), and social practice(s) constitutes dynamic embodied

agency). We may call this the "*co-constitution criterion*" (hereafter "CCC").

2.4   *Turner's Criticism of Social Practices*

In order to show how Gallagher's and Schatzki's accounts fail the CCC, I will use

Turner's criticism of social practices. In *The Social Theory of Practices*, Turner outlines

an argument that shows how any attempt to make good on the "promissory note" left by

those who invoke the concept of "social practice" must inevitably fail. In terms of

explanatory fecundity, if this negative argument succeeds the concept of practice, along

with a constellation of similar concepts (traditions, customs, institutions, habits, etc.),

must be cast to the junk-heap of philosophically useless concepts. Turner's argument

amounts to the claim that to explain an action as instantiating a practice is precisely to fail

to adequately satisfy the reasonable minimal criteria for explanation. He differentiates

the problems faced by notions of practice along three axes:

> (I)   [Transmission]: How practices can be *transmitted* between and among
> individuals within a society, culture, or tradition,

---

[178] Wittgenstein was the paradigmatic case of what would later be called an 'ordinary language' philosopher. His intention, at least by the time he worked on *Philosophical Investigations*, was to tidy up the way we think about language by analyzing how we actually use it in concrete situations. For Wittgenstein then, there is no problem of explaining how metaphysics of agency and social practices relate—namely because Wittgenstein simply was uninterested in metaphysics. If it turns out that an analysis of our use of language reveals inconsistencies between how we talk about agency and how we talk about social practices—this just points to a philosophical confusion—a bungle that only philosophers could have gotten themselves into—rather than some deeper underlying problem. Wittgenstein doesn't attempt to clear up such 'philosophical confusions' but seems to think that revealing them as such is sufficient.

(II) [Diachronicity]: How practices can be said to both *persist* and *change* over time in a coherent way, and

(III) [Synchronicity]: How individuals that *differ* in their situational experiences can participate in a practice if nothing is *shared* between them.

His argumentative strategy is to show that any adequate theory of social practices must simultaneously solve all three problems, and that attempting to solve the problems on any one axis results in exacerbating problems along another or both of the others.

If someone takes transmission (I) seriously as a problem, it betrays the supposition that there are objective contents to be transmitted, essentially sedimenting and reifying the contents practices[179]. This in turn makes the persistence in (II) a bigger problem. On the other hand, if one takes the issue of diachronic change in (II) seriously, it becomes difficult to articulate exactly *what* could be shared by social practitioners, and what kind of cognitive mechanism could enact the transmission of such continually dynamic practices[180]. Finally, when we take the issue of synchronic differentiability (III) seriously, we seem to be left with nothing to transmit in (I), and no way to track persistence or change in (II). Ultimately, Turner concludes that:

> [T]he idea of the special persistence of *mores* or traditions was thought to necessitate the hypothesis of the existence and transmission of some sort of collective object. But there is no need for any such collective object. The same kinds of persistence can occur entirely through individual (and possibly literally different) habits that arise in the individual as a *consequence* of the emulative performance of particular activities, observances and the like. (*STP* 98-99)

---

[179] In chapter 3 I argue for a process-based ontology of systems that is incompatible with the presumed substance—or reified object—based approach that underwrites the problem of transmission. By instead taking a process approach, the problem of transmission is averted, though a different account of communicability must be provided. The interactive dynamic systems approach I describe in chapter 3 does precisely this.

[180] Again, this is an ontological issue – and one that is sufficiently addressed via Dynamic Systems Theory.

The solution is to do away with reference to social practices altogether, and in their explanatory stead talk of "individual habits". These habits, as we will now see, are nothing more than learned bodily skills—alterations to one's body-schema and body-image[181].

All parties involved (Gallagher, Schatzki, Turner, and even Wittgenstein) value the explanatory relevance of dynamic embodiment in agency. In Turner's *The Social Theory of Practices* these commitments are displayed in the following passages. For instance, In his discussion of "the Mauss Problem", he writes:

> Marcel Mauss, the nephew and student of Durkheim, provides the classic case of the discovery of a practice in his essay on techniques of the body. In this essay he recalls lying ill in a hospital in New York:
> I wondered where previously I had seen girls walking as my nurses walked. I had time to think about it. At last I realized that it was at the cinema. Returning to France, I noticed how common the gait was, especially in Paris. The girls were French and they too were walking in this way. In fact, American walking fashions had begun to arrive here, thanks to the cinema. ...Mauss could distinguish the walk as habit because he could say that the difference in walks he had noticed was not a natural difference, and he could say that it was not a natural difference because he could give a historical account of it. He started, so to speak, within a culture with its expectations...An exterior performance, such as a manner of walking, is the result of the compositions of many causes – body properties, shoes, training, and setting. The 'same' external walk can be produced by various combinations of causes.[182]

Here, Turner shows how Mauss' assumptions about the different ways in which embodied activities are learned (either through individual lived experiences or through

---

[181] To anticipate the account I give in chapters 3 and 4, we may describe habits as behavioral "attractors" established through various interactive feedback mechanisms between an individual agent and her local environment. The patterns that underlie such behaviors are determined both by the agent's self-guiding norms (developed through the agent's embodied sensitivities to the various affordances in her local environment) as well as the negative and positive social-normative feedback provided by others. I take the issue of pattern-detection – and therefore the issue of identifying habits and norms – to be a primarily epistemic issue (for more on this, see my discussion of "systems" in chapter 3).

[182] Turner, Stephen P. (1994) *The Social Theory of Practices*. Chicago Press. 1994. pp. 21-22.

manifold forms of social conditioning), factor fundamentally in social practices or

'habits'. Along the same lines Turner later writes:

> Two people may learn to ride a bicycle equally well, in the sense that they are able to perform the same tricks on the bicycle, or guide the bicycle through the same manoeuvres. But each person, in acquiring this skill, has done so with a different instrument, their own body, which they have trained in a different way. The 'differences' may be seen from the history of their teaching themselves the skill, or learning it...Not only may they not be identical, there is every reason to suppose they are not identical.[183]

Quite similar to Gallagher's point about body-schema constraining and affording

different possibilities, Turner here emphasizes the malleability of the body as a skill-

bearing locus, and that the status of one's possession of a skill depends solely upon the

status of the body—which changes in ways that inform that skill. Additionally, this

passage also shows that Turner doesn't distinguish as Gallagher does between pre-noetic,

and perceptual learning. For Gallagher one can either consciously or unconsciously train

himself to ride the bicycle—bicycle riding is either a body-schematic skill or

intentionally manipulated through one's body-image. The distinction is appears

extremely important here because it signifies the difference between success and failure

for Turner's general argumentative scheme. If the possibility of intersubjective

communication depends crucially on conscious discursive expression and skilled bodily

learning occurs only at the pre-noetic or body-schematic level, then Turner's argument

for the impossibility of sharing the skilled practice of bicycle riding via discursive

communication succeeds. However, if the possibility of intersubjective communication

depends on *conscious*, attentive, discursive communication, and skilled bodily learning

---

[183] Ibid. p. 58.

can be expressed consciously and discursively, then there is absolutely no problem with *sharing* practices. While this purported solution may seem enticing, it would be prudent to consider that social practices are themselves complex and dynamic, and it no trivial task to explain the changes that practices can undergo, and how such changes in practices relate to the concrete interactions that a multiplicity of agents enact.

Alternatively (and I think rightly), it seems Gallagher would argue that intersubjective communication must occur pre-noetically between two or more subjects' *body-schemata*[184]. Turner's language here seems to suggest that many of the aspects of learning the skills of bicycle riding can only occur consciously (at the level of *body-image*). If this is the case, Turner's argument once again succeeds: some of what is necessary for skill/practice transmission (specifically one's idiosyncratic body-image) cannot be transmitted. Later in *The Social Theory of Practices*, Turner seems to problematize learning by transmission in exactly this way:

> [consider] for example a robot replacing a lathe-operator's body together with a traditional mechanical lathe. The lathe-operator obviously does not possess the 'knowledge' to run the robot's body. The lathe-operator's embodied achievements and powers (to perform a specific range of tasks) are what is being emulated...In this case something – a computer-driven robot – is emulating an activity, but obviously not literally following the same rules. But if something – a machine driven by a computer – that is not following the same rules can emulate, why can't this same kind of relationship, of emulation, occur between persons?[185]

Here, what may at first blush appear to be an argument for a multiple-realizability thesis

---

[184] The picture here can very quickly become quite complex. There is regular traffic between body-schemata and body-image for any given agent. We could ask whether the issue of communication between two or more agents is a matter of body-schema or body-image (or both). Phenomenally it may appear to agents that expressiveness in ones body-image is what is being transmitted and received. However this may simply be the result of agents identifying their body-image with their body-schema in a certain closely-coupled way.

[185] Turner, Stephen P. (1994) *The Social Theory of Practices*. Chicago Press. 1994. p. 98.

consistent with the functionalist theses of Putnam and Fodor amongst others, actually

suggests a deeper point: that there can be no hope for multiple realizability. Rather, each

differently-shaped, differently-trained body cannot help but run its own unique, non-

portable "program"—because each body is necessarily different from others. One cannot

"possess the 'knowledge' to run [another] body". Turner's argument undermines the

expectation that intersubjective communication can be achieved through emulation,

regardless of whether it occurs pre-noetically or consciously. The negative result

produced by Turner's argument is strong enough to eclipse the importance of Gallagher's

body-schema/body-image distinction because—if Turner is correct—it doesn't matter

whether communication is supposed to occur between agents' body-schemata or their

body-images; since the differences in *both* abrogate any possible transmission of

expressiveness from one agent to another.

2.5   *Gallagher's "*Interactionism*" and Turner's "*Emulationism*"*

Interestingly, while discussing theory theory (TT) and simulation theory (ST) in

his *Brains/Practices/Relativism*, Turner supports what he calls an "emulationist" theory.

Theory theory is the idea that one's communally-shared, culturally-inherited

understanding of folk psychology constitutes a theory of mind, while simulation theory is

the idea that agents 'simulate' the mental activities of others via their own mental

activities. The issue at stake here is different than those discussed above however. By the

time we get to discussing theories of "mind reading" (coming to understand the thoughts,

desires, emotions, etc. of others) such as TT and ST, we've moved beyond theories of

social communication. Indeed, both TT and ST share some key assumptions that

Gallagher later goes on to question in his "Two Problems of Intersubjectivity". There, Gallagher treats concerns with three assumptions implicit in both Theory Theory and Simulation Theory. The first of these assumptions is that minds are essentially individual and autonomous—there is a gap in accessibility between one subject and others. This is taken to imply that to solve the "problem of other minds" one needs to *infer* the mind states of others. Second, the inference of others' minds occurs essentially *3rd personally* as an observer from the outside looking in. And third, that the above mentioned modes of inferring how others think *explains a vast majority* of the instances of inferences about what others think. Gallagher instead refutes at least the third claim—that if one does anything like the above, it is only in somewhat rare circumstances. Alternatively, Gallagher proposes a third theory of mind option, his "Interaction Theory"[186], which states that we *directly* perceive others' intentions in their dispositions or embodied behavior, trading the 3rd person observational assumption, for an unmediated 2nd personal stance.

Gallagher (again, I think rightly) refuses to make the argumentative move[187] that places individual, embodied agents out of direct perceptual contact with each other. This is important regarding the critical arguments presented in Turner's *The Social Theory of*

---

[186] It should be noted that the theory of Dynamic Embodied Agency that I develop here bears a close kinship with approaches that have been called "Interactivist". My approach shares much in common with both "Interactionist" and "Interactivist" theories; but these two are quite different. As described on the Institute for Interactivist Studies official website: "Interactivism is a metaphysical and theoretical approach to understanding phenomena of biology, mind, persons, and social reality. At the metaphysical level, it emphasizes a strict naturalism and a process metaphysics as essential to the understanding of emergence…The term "interactivism" derives from the model of representation as emergent in interactive systems." (http://www.lehigh.edu/~interact/ retrieved 2/2/2013)

[187] That is, the argumentative move assumed by Theory Theoretical and Simulation Theoretical models of the mind that agents are more-or-less isolated from each other and autonomous.

Practice because it appears to sidestep his problem-schematic[188] (a possibility ruled out by Turner). However, we shall see that Gallagher's argument for "direct interaction" does not evade Turner's critical arguments. Initially, Turner seems to agree with the basic aspects of Gallagher's interaction theory, writing that "It helps [Turner's "emulationist" approach] if the basic material with which "understanding" develops includes some common starting points [..or..] basic emotions, and there is important evidence that there are some universal feelings with universal facial expressions"[189]; also "all that understanding another person means is to be able to *interact*, and this means to play the relevant roles,"[190] though he continues the sentence in a way that betrays a major difference from Gallagher: "and to take the attitude of the other in the relevant stereotyped ways, and to employ the significant symbols in this activity of attitude taking"[191]. He continues to elucidate a more or less *simulationist* position that should now sound familiar:

> What we 'acquire' is no more than the results of our own attempts to interact on the basis of our hypotheses about the attitudes of others and the feedback that enables us to improve our attempts to take the attitudes of others is the success and failure of our interactions. There is no more than this to understanding and specifically nothing in the way of a 'system of conceptions' that must be 'inherited' in order for us to understand.[192]

Because Turner doesn't make use of a distinction between body-image and body-schema,

---

[188] Turner's problem-schematic is that all explanations invoking notions of 'social practice' must necessarily face the problems of transmission, change, and sharing—and that facing one means exacerbating another. Gallagher "sidesteps" this scheme by eliminating the mediation that underlies the problems of transmission, change, and sharing. These are only problems when agents' interactions are mediated—when there are boundaries preventing direct access to others. By hypothesizing direct access to others' bodily states, Gallagher successfully circumvents these concerns—perhaps trading them for others.
[189] Turner, Stephen P. (2002) *Brains/Practices/Relativism*. Chicago Press. 2002. p. 67.
[190] Ibid. p. 68, emphasis added.
[191] Ibid. p. 68.
[192] Ibid. p. 68.

his picture of communicative interaction involves one agent discursively making inferences about another. On the other hand, Gallagher need not hypothesize that communicative expressibility trades in, or depends upon body-image at all; and if body-image is not implicated in communication, then there is need for neither discursivity nor inference. The problem that disallows Gallagher's interaction theory exemption from Turner's criticism of social practices is that there is no *principled* way of distinguishing it from Turner's indirect, inferential account of making sense of others. If the direct access agents have of others bodies occurs at the pre-noetic, schematic level of bodily expressiveness (and much of the work in Gallagher's book is at pains to motivate and support precisely this claim) and does not transgress the threshold of conscious inference, then there can be no way to guarantee the directness of that access. Put perhaps more pointedly, the kind of expressiveness to which agents have direct access in others' bodies is exactly the kind of expressiveness available in non-agential objects, animals, plants, rocks, and T.V. screens. Interaction theory collapses into mundane simulation or emulation theory in that it neither requires nor can help itself to a theory of *mind*. The direct interaction is not between two or more agents (or minds) but is between two or more body-schemata; and there is no obvious sense in which body-schemata *are, in themselves, minds* (or agents). What is needed is some argument that shows how body-schema is *sufficient* (not merely *necessary*) for agential (or "minded") action. If all I am—as a mind—is pre-noetic body-schema, what are we supposed to make of the noetic, the discursive, and the representational? Are these not also properly constitutive of the mind? It seems here we have again run up against Adams' and Aizawa's problem of the

107

"mark of the cognitive"[193]. At any rate, they are precisely what Gallagher proscribes in

his Interactionist account. In effect, he closes both the epistemic and metaphysical gaps

between different agents' bodies (and body-schemata), but in so doing widens a gap

between body-schema and body-image. Perhaps lived bodies interact with each other

directly as Gallagher says, but this provides no reason for eliminating the inference an

agent must make in relating his body-schema to his body-image. Effectively, all

Gallagher has done is shift the indirect inference from between isolated agents' bodies to

between isolated body-schema and body-image within an agent. Indirect and inferential

access occurs not *between* agents, but *within* agents. Gallagher believes his direct

Interactionist account evades the issues caused by mediation but it instead merely

internalizes them.

2.6   *Schatzki's Social Practice Theory of Agency*

In league with Gallagher, Schatzki argues for a social theory of practices in which

agents directly perceive each others "conditions of life" (Schatzki's term of art for the

expression of "how things stand or are going for someone"—basically one's general life

situation[194]). Schatzki echoes both Gallagher's and Turner's de-privileging of the 3rd

personal observer stance in favor of a 2nd personal interactive one. He writes "a

---

[193] That is, the question "what is to count as cognitive?" or "in virtue of what is something a mind?".
Adams and Aizawa ask for the necessary and sufficient conditions for counting as a mind; and while all
of Turner, Schatzki, and Gallagher resist taking the bait, it seems none succeeds in evading this
problem.

[194] I am tempted here to compare Schatzki's "conditions of life" with Gallagher's "body-image", but I
think an identity is too strong. They are in some weaker sense analogues—both are a matter of how one
sees one's self. However, Schatzki's "conditions of life" may be broader, and contain the concrete, pre-
noetic bodily and physical conditions of one's situation as well. Since he doesn't discuss the phrase in
these terms it is difficult to say. Nonetheless, when he uses it, it tends to refer to the way in which one
sees their life (in a diachronic sense). It is an assessment of how one's life is going, here and now.

functional adult's extensive bodily repertoire of doings and saying is social in the sense

of being acquired through learning and training in the context of others' activities...An

expressive body that manifests, signifies, and constitutes conditions of life is thus a social

product"[195]. He continues by fleshing out four different ways that individual agents are

"characterized *by virtue of* their participation in social practices"[196].

The first feature Schatzki outlines is that of "*being in*" a particular, contextualized

"condition of life". He says that "to be in a condition of life is for things to stand or to be

going some way that is expressed in doings, sayings, sensations, and images." First to

note with the idea of being in a condition of life, then, is that it does not distinguish as

Gallagher does between body-schema and body-image. For Gallagher, in order for

something to be 'expressible in doings' is for it to be proprioceptively capable—it

implies having available a pre-noetic body-schema through which doings are

meaningfully expressed (again, pre-noetically) amongst embodied agents engaged in an

act of participatory sense-making[197]. Even sayings and sensations can be thought to be

expressible though body-schemata. Since Gallagher takes expression to occur at the *pre-*

noetic level, it should not be possible for conditions of life to be expressible via (body)

images. Indeed, in listing "sayings" and "images" what Schatzki suggests is that

conscious conceptions of body-image are also *directly expressible*—a possibility ruled

out by Gallagher (body-image is, if anything, *transmissible,* not directly perceived

---

[195] Schatzki, Theodore. (1996) Social Practices: A Wittgensteinian Approach to Human Activity and the Social. Cambridge. p. 70.
[196] Ibid. p. 70.
[197] The phrase 'participatory sense-making' is a term of art introduced by Hanne De Jaeger and Eziquel Di Paolo. Cf. e.g. De Jaegher H, Di Paolo EA (2007) Participatory sense-making: an enactive approach to social cognition. Phenomenology and Cognitive Science 6(4): 485–507

amongst embodied agents)[198]. At any rate, Schatzki's point in highlighting the feature of being in a condition of life is to emphasize the word *in*. He writes that "understanding particular conditions, as much as understanding types thereof, is *part of the background* against which behavior and inner episodes express particular conditions"[199]. For Schatzki and others (notably Searle), a "background" is the pre-representational, meaning-bearing substratum from and according to which individual agents situate their particular understandings of their own, and others conditions of life[200]. Individuals are always already immersed in and structured by the social conditions given by—and constitutive of—their environments.

Schatzki however attempts to sidestep some of Turner's concerns by incorporating considerations of complexity and ambiguity in conditions of life. He puts the problem thusly:

> Strictly speaking, understanding determinately institutes specific conditions only in conjunction with the entirety of the context of behaviors and inner episodes occur. These contexts, furthermore, are invariably complicated, embracing events in the immediate settings of action, wider social situations and practices, past and future behaviors and inner episodes, other conditions of life, and physical states of bodies. What is more, even given specific expressions and the totality of contexts in which they occur, conceptual understanding might not unambiguously determine that a person is in some specific condition. Common locutions for mentality and activity are extremely flexible, and different words can often equally well capture how things stand or are going for someone at some point.[201]

---

[198] That is, when Schatzki says that being in a condition of life is "for things to stand or to be going some way that is expressed in doings, sayings, sensations, and images" what he has in mind is that doings, sayings, sensations, and images are shared amongst those who share the same life condition (and for Schatzki, this is not an empty set. Turner would disagree).

[199] Schatzki, Theodore. (1996) *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge. p. 70 (emphasis added)

[200] cf. e.g. Turner, Stephen P. (1994) *The Social Theory of Practices*. Chicago Press. 1994. pp. 35-41.

[201] Schatzki, Theodore. (1996) *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge. p. 73

Schatzki is here concerned with the problem of underdetermination[202]. Any explanations that must make essential reference both to direct perceptions of others mental conditions as well as a background of social intelligibility will need to address the issue of how *particular* expressions can be honed-in upon. This concern is similar to Wittgenstein's preoccupations in the celebrated private language argument sections of *Philosophical Investigations*, where we [philosophers] are faced with the problem of verifying or explaining *which* rule is being followed when one would like to "the" rule. Of course, on some readings of these sections, Wittgenstein's purpose is to show that, in normal everyday doings and uses of seemingly rule-governed behavior, this problem precisely *doesn't* arise—it is only a problem when we set to *explaining* what we do. Schatzki makes the same rhetorical move. He notices that "in the rough and tumble of real life we are of course rarely appraised of all contexts of behavior"[203] and instead we must simply rely on whichever background criteria arise in the particular considerations of the context—we recognize whichever "behavioral phenomena by virtue of which it makes sense to say that someone is in a given condition"[204]. This, of course, does not give us transmission of expression across the intersubjective void between isolated agents; nor does it allude to "intending to", or "triangulating amongst" (a) one's understanding, (b) the expressivity of the other's body, and (c) a Platonic "background" (as Searle or

---

[202] This problem of "underdetermination" is precisely what Kripke (1982) and Brandom (1994) – in dealing with Wittgenstein's remarks on rule-following – describe as the "gerrymandering problem" – the problem of being able to, post-hoc, re-draw the boundaries of the terms of discussion so as to make them fit with whatever inference is being made. This problem more generally applies to "transcendental" inferences to the "conditions for the possibility" of factual states of affairs. Transcendental inferences, because they're always undetermined, can always be "gerrymandered". For a similar characterization see Turner's *Brains/Practices/Relativism*, pp. 122-4.

[203] Schatzki, Theodore. (1996*) Social Practices: A Wittgensteinian Approach to Human Activity and the Social.* Cambridge. p. 74.

[204] Ibid. p. 74.

Davidson might have it); and importantly it doesn't necessitate direct perceptual apprehension of one another (as with Gallagher). Rather, Schatzki's more modest understanding of intersubjective communication bears a striking resemblance to Turner's position in *Brains/Practices/Relativism*—namely that "What we 'acquire' is no more than the results of our own attempts to interact on the basis of our hypotheses about the attitudes of others and the feedback that enables us to improve our attempts to take the attitudes of others is the success and failure of our interactions"[205]. Compare this to Schatzki's claim that:

> People's evaluations of the bearing of specific context also often deviate. So our judgments of others' conditions diverge, and we occasionally challenge a person's self descriptions...[one] is of course familiar with life conditions and her situation, but this familiarity need not be explicitly drawn upon. She just acts. And what conditions are thereby expressed depends on what she does and says, the contexts in which she acts, understandings of life conditions, and possibly whatever inner episodes (if any) occurred.[206]

If this was all Schatzki thought was necessary for the constitution of persons by way of social practices, his views would be in line with both Turner's and Wittgenstein's. However, he includes three additional features which he thinks fundamentally bases explanations of persons and intersubjective interaction in social practices.

He tells us that the second feature of social practice constitutive of individual agency is our "abilities to describe, explain, and report one's conditions to others"[207]. This equates fairly straightforwardly with the pre-noetic, body-schematic expressibility that underwrites Gallagher's notion of direct perception of others. On the other hand,

---

[205] Turner, Stephen P. (2002*) Brains/Practices/Relativism*. Chicago Press. 2002. p. 68.
[206] Schatzki, Theodore. (1996) *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge. pp. 74-75.
[207] Ibid. p. 75.

Schatzki says that when it comes to describing, explaining, and reporting one's conditions to himself "he usually ascribes them on the basis of nothing at all...there is nothing articulable or designatable in the experience of a person whose understanding is formed within social practices that informs him what his conditions are"[208]. Gallagher however certainly allows for the possibility that body-schematic knowledge can cross over into and become body-image knowledge, claiming that "there are reciprocal interactions between pre-noetic body-schemas and cognitive experiences, including normal and abnormal consciousness of the body"[209]. Unfortunately Gallagher does not do much to flesh out precisely how body-schema and body-image *do* interact[210]. Even more puzzlingly, this seemingly magical ability to unproblematically transmit one's life conditions to others seems to directly oppose the non-totalizing, complex, ambiguities of the first feature Schatzki describes.

The third feature is the compliment to the second—"the ability to identify others' conditions"[211]. Here Schatzki distinguishes as others (including Gallagher) do between *perceiving* and *inferring*. Following Wittgenstein, he says for instance that we simply do not infer other's pain states:

> "We *see* emotion." - As opposed to what? - We do not see facial contortions and make inference from them (like a doctor framing a diagnosis) to joy,

---

[208] Ibid. p. 76. One way of understanding this is that, for example, when I feel morose, I do no give myself reasons for the feeling—I do not feel morose on the basis of anything at all—I simply feel that way (directly).

[209] Gallgher, Shaun. (2005) *How the Body Shapes the Mind*. Oxford. p. 35.

[210] One may see parallels between this difficulty in Gallagher, and the difficulty Kant had in expressing how precisely schematism is supposed to succeed at taking non-representational intuitions to produce representational understanding in the intellect. If they're distinct and separate kinds, exactly how do body-schema and body-image interact? This problem is exacerbated the problems of inferences just discussed.

[211] Schatzki, Theodore. (1996) *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge. p. 76.

grief, boredom. We describe a face *immediately* as sad, radiant, bored, even when we are unable to give any other description of the features. - Grief, one would like to say, is personified in the face."[212]

Schatzki continues by explaining how the ability to perceive others' life conditions depends on a background (pre-noetic or schematic) "understanding" of the social practices in which such expressions can arise meaningfully. This understanding is acquired through the (often tacit) recognition of patterns of behavior. This pattern recognition is not some kind of inferential calculus; working out a complex regression analysis of previous or stereotypical life situations to compare degrees of similarity of features[213]; but rather such previous experiences factor in precisely as a concrete historical conditioning of the agent's body through habituation and learning. In particular contextual situations, experience is always novel and can only be made intelligible by drawing from the historically, and socially established background patterns of behavior[214,215]. Recall that all this is supposed to occur in the moment of direct perception of another's bodily expressiveness. The problem here is that more of an account for the mechanics of learning needs to be given (even at the pre-noetic level of body-schema)— how does one first come to recognize the patterns beset by the background of social

---

[212] Wittgenstein, Ludwig. (1967) *Zettel*. Trans. G.E.M. Anscombe., G.H. von Wright, G.E.M. Anscombe ed.'s. University of California Press. p. 225.

[213] Presumably contra Brandom's "inferentialism".

[214] Schatzki, Theodore. (1996) *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge. pp. 77-78.

[215] But this background is itself dynamic. Compare here to Derrida's notion of iterability. When one uses a word in a particular context, the use of this word both derives its meaning from past uses and contexts, but also from its novel application here and now. Thus, every time a word is used, it cannot mean the same thing as previous uses for two reasons: (1) since its last use, it has accumulated one more use-context, which alters its meaning, and (2) since this situation is novel, its meaning is indexed to this very situational novelty. Thus, through repetition (using a word, time and again) its meaning is "iterated". Schatzki's "background" must work the same way.

practice[216]—e.g. how does the background *become* the background? Like Turner, Schatzki's explication of inferential identification of others' conditions largely follows a simulationist tack[217].

The last feature of individual experience that exhibits how it is constituted by social practices is "possession of 'convictions' that, in Wittgenstein's words 'hold fast' in a human life"[218]. Citing passages from *On Certainty* (mainly 88-105) Schatzki explains how one's convictions maintain their normative force without ever being explicitly taught or discursively formulated; "rather, people's understandings of the concepts of belief and doubt, in conjunction with their wider ability to grasp particular languages, is simply such that when certain statements are made, perhaps for the first time ever, people are not in the position to doubt them"[219]. Presumably these convictions function as the persistent, normative, structure-giving elements of the pre-noetic social background of practices. In this case, the reason explanations 'bottom out' at practices and we must cease attempting explanation, is because there simply is nothing more to articulate about our current

---

[216] Especially when this "recognition" is often tacit. In a sense, the problem here is akin to the poverty of stimulus problem Chomsky addresses with his theory of generative grammar. Infants seemingly lack sufficient exposure to well-formed grammatical utterances—yet they're capable of linguistic expression far beyond what their exposure affords. What could explain this? Chomsky has made a career of fine-tuning his theory, but his general response is a transcendental inference that there must be some kind of innate "language acquisition device" or "generative grammar"-supporting structure in the brains of competent language-users. Cf. Chomsky, Noam. (1980). *Rules and Representations*. Oxford: Basil Blackwell.

[217] Schatzki's formulation of "background" is (compared to Searle's) quite nuanced and helpful. The Dynamic Systems Theory (DST) approach I outline in chapter 3 agrees in large part with Schatzki's. What is needed is an explanation of the processes or mechanisms whereby a "background" can be established or "bootstrapped". The model Richard Campbell describes (considered at length in chapter 3) can be used to do precisely this—it gives a basis for a naturalistic model of dynamic embodied learning that provides sufficient conditions for the kinds of social resources required for establishing such a "tacit" background.

[218] Schatzki, Theodore. (1996) *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge. p. 80.

[219] Ibid. p. 82.

actions or practices—as Wittgenstein writes "If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what I do'"[220]. To explain convictions grounded in a non-discursive background is to reach beyond what is epistemically available. In this way we can see how the conviction of acting in a certain particular way—in acting and expressing oneself according to social conventions or practices—is understood as 'reaching bedrock' or being unable to give further explanation or justification. However, what adds to confusion here, at least in considering pre-noetic constraints or criteria for action (including speech action), is in trying to differentiate between consciously-held *convictions* on the one hand, and mundane non-discursive, bodily-learned *skills* (like riding a bicycle) on the other. Both are presumably normative in the relevant ways. There can be no hard line in understanding "this is simply what I do" as a matter of *conviction* or of individually learned, historically habituated, skilled bodily activity[221]. However, I suspect that nothing in Schatzki's account hinges on this issue.

It should now be quite clear that Schatzki's claim to co-constitution between individual embodied agents and the background of social practices is merely lip-service. His project is straightforwardly one in which social practices unidirectionally constitute the embodied individual, and not the other way around. Earlier I said that if it turns out that the only defensible position was (3-C) (embracing the unbridgeable gap between

---

[220] Wittgenstein, Ludwig. (1953/2001*) Philosophical Investigations*. Trans. G.E.M. Anscombe. 3rd Edition. Blackwell. §217

[221] Rather, if there is a difference, it must be a difference of degree. In *How the Body Shapes the Mind*, Gallagher argues for a difference of degree between body-schema and body-image, though he does not spell out exactly how this is supposed to work out. I suspect he realizes it body-schema and body-image must lie on a continuum, but cannot give good grounds for this and maintain the significance of the distinction simultaneously. Further, we have already seen that there appears to be an important sense in which body-image and body-schema are epistemically distinct.

embodied cognition & social practices), we could accurately call it the "Wittgensteinian Approach". While such an approach, under such conditions would be Wittgensteinian, I do not think it is the only defensible position.

All three of Gallagher, Schatzki, and Turner recognize this as well. The only defensible position is the one embraced by both Gallagher and Schatzki—which includes both bottom-up and top-down constitution of embodied agents and social practices[222]. It is a *phenomenological* fact that for each individual agent, we enter a ready-made world socially imbued with significance. As Wittgenstein recognized, however, socially imbued significance does not in-itself constitute *meaning*. For this, we need actual agents engaged in concrete experiences of bodily encounter in and with the world. Gallagher's distinction between body-schema and body-image does some work in helping to untangle some of the details of this encountering, but it does not go far enough in bridging the explanatory gap between phenomenal experience and the significance of social practices. Specifically, the body-image/schema distinction fails to adequately address the top-down (social practice) constitution of embodied subjects. Schatzki's "Wittgensteinian" theory of social practice puts the perceiving agent in a role like that of a patron in the Cartesian theater—significance is bestowed from without, from the inherited background of intelligibility always already given. Agents may come to participate in meaning-making, but only derivatively as a result of their absorption of the social-practice background. Thus Schatzki's theory fails to adequately address the bottom-up (phenomenal) constitution of embodied subjects.

---

[222] That is, the co-constitution criterion, or CCC.

I take it that Schatzki does much work in motivating the importance of giving a social-theoretical account of practices, and that Turner succeeds in elucidating the kinds of problems that have traditionally plagued such explanations. But I think Gallagher; with his more recent work on Interaction Theory and Mental Institutions[223] is beginning to give the appropriate attention to a properly phenomenological account of social institutions and practices.

2.7   *Attunement [or* **recursive self-maintenance]**

In the balance of this chapter, I will examine the conditions for the possibility of individual agents intersubjectively sharing what Schatzki calls "conditions of life". If Turner's skeptical arguments prove correct, we should not be able to succeed at securing the individually necessary and jointly sufficient conditions required for sharing. Schatzki's notion of "conditions of life" maps closely to Wittgenstein's much discussed use of the term *Lebensformen* or "forms of life". There is an entire cottage industry in discussing what "forms of life" could mean to Wittgenstein; but for the purposes of this exegesis, we can generally understand the notion of a "form of life" as the socially and culturally established practices or "ways of living" that comprise the "background" upon which the intelligibility of an agent's actions gains purchase.

Wittgenstein addresses the problem of intersubjectivity somewhat circuitously through his investigations into rule-following. He hinges the possibility of following rules on pre-reflective "*agreement*" (*Übereinstimmung*) which works to constrain the

---

[223] Gallagher, S. and Crisafi, A. (2009). Mental Institutions. *Topoi* 28 (1):45-51.

hermeneutic space of understanding amongst language-users[224]. Wittgenstein recognizes

the difficulty of this task in showing how describing one human's behavior must

presuppose an entire comples social background or *Lebensform(en)*:

> How could human behavior be described? Surely only by showing the actions
> of a variety of humans, as they are all mixed up together. Not what *one* man is
> doing *now*, but the whole hurly burly, is the background against which we see
> an action, and it determines our judgment, our concepts and our reactions.[225]

At first blush then, it seems as though one must give an account of how one man is

capable of 'accessing' this background in order to show how agents can come to *agree* in

their judgments, concepts, and reactions to (amongst other aspects of *Lebensform*)

following rules. Given that Anglophone Wittgenstein scholarship is fraught with

interpretive and translational issues, it would be wise to inspect whether the term

'agreement' is a good translation for *Übereinstimmung*. The root *stimmung* translates to

"*mood*", "*disposition*", "*tuning*", or "*morale*". From the contexts in which Wittgenstein

makes use of *übereinstimmen* or *Übereinstimmung*[226], it is clear that "agreement" is more

---

[224] Wittgenstein, Ludwig. (1953/2001) *Philosophical Investigations*. Trans. G.E.M. Anscombe. 3rd
Edition. Blackwell. §224.

[225] Wittgenstein, Ludwig. (1980) *Remarks on the Philosophy of Psychology, vol. 2*, G.H. von Wright and
H. Nyman (eds.), C.G. Luckhardt and M.A.E. Aue (trans.), Oxford: Blackwell. p. 629.

[226] cf. Wittgenstein, Ludwig. (1953/2001) *Philosophical Investigations*. Trans. G.E.M. Anscombe. 3rd
Edition. Blackwell. §224, 241-2, 429, PPF 346-7, 351-2. I've highlighted in boldface the English
analogues of "*übereinstimmen*" and "*Übereinstimmung*":
§224(e). The word "**accord**" and the word "rule" are related to one another; they are cousins. If I teach
anyone the use of the one word, he learns the use of the other with it.
§241(e). "So you are saying that human **agreement** decides what is true and what is false?" - What is
true or false is what human beings say; and it is in their language that human beings **agree**. This is
agreement not in opinions, but rather in form of life.
§242(e). It is not only **agreement** in definitions, but also (odd as it may sound) **agreement** in judgments
that is required for communication by means of language. This seems to abolish logic, but does not do
so. - It is one thing to describe methods of measurement, and another to obtain and state results of
measurement. But what we call "measuring" is in part determined by a certain constancy in results of
measurement.
§429(e). The **agreement**, the harmony, between thought and reality consists in this: that if I say falsely
that something is red, then all the same, it is red that it isn't. And in this: that if I want to explain the
word "red" to someone, in the sentence "That is not red", I do so by pointing to something that is red.

appropriate than "moodiness", "disposition", or "morality".

Interestingly, when it comes to the option of "tuning", the rendering "attunement" is less straightforwardly incorrect. "Attunement" and "agreement" certainly do not mean exactly the same thing (i.e. they do not participate in the same language-games in the same ways), though they are similar in important ways. For example, we can say that one is "attuned to nature"; and this would be quite different than what we mean if we said that one "agreed with nature". However in the case of the latter, there is a similarity between a colloquial meaning of 'agreement' and the notion of attunement. When we say that someone "agrees with nature" we do not usually mean that there are two psyches (the person and Nature) and that they've come to some sort of discursive accord. Rather, the agreement is in fitting or *attuning* one's *mood* to that which jibes harmoniously with the natural environment around her.

Charles Taylor identifies an important, and I think accurate, aspect of the ways in which individuals come to "agree" (or perhaps "attune themselves") with communal practices:

> The background understanding we share, interwoven with our practices and ways of relating, isn't necessarily something we partake in as individuals.

§346(e). Does it make sense to say that people generally **agree** in their judgements of colour? a What would it be like if it were different? - One man would say that a flower was red, which another called blue; and so on. - But with what right could one then call these people's words "red" and "blue" our 'colour-words'? - How would they learn to use these words? And is the language-game which they learn still the one we call the use of 'colour names'? There are evidently differences of degree here.
§347(e). But this consideration must apply to mathematics too. If there weren't complete **agreement**, then human beings wouldn't be learning the technique which we learn either. It would be more or less different from ours, perhaps even up to the point of unrecognizability.
§351(e). There is such a thing as colour-blindness, and there are ways of ascertaining it. There is, in general, complete **agreement** in the colour statements of those who have been diagnosed normal. This characterizes the concept of a colour statement.
§352(e). There is in general no such **agreement** over the question of whether an expression of feeling is genuine or not.

That is, it can be part of the background understanding of a certain practice or meaning that is not mine but ours; and it can indeed be 'ours' in a number of ways: as something intensely shared, which binds a community; or as something quite impersonal, where we act just as 'anyone' does. Bringing in the background allows us to articulate the ways in which our form of agency is nonmonological, in which the seat of certain practices and understandings is precisely not the individual but one of the common spaces between.[227]

Whereas "agreement" suggests an active, conscious, discursive, and thematic normative judgment, "attunement" can engage in normative practices pre-reflectively through the shared background of "common spaces"[228]. The notion of attunement has a particularly beneficial feature when we consider the metaphor of an orchestra tuning itself for a recital. Each instrument enters the cacophony already having been tuned—perhaps for a previous recital. No instrument is too far out of alignment, and if it was the room full of trained musicians could easily identify that fact, and use their developed skills as musicians to resolve the problem. Myriad practices are involved in the activity of bringing an orchestra into tune (the learned, proprioceptive, skilled bodily practices of the musician acquainted with her instrument, the ingrained music-theoretical skill of the musician, ways of arranging and sitting as an orchestral group, the recognition of the conductor as having authority, etc.). The activity of tuning, which is itself a social practice, depends on an open and unbounded structure or "framework"[229], or "background" "web of practices"—all of which are continually and dynamically engaged in normative adjustment[230].

---

[227] Taylor, Charles. (1995) *Philosophical arguments*. Cambridge, MA, Harvard Univ. Press. pp. 76-77.
[228] Dreyfus, Hubert. (1991) *Being In The World*: *A Commentary on Heidegger's Being and Time Division I*. MIT Press. p. 144.
[229] Cf. Wittgenstein, Ludwig. (1953/2001) *Philosophical Investigations*. Trans. G.E.M. Anscombe. 3rd Edition. Blackwell. §240.
[230] The idea of "attunement" presented here coincides with the relatively robust processes of **recursive self-**

Nevertheless one must be careful, as the language of a "shared" or "common" space "between" or amongst intersubjective interlocutors potentially suggests a *consensus*-driven process through which all involved parties come into normative alignment or "attunement". William Lynch (following David Bloor and Henry Collins) writes that the concept of *Lebensform* "introduces a sociological element only at the cost of assuming that social life is consensual at its core and that the limits of consensus define the limits of community"[231] while Norman Malcolm avers that "[Wittgenstein held] that in the absence of a consensus of action, there would be no concept of a rule"[232]. E. F. Thompkins similarly claims that "the meaning of the words on which for example our judgment of colors depends is a matter of communal consensus; of the way people live their lives, not of transcendental truths enshrined in language"[233]. These three accounts are representative of the general tenor of humanist arguments for a socially constructed, Wittgensteinian account of normative force meant to be robust enough to adequately characterize rule-following. But as philosophers of science—such as Sandra Harding, Donna Harraway, as well as David Bloor—have argued, consensus accounts of belonging, while seemingly egalitarian on paper do not a community make.

In this vein, Schatzki does a good job in giving a Wittgensteinian account of

---

maintenance explicated in chapter 3. A recursively self-maintenant system is one that is "densely" coupled to its local environment in feedback processes that allow the system to interact with the environment in such a way as to maintain itself by differentially adjusting both its own homeostatic processes as well as the conditions of the local environment. The local environment is comprised of all the factors to which the system is sensitive (e.g. the set of all affordances and constraints). For Dynamic Embodied Agents (DEAs) this includes physical, biological, psychological, and social (amongst others) factors.

[231] Lynch, William T. (2005) "The Ghost of Wittgenstein: Forms of Life, Scientific Method, and Cultural Critique". *Philosophy of the Social Sciences*, 35:139. p. 146.

[232] Malcolm, Norman. (1995) *Wittgensteinian Themes*.Cornell University Press. p. 169.

[233] Thompkins, E.F. (1990) "A Farewell to Forms of Life". *Philosophy*, Vol. 65, No. 252, p. 190.

"being one of 'us.'" He writes:

> Being one of us means, broadly, that a person speaks and behaves intelligibly to us...this means that what a person says and writes when confronted with novel and unusual situations is immediately understandable to the rest of us. When someone departs from familiar usage, he can make himself understood relatively easy with a succinct explanation of the departure...being one of us implies...that her actions make sense to us in the contexts in which they occur and that we grasp which life conditions they express...when these things are not understood, a person still counts as one of us so long as her behavior becomes intelligible once we learn of reasons for her action or about the contexts in which she acts.[234]

However, Schatzki also recognizes the trouble with this type of account. He acknowledges that even with those nearest and dearest to one another, each individual agent "never fully understands" and is never fully understood by anyone. Crucially, he notes,

> Just where someone begins to become unintelligible and thus no longer one of us, is a contingent and shifting matter...A 'we's internal boundaries are similarly flexible and porous, since the border between sanity and insanity, between those who behave intelligibly and those who, despite [similar upbringings & experiences] do not act and speak like us, is subtle and shifting. Foreignness, sanity, and we-ness are all matters of degree.[235]

Unfortunately he goes on to shrug these concerns off as mundane troubles already pre-reflectively surmounted, and hence they are issues of little importance. Contrarily, I do not think we can be so sanguine. These considerations of radical alterity need to be addressed head-on if humanists such as Schatzki and Taylor are to succeed at maintaining their position. The differences of degree amongst peers within a community may be minute, but this does not erase the substance of those differences—especially as they factor significantly into concrete communicative relationships. The practices constituting

---

[234] Schatzki, Theodore. (1996) *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge. p. 66.
[235] Ibid. pp. 66-67.

the background of intelligibility for a "form of life" are abundant, as are the manifold

ways in which those practices can relate and mutate within supposedly shared communal

spaces, but the differences in degree which comprise the varying bodily encounters of

individual agents with *their* worlds, compounded by the myriad and chaotic continual

shifts in 'attunement' vastly outnumber them, taken jointly.

Without a satisfactory answer to the problems raised here, it is difficult to

motivate an affirmative response to the question of whether the possibility of meaning is

predicated on individual agents *sharing* in a communally-determined form of life whose

normative force naturally applies to all and only humans. One may grant that individual

agents normatively "correct"[236] their actions by employing any number of practices,

habits, rules, or behaviors, etc. to bring themselves into "attunement" with the uses (or

rules) in which they encounter (either pre-reflectively, or consciously) others acting. In

§242 of *PI* Wittgenstein writes, "If language is to be a means of communication there

must be agreement [attunement] not only in definitions but also (queer as this may sound)

in judgments."[237] I think that there is probably agreement (or attunement) in judgments

even without the kind of interactive, transmissive communication critiqued by Turner.

Judgment needn't be shared collectively; rather it can merely show up phenomenally to

dynamic embodied agents *as collectively attuned*. This is all it means for agents to agree,

whether it is in their pre-reflective embodied (body-schematic) practices or in conscious,

---

[236] On the DEA account, this process of "correction" is the recursive self-maintenance of a social system qua social.

[237] Cf. Wittgenstein, Ludwig. (1953/2001) *Philosophical Investigations*. Trans. G.E.M. Anscombe. 3rd Edition. Blackwell. §242.

discursive, linguistic utterances[238].

      While no particular judgments must be understood as collective, other pressing questions remain: is collective sharing of judgments, expressions, emotions, or social practices possible? How are we to understand the *extended*, *dynamic*, and *embodied* aspects of dynamic embodied agents in relation to sharing social practices? What model of communication is entailed or required by this picture of agency, and what are the epistemic bases or consequences of this model? These questions will be addressed in the two remaining chapters.

---

[238] It should be clear by now that I do not distinguish sharply between pre-reflective embodied skills, and reflective embodied skills (like using language). There are many ways in which pre-reflective skills can become reflective, and in which reflective skills become pre-reflective. An analogy to Hubert Dreyfus' discussion of the Heideggerian concepts of "readiness-to-hand" and "presence-at-hand" may be helpful here. The proper use of equipment must be learned. The process by which equipment goes from showing up to Dasein opaquely (not as equipment for Dasein, but as obstinate, or in Dasein's way, keeping Dasein from transparently coping) is precisely a process of embodying a skill—of transforming a reflective engagement into a pre-reflective engagement with that equipment. Likewise, when that equipment fails, or the pre-reflective engagement between Dasein and the equipment breaks down, the relation shifts from transparent, embodied, pre-reflective tool-use to opaque, obstinate, reflective frustration of Dasein's worldly coping.

Chapter 3 – Dynamic Systems Theory

It will behoove us at this point to more fully flesh-out what exactly Dynamic

Systems Theory (DST) is, and is not, and how it has developed into its current

manifestation. DST is born from many interrelated fields of research (such as sociology,

anthropology, communication theory, social philosophy among others) and coalesces as a

mode or method or approach in researching in these various fields. Its approach has

changed somewhat over its relatively short history but the basic idea that motivates DST

is an effort to understand higher order patterns as a function of lower order patterns; to

see more general systems in terms of the interworkings of their dynamic component

parts. I hasten to add that even this extremely general characterization does not pass

without opposition from under the umbrella of DST approaches. To generalize DST in

the way I just did, argue these critics[239], already implies that there is some hierarchy of

system and component; that micro-systemic processes are nested within increasingly

higher-order macro-systemic processes, and that there are clear, perhaps system-

functional divisions defining each component function within the greater system. Indeed

one of the disciplinary origins of DST follows a trajectory beginning with the early

structuralists in social theory such as Emile Durkheim, Marcel Mauss, and Talcott

Parsons. These structuralists advanced theories that relied on the relative modularity and

decomposability of hierarchically structured systems.

---

[239]E.g. Bickhard, 2009.

3.1   The Etiology of DST: Structuralism

A system is *modular* if it is composed of multiple parts; each of which has a function that, counterfactually, it could achieve on its own without being connected to the greater systemic whole. A module contributes its functionality to the system not as a result of its interconnections with other modules but because of its intrinsic functionality. A system is *decomposable* if it can be taken apart or broken down into its various components without erasing the system-independent functionality of those components. The overall system may lose its more general functionality by being decomposed, because the more general system depends upon a particular *configuration of relations* between its components in order to accomplish the functions of the greater system. Thus, built-in to the ideas of modularity and decomposability is a particular sense of *emergent properties*; the undecomposed system exhibits or possesses some functional property beyond those exhibited or possessed by each of its modular components. Emergent properties are those properties that are exhibited in a system at a macro-level, but that are not exhibited by or explanatorily reducible to any of the properties of micro-level components that comprise that system. A modular and decomposable system is one that performs some emergent function or exhibits an emergent property in virtue only of the *configuration* of and *relations* between its components[240]. The analysis or reduction of systems in terms of just modules (or even of their relations) in isolation of their

---

[240] It is important to recognize that such a system in order to be considered a system at all must possess some system-identifying property or properties that identify the organized collection of modules as belonging and excludes that which is not part of the system. Without such an identification, the purported 'collection of modules' cannot be thought of as a system. Thus, all systems – in order to be systems at all – must have some emergent system-identifying property. All systems *qua* system are emergent.

incorporation into the functional whole of the system then works to obscure the wider

system-level emergent properties, and gives an inaccurate view of the system. Finally, a

system that is *hierarchically structured* is one whose component modules are nested

within each other based upon some organizational principle; often based on degree of

functional generality. Thus more specific functional modules are subsumed under more

general functional modules, which are in turn subsumed under even more general

functional modules until the overall system is sufficiently composed[241,242,243].

    This early structuralist picture provided a framework upon which sociological,

anthropological, economic, philosophical, and communicative theories could be

expressed and provided fruitful research programs for these disciplines. Of course, such a

framework carries with it commitments to conceptual dependencies like modularity,

decomposition, and hierarchy, which even in these early stages were questioned[244]. The

stringency of such constraints as modularity, decomposition, and hierarchy to the

framework could be loosened, but not without significant costs. For example, a property

like decomposability could be loosened by treating it as matter of degree of

decomposability rather than in kind; modules could be understood as more or less

---

[241] For a full expression of the structuralist conception of systems see Parsons, T., & Shils, A., (eds) (1976) *Toward a General Theory of Action*, Harvard University Press, Cambridge

[242] For an account of the analysis of the concept of functions in scientific explanations see Cummins, Robert C. (1975). "Functional analysis". *Journal of Philosophy* 72 (November):741-64

[243] M. H. Bickhard (in J. Seibt (ed.), 2003) distinguishes between parts (or modules) of a system having versus serving a function. He argues that etiological models of function construe the order of dependence such that the function that some module "serves" for a system is derivative of that module "having" that function. By contrast, Bickhard's "interactivist" model reverses the order of dependency so that a module's having a particular function is the result of the overall systemic functionality achieved through a module's serving the function it does. Bickhard argues (convincingly) that the etiological approach that underwrites most functionalist / representationalist accounts of the mind (what he calls "encodingism") suffers from an incoherence in the very idea of a function; caused in part by misunderstanding this order of dependency in having and serving functions.

[244] Cf. e.g. the idea of "interpenetration" in Parsons, T., & Shils, A., (eds) (1976) *Toward a General Theory of Action*, Harvard University Press, Cambridge, p. 109

decomposable. They could be fully, or nearly decomposable, or non- or nearly-non-decomposable. However, the introduction of intensity brings with it the drawback that the utility of the decomposability distinction is either weakened or completely compromised. With modularity, the problem is exacerbated because a module is defined by functionally localizable boundaries; but if modularity is a matter of degree, then so are the corresponding boundaries. And if boundaries are not demarcated by clear and distinct differences in kind, we now must face the difficulty of potentially being unable to define modules *at all*[245].

Parsons' structuralist social theory saw individual persons as modular components comprising social entities like institutions, social norms, and practices. However individuals may be somewhat more versatile than dedicated functional modules in that they are capable of taking up or occupying various functions; what Parsons called social roles[246]. Often what role (or "function" as it may be) a person plays within the greater system of, say, a social institution like the university system is dictated by other social roles in which a person finds herself. Thus, according to Parsons what social roles a person occupies, or takes up, is determined by a recursive operation involving the other social roles she already plays[247]. This process is dictated in part by cultural norms, and also in part by the particular circumstances (within that culture) in which the person finds herself. When a person is playing a social role she is modularly fulfilling a social function. There are more general social functions (like being a modern woman, or being

---

[245] This is one rather functionalist characterization of modularity. See later discussion on C. Hooker for additional detail.
[246] Parsons, Talcott (1951) *The Social System*. Routledge, London
[247] Parsons, Talcott (1964) *Social Roles and Personailty*. The Free Press, New York.

an American) and more specific social functions (like being a second-string wide receiver for the Green Bay Packers for the 2012-13 NFL season). What makes persons and their social roles somewhat more interesting is that they are versatile; someone else could play the role of second-string wide receiver for the Green Bay Packers in the 2012-13 NFL season. In a sense, persons are more or less *general* modules in Parsons' conception of social roles. They are relatively interchangeable across a wide variety of possible social roles. More important to Parsons was explaining how social norms, and thus social roles are transmitted among individuals within a society; and more generally how these interactions work to shape the character of the wider social system. He called the process of norm transmission "socialization", but never gave a full explanation of the mechanism(s) through which socialization was meant to occur[248]. This amounts to recognizing the types of relations between modular components in a system, while failing to explain through what media those relations take place.

3.2   *Dynamics, Chaos, and Complexity*

Structuralisms like Parsons' had their heyday during the first half of the 20th century, and laid a considerable part of the conceptual basis for later developments in DST. Concurrent with Parsons' sociological structuralism, other disciplines (including especially mathematics, theoretical physics, economics, biology, ecology, cybernetics, and chemistry, among other disciplines, sub-disciplines, and interdisciplinary areas of research), sometimes independently, sometimes cooperatively developed approaches to the analysis of their respective subject areas that increasingly focused their attentions on

---

[248] Cf. Parsons T., Bales, R. (1955). *Family Socialization and Interaction Process*. Routledge, London

the interactions and complexities of the processes that had plagued their research.

Initially, these approaches drew heavily upon the influx of largely successful linear and

numerical methods of data analysis that dominated wartime and post-WWII engineering

problems; and focused primarily on the conditions of system equilibrium. In 1954

Ludwig von Bertalanffy, Anatol Rapoport, Ralph W. Gerard, and Kenneth Boulding

established the Society for the Advancement of General Systems Theory[249], which sought

to bring into conversation the then disparate systems-based approaches into a unified

scientific program. In his 1968 *General System theory: Foundations, Development,*

Applications Bertalanffy wrote that:

> There exist models, principles, and laws that apply to generalized systems or
> their subclasses, irrespective of their particular kind, the nature of their
> component elements, and the relationships or "forces" between them. It seems
> legitimate to ask for a theory, not of systems of a more or less special kind,
> but of universal principles applying to systems in general. (p. 32)

The "General Systems" framework held that it is fruitful to understand any subject of

investigation as "systems" of relations between components, and that such systems

exhibit common, generalizable and formalizable patterns, behaviors, and properties.

Eventually, however, this attempt to unify the special sciences under one generalized set

of universal principles encountered both political and conceptual obstacles[250].

In the 1970s through the 1990s advancements in computer technology made

---

[249] Later renamed the Society for General Systems Research, then the International Society for the Systems Sciences (ISSS). Other influential early general systems theorists included Ervin Laszlo, William Ross Ashby, Margaret Mead, Gregory Bateson, C. West Churchman, Norbert Wiener, William Ross Ashby, John von Neumann, Heinz von Foerster, Aleksandr Lyapunov, Henri Poincaré, Howard T. Odum, René Thom, among others.

[250] Many of the original contributors to the project felt that it had been co-opted by parties interested in using its approach to justify Cold War political ends. For further discussion see Hull, D.L. 1970. "Systemic Dynamic Social Theory." *Sociological Quarterly*, Vol. 11, Issue 3, pp. 351–363.

modeling systems easier and this in turn revealed some of the infelicities in earlier,

structural systems theories. For instance, each of the assumed properties I listed earlier;

modularity, decomposability, hierarchy, and linearity; were shown to often be

unreasonable or inapt when applied to many of the systems that researchers were

interested in, especially systems involving social interactions[251]. Instead of being

relatively modular, social systems are made of densely and complexly interdependent and

interrelated parts; and those parts weren't so rigidly modular, they were more or less

plastic or malleable. They were dynamic. Further, the composition of social systems

tended not to be a neat, organized hierarchy of particular functions subsumed under more

general functions; rather there were denser, complex interrelations and interdependencies

across various levels of functional generality[252]. In the late 1980s and early 1990s, aided

by still further advancements computing power, researchers in chaos and complexity

theories began modeling rudimentary systems (including social systems) designed with

sensitivities to these complex interdependencies.

     Research in *chaos theory* focused on how relatively simple systems; systems with

relatively few, or simple components; could exhibit unexpected or emergent behavior.

Research in *complexity* focused on how relatively complex systems; systems with very

many, often very different components that relate to each other in highly complex ways;

could exhibit surprisingly regular behavior. In both camps however, researchers tended to

be more interested in the lawlike generalities they could extract from their models than in

---

[251] Cf. e.g. especially Laszlo, Ervin. (1996). The Systems View of the World.Hampton Press, NJ and
Bánáthy, B (1996) Designing Social Systems in a Changing World. Plenum, New York.

[252] One good source for this is Gregory Bateson's 1972 Steps to an Ecology of Mind which outlines the
ways in which social relationships fail to exhibit linearity.

the particular dynamics of specific research domains. As with the General Systems theorists, and somewhat ironically, they were more interested in finding the general laws that govern chaotic and complex behaviors across all domains, rather than being sensitive to the potential peculiarities of the domains they researched. On the one hand, the focus on mathematical laws of complexity revealed the importance of temporality and dynamics. Important systems like social norms, institutions, and practices were no longer seen as static hierarchies composed of relatively isolable modules, but could now be modeled with dynamic relations that themselves change as a function of other relations, and are sensitive to time. Such systems could now be understood as irreducibly *processual*. On the other hand, this insight into "sensitive dependence" should have highlighted the necessity to index particular systems to particular (social, historical, cultural) conditions; but this was eclipsed by the more universalizing tendencies that guided such programs.

Along with a new focus on system dynamics and change came emphases on a collection of interrelated properties that related to the complex and dynamic interworkings of system processes.

Recently[253], C. A. Hooker has compiled a list of the more central of these properties that significantly serves to introduce the fundamental concepts of dynamic systems, chaos, and complexity theories:

> Nonlinear interactions; non-additivity;
> Irreversibility;
> Constraints — holonomic and non-holonomic;
> Equilibria and Stabilities — static and dynamic;

---

[253] Hooker, C. A. (2011). *Philosophy of complex systems*. Oxford, U.K: North Holland.

Amplification; sensitivity to initial conditions;
Finite deterministic unpredictability;
Symmetry breaking; bifurcations; self-organisation; emergence;
Constraints — enabling and coordinated;
Intrinsically global coherence;
Order; organisation;
Modularity; hierarchy;
Path-dependence and historicity;
Constraint duality; super-system formation;
Coordinated spatial and temporal differentiation with functional organization;
Multi-scale and multi-order functional organisation;
Autonomy; adaptation; adaptiveness; learning;
Model specificity/model plurality; model centredness;
Condition-dependent laws.[254]

It will be instructive here to rehearse in brief detail the definitions of some of these terms.

To this end I will follow closely the exposition that Hooker lays out in his "Introduction

to Complex Systems: A"[255]

> **Nonlinearity**: "An interaction is nonlinear for some variable v if the
> interaction force does not vary proportionately to v."

Nonlinearity is perhaps best characterized negatively with respect to linearity. *Linear*

interactions are those that preserve vector additivity and scalar multiplication. Examples

of linear transformations include rotation, reflection, scaling, shearing, squeezing, and

projection. In a general sense, a linear characterization of a system is one that assumes

that any arbitrary local neighborhood of the functions describing a system can be

expressed in the form: $f(x)=ax+b$ (viz. as a line). Nonlinear systems are then those that do

no preserve vector additivity or scalar multiplication, and whose system functions cannot

be expressed in the form $f(x)=ax+b$ for arbitrary local neighborhoods. Nonlinearity will

be discussed in considerable more detail in chapter 4.

---

[254] Ibid. pp. 20-1.
[255] Ibid. pp. 21-40.

Irreversibility: "A process that is reversible can also be run backwards while still satisfying the same laws. Classical dynamics is time-reversible in this sense. Every dynamically possible process running forward is equally possible running backwards. But virtually all real processes are dissipative...so that they cannot be run in reverse. They are inherently open systems that require an influx of resources or energy in order to persist."

Irreversible processes are nonlinear transformations. Linear transformations can be "undone" by reversals because they preserve vector additivity and scalar multiplication. Some consequences of the irreversibility of processes are that they break time symmetry, they make possible emergent behavior and properties, and they allow for novel system states and trajectories.

Constraints: "Constraints on a dynamical process are those limitations on the relationships among its variables that arise from the imposed conditions in which the process takes place."

Dynamic systems are constrained by the conditions of the surrounding environment—usually some other system(s). Dynamic systems are constrained by being differentially coupled to external systems. The influence exerted by conditions external to a system S give the system's **parameters**. If the external conditions are dynamic, then the system's parameters vary accordingly. If the conditions are static (e.g. if the environment is appropriately controlled) then the system's parameters are fixed. Dynamic systems also include **variables** which change depending on the characterization of the system given its (changing) parameters and constraints. Thus, a dynamic system is defined by the (functional) differential relations between the parameters and variables. The functional relations that define a dynamic system are an expression of the system's constraints – *a system just is its constraints*. A dynamic system's **state** or **phase** space is a space where all possible states of the system are represented. Each possible state of the system

corresponds to a unique point in the state space. For dynamic systems differential

equations giving the relations between initial conditions, variables, and parameters

determine the **evolution** of the system, represented by a plotted trajectory in the state

space. Changes to initial conditions, variable, and parameter values each alter the

evolution of the system, and are expressed as different trajectories or evolutions in the

state space.

Hooker differentiates between *limiting* and *enabling* constraints, as well as

holonomic and *non-holonomic* constraints. **Limiting constraints** are the most common

understanding of the concept of constraint – they give the boundary conditions for the

system, they shape which states and trajectories the system can take by prohibiting all

other possibilities and reduce the system's degrees of freedom by limiting dynamical

trajectories to sub-sets of the system's state space. **Enabling constraint** is the reverse

side of the limiting constraint "coin". By giving the boundary conditions for the system,

constraints also positively define the capabilities and possibilities of the system. "By

coordinatedly decreasing degrees of freedom they provide access to dynamical

trajectories inaccessible to the unconstrained system." These two concepts map

respectively to J. J. Gibson's terms "constraint" and "affordance". **Holonomic**

constraints are conservatory rather than dissipative. Constraints are holonomic whenever

a system is isolated or "closed" (rather than open). By comparison, "open" systems that

involve dissipation are called **non-** or **a-holonomic**.

> **Equilibrium**: "Qualitatively, some aspect *A* of a dynamical system is in
> equilibrium if (and only if) there is no net force acting on the *A* aspect of the
> system (its *A* forces are in balance) and there are thus no net 2nd order rates of
> change (accelerations) in *A*."

136

It is worth noting that this definition of equilibrium is rather broad. It characterizes a system as being in equilibrium whenever *net* forces are "in balance". It is an open question here whether "net" is meant synchronically, diachronically, or in both ways. If the "no net forces" condition allows for diachronic variation (depending on the timeframe selected) periodic oscillations (e.g. the motions of a perfect pendulum) could count as in equilibrium. If "net" is meant non-diachronically, then as long as the pendulum has some nonzero angular momentum the motion cannot be considered equilibrious (because measured non-diachronically there would always be some net force of acceleration).

Hooker further distinguishes between *static* and *dynamic* equilibria. **Static equilibrium** is when "the time invariance concerns *state* parameters and variables (A = system state). Static equilibria require no energy input or output to persist, e.g. a crystal at rest." **Dynamic equilibrium** is when "the time invariance concerns *process* parameters and rate variables (A = system processes). Dynamical equilibria typically require an irreversible ordered energy (negentropy) flow to sustain them, together with appropriate waste (degraded or entropic) outputs". Thus dynamic equilibrium requires that the process parameters or rate variables that comprise A be systemically ***open***, while static equilibrium requires that such variables are isolated or systemically ***closed***.

Of course, no system that matters to any agent is systemically closed, so it is reasonable to restrict systems research on the dynamics and relations of open systems. There is then a question about how such systems can achieve or relinquish equilibrium; how systems can become stable or unstable. Here Hooker distinguishes between three possibilities:

137

> **Stability**: "An equilibrium in some aspect *A* is stable, with respect to some class of disturbances (perturbations) *D*, if (and only if) its response to any disturbance from *D* is to soon return near (including exactly) to its original *A* condition under its own dynamical processes and remain there."

> **Instability**: "An equilibrium is unstable to a class *D* of disturbances if it does not return near to its original A condition."

> **Meta-Stability**: "An equilibrium is meta-stable to *D* if it is stable for some disturbances from *D* and unstable for others."

Here, because dynamic systems are open, system stability is subject to outside influence or **perturbation**. Stable systems are those that are relatively unaffected by perturbation; unstable systems are those that are significantly affected by perturbation; and meta-stable systems are sensitively affected by perturbations. However, what needs further explication is the idea of returning near (including exactly) original *A* conditions under its own dynamical processes. This is the basic contour of the concept of an *attractor*:

> **Attractor**: "An attractor is the closed set of states a system repeatedly traverses when at equilibrium."

> **Attractor Basin**: "an attractor basin is the set of states a system can pass through while still returning to its attractor."

Because Hooker has defined the concepts of "equilibrium" and "stability" rather broadly, a system needn't exhibit static or periodic behavior in order to be at equilibrium. As long as a system's trajectories or states remain within an *attractor basin* (even under perturbation), that system will be at a stable, dynamic equilibrium. Hooker also acknowledges that there are different kinds of attractors; mentioning three possibilities:

> **Point Attractor**: "A point to which a system tends; if the system is at that point, it stays there."

> **Cyclic Attractor**: "A periodic orbit. The system tends toward some cyclic or periodic behavior. If it already exhibits that behavior, it continues to do so

indefinitely."

> **Strange Attractor**: "An attractor in which the evolution of the system state space is bounded but aperiodic. Strange attractors are "chaotic" and sensitively dependent on initial conditions."

Point attractors are the kinds of systems traditionally associated with equilibrious behavior. A system with a point attractor dynamically or progressively tends toward *static* equilibrium. Cyclic attractors are those that exhibit periodic behavior (e.g. a perfect pendulum). Importantly, because their dynamics can be expressed by finite polynomials, the behaviors of both point and cyclic attractors are *linearizable* while the behaviors of strange attractors aren't. Strange attractors also exhibit the three properties characteristic of **chaos**:

1. The system's state or phase space is **bounded**.
2. The system's behavior is dynamic and **aperiodic**.
3. The system's evolution over time is **sensitively dependent on initial conditions**.[256]

Strange attractors are bounded by the limits of the attractor basin – the state space for which initial conditions remain on the attractor. However, such boundaries may be diffuse, dynamically complex, or not well-defined (or well-definable). For this reason it may be better to characterize (1) instead as:

1*. The system's state or phase space is **not unbounded**.[257]

To this then, we may add three additional concepts that describe dynamical attractor

---

[256] There has been surprisingly little consensus in the literature on the characteristic properties of chaos. Robert C. Bishop's entry on "Chaos" in the Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/chaos) highlights the issue of lack of definitional consensus well; however his discussion does not address the specific three conditions given above—which; when taken in conjunction as severally necessary and jointly sufficient; I believe circumvents each of the worries he raises.

[257] However, not much for our purposes turns on this.

systems:

> **Attractor Landscape**: "a system's dynamical signature, expressing its dynamical form."

> **Structural Stability**: "A system that remains within a single attractor landscape is structurally stable (= autonomous dynamics in mathematical parlance)."

> **Structural Meta- or In-stability**: "A system that does not remain within a single attractor landscape (= bifurcate in mathematical parlance[258])"

In many systems there are many relevant aspects A_n that relate to and affect each other. These are expressed as functional relationships between the variables and parameters of a system. The higher the number of variables, parameters, and functional relations (interdependencies), the more ***complex*** the system is[259]. A system's attractor landscape designates the conditions under which a system evolves. Often, however, this landscape is itself non-static – it changes along with the evolution of the system. Further, there may be **dynamical coupling** relations that designate how two or more systems affect each other's attractor landscapes. The idea of "structural stability" is equivalent to the notion that a system is not dynamically coupled to any other systems. A structurally meta-stable system is one that is coupled to some other system(s), but whose behavior remains stable through that coupling—essentially re-shaping the attractor landscape to include the influence of the other system(s). A system that is structurally unstable is one whose attractor landscape "collapses". Structurally unstable systems can either re-stabilize, or

---

[258] **Bifurcation**: "A bifurcation occurs when a structural instability in a system leads to a change in its dynamical form, that is, a change in the structure of its attractor landscape." However, these specifics aren't important to the overall picture I am presenting here.

[259] We can think of a complex system as constraints differentially constraining constraints. On the one hand because of the systemic interdependence in complex systems, small changes to one parameter or variable can have wide systemic effects. On the other hand, depending on the particular dynamics and relations of the complex system, such changes to parameters are more likely to be dampened rather than amplified by these interdependencies—though both remain possibilities in principle.

remain unstable. The process of bifurcation (see footnote below) is often understood as a

"period-doubling cascade" of rapid oscillations between structurally stable cyclic

attractors and instability resulting ultimately in unstable-though-bounded systemic

behavior (viz. a strange attractor).

Most highly-complex, coupled, dynamic systems exhibit *sensitive dependence* on

initial conditions.

> **Sensitive Dependence on Initial Conditions (SDIC)**: "small differences in
> system state or trajectory are amplified into large differences in subsequent
> system state or trajectory."

Left unconstrained and unbounded, SDIC produces widely divergent system behavior.

Constrained and bounded however, SDIC deterministically produces *holistically* coherent

patterns of behavior but in-principle unpredictable and random-appearing (though

deterministic) micro-level behavior. Such holistic behavioral patterns themselves may

constitute a macro-level constraint which can function as systemic parameters in

dynamical coupling[260]. This has been expressed in the literature by the term "self-

organization":

> **Self-organization**: "Self-organization occurs when a system bifurcates,
> sufficiently under its own dynamics, to a form exhibiting more ordered and/or
> more complex behavior. Self-organization occurs where (and only where) a
> system bifurcates, sufficiently under its own dynamics, so as to bring to bear
> an additional system-wide constraint (or at any rate an additional multi-
> component, that is, relatively macro, constraint)."

A system that *self-organizes* is one whose constitution is such that by doing what it does

---

[260] For systems that are **constituted** through their dynamical coupling with the systems and conditions in
their local environment (e.g. open dynamic systems), the influences of these [emergent] macro-level
constraints can affect the conditions of the local environment – which can, in turn, differentially
function to change the system's [micro-level] parameters . Thus, through dynamic coupling emergent
macro-level constraints can effectively exhibit **downward causation**.

and by being what it is, it brings about change to continually (re)make itself. This sounds a bit more conceptually complicated than it is. A self-organizing system is dynamic; it changes over time; but it changes in a way that is determined and regulated internally by the system itself; it needs to change in order to maintain its identity as the system that it is. The "additional system-wide constraint" is said to *emerge*, in part because it genuinely constrains system dynamics, but also arises solely out of its "lower level" system dynamics.

> **Emergence**: "the appearance of a phenomenon that could not have been predicted from knowing only the pair-wise dynamical interactions of components."

An instructive and often used example in the literature about emergence and self-organization is the thermodynamically open physical system of a candle flame[261]. In order for this candle flame system to remain a candle flame it must continually combust oxygen from the local environment and exhaust smoke into the environment. The wax must continually undergo thermodynamic and chemical reactions in order to maintain the flame. The candle is "self-organizing" because it is its own source of the dynamic regulation which maintains its identity as a system. This is slightly different than Maturana and Varela's concept of *autopoiesis* which pertains to systems that continually *reproduce* themselves as a mode of propagation. A system can be defined by autopoiesis whenever the processes of reproduction are what make the system what it is; whenever the system gains its identity through its reproduction. Maturana and Varela developed the notion of autopoiesis specifically as a way to explain how biological living systems

---

[261] Bickhard (1993)

work[262].

3.3   Richard Campbell's "A Process-Based Model for Interactive Ontology"[263]

It is in this historical context that dynamical systems theorist Richard Campbell

has recently developed a taxonomy of distinctions that differentiates between different

kinds of living and non-living systems[264]. On his "Interactive Ontology" Campbell

presents a series of binary questions to ask about systems, whose taxonomic classification

depends on responses to these questions. It will be instructive and helpful to further

discussion of DST to consider Campbell's taxonomy in some depth. The first question is

whether the processes involved in the system are persistent, or if they're fleeting. If they

are fleeting then the processes that comprise the system do not endure through significant

changes in the environment surrounding system and as a result the system dissipates. If

the processes are persistent they *do* endure through significant changes in the

environment surrounding the system.

The second question is only asked if the answer to the first question is that the

system's processes are persistent. It queries whether the processes involved in a system

---

262 In their 1980 Autopoesis and Cognition: The Realization of the Living, Maturana and Varela define
    "Autopoesis" as: "a machine organized (defined as a unity) as a network of processes of production
    (transformation and destruction) of components which: (i) through their interactions and
    transformations continuously regenerate and realize the network of processes (relations) that produced
    them; and (ii) constitute it (the machine) as a concrete unity in space in which they (the components)
    exist by specifying the topological domain of its realization as such a network."

263 Campbell, Richard. (2009) "A Process-Based Model for Interactive Ontology", *Synthese*, 166: 3, pp.
    453-477.

264 To be fair, Campbell's approach is found within and amongst a constellation of approaches that together
    have been called "interactivist theory". Some notable contributors to this collection of views include M.
    H. Bickhard, L. Terveen, R. L. Campbell, W. D. Christensen, C. A. Hooker, J-C Buisson, G. Stojanov,
    and S. Bruno among others. Campbell's contribution is in some ways here meant to be taken as
    representative of some of the more broad characteristics of "interactivism", but should not be
    understood as taking itself to be so representative.

are cohesive or not. Campbell defines cohesion as "the internal bonds which constrain the behavior of its constituent sub-processes in such a way that the totality behaves dynamically as an integral whole."[265] If the system is persistent but not cohesive it persists through change and time, but it doesn't function to cohere itself into a totality. A leaky faucet is an example of a persistent but non-cohesive system. It produces dripping behavior, but the dripping behavior doesn't contribute to integrating the dripping system as a whole.

Third is the question as to whether the cohesive system is far-from-equilibrium stable or energy-well stable. Systems that are energy-well stable are those that "persist at or near thermodynamic equilibrium, and whose organization can be disrupted only by an input, from external sources, of a critical level of energy."[266] In a sense, energy-well stable systems don't *do* much of anything. An example of an energy-well stable system might be a rock[267]. One may feel that the phrase "far-from-equilibrium stable" is oxymoronic. How could a system both be far-from-equilibrium and also stable? Doesn't "stable" just mean equilibrious? Campbell defines stability differently, however. The kinds of systems Campbell is interested in are *open* systems; systems that constitutionally interact with their local environments. The local environment is not considered a proper part of the system itself, but with open systems there is no *principled* boundary separating

---

[265] Campbell, p. 462

[266] Campbell, p. 463

[267] It should be noted that energy-well stable systems are not necessarily systems at thermodynamic equilibrium. Energy-well stability is rather a distinction based on relative systemic independence (or closedness) – the degree to which the system is insulated from or affected by the various kinds of perturbations occurring in its local vicinity. A rock is relatively energy-well stable because there isn't much in the local environment that affects it. Naturally, it is also important to recognize that stability is also relative to various timescales. A rock is relatively energy-well stable on a human life timescale, but perhaps fairly active on geological timescales.

the system from the local environment, because the system, defined as open, is constituted in part by the character of its interaction with the environment, and so must to some extent include reference to that environment. Such environments are often dynamic; so open systems tend to require some semblance of robustness or stability in the face of change. If they didn't, they would quickly dissipate (they would fail to persist). Thus, stability here means robustness rather than stasis. So what is a far-from-equilibrium, open, cohesive, persistent system? One that is capable of retaining its cohesive processes in the face of significant changes in the local environment.

The next question is whether far-from-equilibrium stable systems "contribute to the persistence of the conditions upon which they depend, or not"[268]. If a system does, Campbell calls the system "self-maintenant", if not, not. A self-maintenant system is one that utilizes processes in order to change the conditions of the local environment to make it easier for it to maintain its coherence. It imposes a sort of buffer to keep the local environment from diverging from conditions consistent with the system's persistence. In order for a system to be self-maintenant it must be sensitive and responsive to an array of potential and actual environmental conditions.

Campbell's taxonomy continues by distinguishing *recursive* self-maintenance from run-of-the-mill self-maintenance. A recursively self-maintenant system is one that "can maintain stability not only within certain ranges of conditions, but also within certain ranges of changes of conditions. That is, they can switch to deploying different processes depending on conditions they detect in the environment"[269]. Thus, recursively

---

[268] Campbell, p. 465
[269] Ibid.

self-maintenant systems are capable of responding to environmental feedback by selecting amongst multiple processes or strategies for coping / self-maintenance. Campbell takes this criterion to be the hallmark of living systems. It is possible that self-maintenant, far-from-equilibrium, open, cohesive, persistent systems aren't living (e.g. a candle flame is an example of a self-maintenant system), but all *recursively self-maintenant*, far-from-equilibrium, open, cohesive, persistent systems are (Campbell's paradigmatic example here is of a paramecium that differentially responds to a sugar *gradient;* comparative changes in the amount of sugar in the local environment[270]). In a certain sense, the decision to define life based on this criterion is purely stipulative. But as a thought experiment we can imagine a totally alien system that happens to be recursively self-maintenant. We would likely count that system as "living", if not intelligent.

One of the most significant questions, however, for the fields I mentioned at the beginning of this section; sociology, anthropology, communications theory, economics, and social philosophy; is the etiology of *norms*. This, perhaps, is the most significant contribution made by Campbell's taxonomy. After recursive self-maintenance; which itself marks a shift from passive interaction with the environment to actively and differentially shaping the environment to suit the system; Campbell introduces the notion of error-detection. He writes "either recursively self-maintenant systems are able to detect that some action they have performed has been in error, or they do not"[271]. Campbell is careful here to distinguish between the system having the *concept* of error, and being able

---

[270] Cf. Campbell, D. T. (1974, 1990).
[271] Campbell, R.(2009), p. 470

to differentially respond *as if* it recognized an error and corrected for it. It is at this stage

that Campbell introduces what he takes to be the first, naturalistic kernel of normativity.

One cannot understate the significance of Campbell's attempt to give a naturalistic

explanation for normativity; as precisely this issue is one that has plagued philosophers

and social scientists since the inception of naturalism itself.

Campbell uses the example of a frog's attempted consumption of a pebble

(presumably having "mistaken" it for a fly[272]):

> Consider a frog, sitting on its lily-pad, which regularly feeds by flicking its
> tongue at flies and other bugs in its vicinity. If this frog flicks at a pebble
> thrown into the air just above its head it will have done something wrong,
> which can be discovered to be wrong by the frog itself. It will have a surprise;
> or at any rate, will experience some discomfort; if it succeeds in catching that
> pebble with its tongue. Even if the frog should swallow the pebble, it will fail
> to eat it. Once more in a minimal sense (for each step taken in building this
> model should presume no more than necessary), the frog will detect that it is
> in error. The error it discovers, however, will not be anything about pebbles or
> bugs; its discovery will be that this was not, after all, a situation offering
> something good to eat. Its tongue flicking and eating action was not
> appropriate in those circumstances.[273]

What is important here is that the frog's action can be "discovered to be wrong" by the

frog; and this discovery comes in the form of a surprise or discomfort. A surprise is a

betrayal of expectation. In order to be surprised, a system must be capable of having

expectations. What are the conditions for being capable of expectation, though?

Minimally, a system needs to, in some way, be able to 'anticipate' changes that may

occur in its environment. Recursively self-maintenant systems must also possess this

capability. So in this minimal sense expectation is simply a matter of being able to

---

[272] Campbell makes use of this classic "frog and pebble" example deriving from conversations on the relationships between error and biological function made popular by Millikan (1986), Dretske (1988), Dennett (1988), and Fodor (1990). The example itself appears to derive from Lettvin et al. (1959).
[273] Campbell, p. 470-1.

anticipate or extrapolate, near-future conditions given current environmental conditions; and there are many completely naturalistic, sub-conceptual level explanations for how this can be accomplished. Surprise, then, is to have anticipated, or extrapolated incorrectly; for the near-future to fail to pare with the anticipation. So far all I've introduced is a minimal sense of correct and incorrect, which is crucial to any conception of normativity, but it isn't sufficient. What is additionally necessary is the added valancing of "right" and "wrong" (or "better" and "worse"). Campbell essentially suggests here that it is the positive ("right") derives from the negative ("wrong"). What makes an action "wrong" rather than merely in error or mistaken is that there is systemic *malfunction*. The system has some functional need that is left unfulfilled by some sub-process specifically taken by the system to fulfill that function. In the example above, the system (the frog) attempts to employ the process of flicking its tongue toward the self-maintaining function of sustenance. It flicks its tongue for the purpose of eating and becoming nourished, which allows it to continue to carry out processes that allow it, as a [frog] system, to continue to persist. When the frog consumes a pebble and that pebble doesn't nourish the frog, its employment of the flicking process toward the function of sustenance has failed. The more general process of system self-maintenance is frustrated. The system has malfunctioned in its self-maintaining processes. Because the system is *recursively* self-maintaining, it can make use of this feedback by differentially responding. As Campbell says "[The Frog's] discovery will be that this was not, after all, a situation offering something good to eat. Its tongue flicking and eating action was not

148

appropriate in those circumstances."[274]

Is failing to self-maintain "wrong"? Campbell's account flips that question on its head; we can define what it means to be "wrong" in terms of failure at system self-maintenance. What makes a system capable of acting "wrongly" is precisely its ability to recursively and differentially respond to future situations. A system cannot be "wrong" unless it is capable of correcting its behavior; of *learning to control* its circumstances. This is the next distinction that Campbell makes in his taxonomy. Combined, error-detection and learning control are meant to constitute a complete naturalistic basis for normative action. In order to learn, however, the system must have available alternate potential actions, and must have the ability to select from among them. Without this capability there can be no sense of "choosing" appropriate (behavior that supports self-maintenance) or inappropriate (behavior meant to support self-maintenance, but that wouldn't, given the circumstances) behaviors for particular environmental conditions. "All that is needed to explain [a system's action] is that potential actions be indicated to and for the [system] by its detecting relevant differences in its environment, and that its internal processes enable it to select (in some sense) between alternative kinds of action as a result"[275].

It is important here to again acknowledge that nothing like thought is as yet playing any role in Campbell's account. Normative behavior is thus not necessarily discursive or conceptual, but rather can occur pre- or sub-conceptually. A system need not *recognize* or *understand* that it is selecting between alternatives to succeed in doing

---

[274] Campbell, p. 471
[275] Ibid.

so[276]. There are two relevant considerations here: the difference between understanding *why* an alternative is the better of the two and determining *that* one is better than the other; and the difference between ascertaining *discursively that* one alternative is better than the other, and ascertaining *pre-* or *non-discursively that* an alternative is better than another. Campbell clearly sees the normativity that comes from error detection and learning to belong in that last category. Further, he sees the other two categories (discursive knowledge *that*, and understanding *why*) as straightforwardly and naturalistically derivative of the last category[277].

Campbell completes his taxonomy with two final distinctions: reflectivity and social embeddedness. The capacity for reflection marks a distinction from the previous taxonomic class (flexible learning & control) by recognizing the significance of a system's ability to self-reflect. This ability has some important antecedent dependencies; specifically discursive understanding and self-recognition (which both, in turn, rely on symbolic representation[278]). Campbell doesn't explicitly emphasize it here, but as we will see, the suggestion that another important dependency for self-reflection is embeddedness in certain kinds of social contexts is also consistent with his approach. So the next major step after adaptive or flexible learning & control is actually the ability for a system to

---

[276] Again, this is an extremely significant departure from the representationalist tradition. After Frege, the Logical Positivists had established that normativity resides exclusively within language—normativity is essentially discursive, and cannot be otherwise. By showing a biological origin for normative behavior, Campbell extracts normativity from discursivity and places it within the processes of recursively self-maintenant systems capable of learning and error-detection.

[277] Campbell sees representation as a natural phenomenon that emerges through the kinds of systemic relations that arise in certain types of biological systems with the right kinds of relational configurations.

[278] More on this in chapter 4.

thematically conceptualize or represent the world[279]. Campbell doesn't dwell much on this point for a couple of reasons. First, because he wishes to give a naturalistic account of normativity, consciousness, and social institutions that doesn't rely heavily on representations and the metaphysical "baggage" that representationalist approaches have traditionally carried that has made their brand of normativity so difficult to naturalize[280]. Second, he thinks that the notion of self-reflection itself sufficiently captures the relevant dependency for his final taxonomic distinction of participation in social activity.

To explain what he means by self-reflection, Campbell first distinguishes between what he takes to be two varieties of consciousness:

> Primary consciousness, which we share with the animals, is simply a contentful flow, an experiential flow, but the only way that the *qualities* of that experiencing could themselves be experienced is if there is a second level of the overall system that is interactively, contentfully, experiencing the awareness level of experiential flow. We have to do here with *reflection*. Such a meta-level of experiencing has in fact evolved; it is a characteristic feature of humans. And there is no intrinsic reason why such iterations of experiencing should stop at the number two. We humans can be aware that we are conscious of primary experiencing, and so on.[281]

Thus, "primary consciousness" is a kind of perceptual awareness of one's surroundings including one's place in those surroundings, and the various affordances and constraints that a situation avails. By contrast, further meta-level consciousness is awareness of the quality or character of primary consciousness. In order for a system to possess the capability to self-reflect, the system must minimally have second-order consciousness;

---

[279] To be clear, this is not the same thing as self-reflection. Thematic conceptualization is conceiving of experience as being about something. The point here is that reflectivity requires taking one's self as the object of contemplation; and in order to do this, more generally the system must be minimally capable of having objects "show up" or be represented.

[280] Namely the seemingly unbridgeable metaphysical gap between facts and norms that representational epistemic systems seem to entail.

[281] Campbell, R. (2009) p. 474. Cf. also Bickhard, M. (1993) on primary and secondary representations.

the ability to reflect on at least some aspects or qualities of its primary consciousness[282].

This second- or higher-order consciousness should be recognizable as the fundamental

source of phenomenology. It is precisely the ability to attend to the qualities of perception

or experience; how experiences "show up" to the experiencer; rather than the objects of

experience themselves. Through attending to the quality of experience, a system exhibits

the ability to recursively monitor and adapt to itself as its own experiential environment,

rather than merely monitoring and adapting to changes in the local environment. Such a

system can take itself, and its ability to take itself, as the target for self-maintenance.

Through this process of self-reflection, the notion of selfhood is borne-out[283].

   The terminal taxonomic distinction for Campbell is whether the system possesses

the ability to engage in social endeavors such as cultures, social institutions, norms, and

practices. At this stage Campbell dwells a bit on the notions of emergence and downward

causation. Earlier, Campbell defined emergence, stating that "whenever a complex of

processes organizes itself into a new cohesive system by forming internal bonds that

involve nonlinear forces, the resultant entity has emergent properties and powers. The

result is the familiar picture of a multi-layered model of the world as stratified into

different levels, in a micro-to-macro hierarchy"[284]. Central to this definition of

emergence is the notion of nonlinearity, which will be the subject of the next section, so I

would like to table extended remark on that for the time being. However, for current

purposes there are two other things important about this definition of emergence.

---

[282] That is, meta-level consciousness just is secondary consciousness. They are both consciousness of
    consciousness, or consciousness taking itself as the object of its consciousness.
[283] This issue will be dealt-with more thoroughly in chapter 4.
[284] Campbell, p. 463

First is the organization of processes into layers or stratified levels "in a micro-to-macro hierarchy". Emergence, by definition, relies on this hierarchical micro-to-macro taxonomic sorting of the world. The idea of emergence is that the relations between lower-order micro-processes that comprise a higher-order system or process provide "added value" to the system in the form of additional functionality of properties that the higher-order macro-system or process possesses, but which cannot be attributed to any of the individual component lower-order micro-processes[285]. What is important is to recognize that it isn't an intrinsic property of systems that we describe them in this hierarchical micro-to-macro fashion; it is a decision to highlight certain mereological features to describe systems in this way. When we recognize that this stratification of taxonomy is a *methodological* choice, rather than a metaphysical commitment, emergence should no longer appear as *metaphysically* dubious. Emergence is necessarily indexed to this micro-to-macro hierarchical way of describing and defining systems; but there is no principled reason why our descriptions and definitions of systems need to be organized in this way. So hierarchical structuring is more a matter of epistemic constraints on what can constitute acceptable explanations and a product of systems-theoretic methodology than it is a reflection of the metaphysical facts[286].

---

[285] Cf. e.g. Kim, Jaegwon (2006b). "Emergence: Core ideas and issues," *Synthese*, 151(3): 347–354.

[286] Things can get complicated here. The above statement suggests a metaphysical realism that needn't be assumed. If, instead, metaphysics and epistemology are wed to each other (the commitments in each entailing and constraining possibilities in the other's domain) systems theory can be thought of as more than merely a methodological approach but also a commitment to a certain metaphysical picture. The skeptical question about whether there can or cannot be parity mismatch between how things really are, and what we can possibly know (and how) looms large. As a matter of parsimony I have adopted the view that nothing that cannot possibly matter epistemically should matter metaphysically. Metaphysical claims are meaningless unless they can be known in some way (in the broadest possible sense of "know", which includes especially non-conceptual and embodied knowledges). The choice to adopt a systems theoretic view is a methodological choice, and thus constrains the conceptual possibilities for

The second consideration about Campbell's definition of emergence is its relation to the idea of "downward causation". The very idea of downward causation frustrates the *causal* well-ordering of the stratified micro-to-macro levels upon which the definition of emergence depends. "Downward causation" can be understood as an inversion of the micro-to-macro causal ordering. The initial idea of a micro-to-macro causal ordering is the "bottom up" reductionist assumption that higher-order macro-processes are composed out of lower order micro-processes. The idea of emergence maintains the causal well-ordering of the reductionist approach, but frustrates the principle of reducibility to component micro-processes. With the addition of the principle of downward causation "higher-order" macro-processes can causally affect "lower-order" micro-processes. This has the effect of fundamentally undermining the ordering principle that organizes processes into higher and lower functions or processes. In this sense, the concept of "downward causation" is self-undermining because it both definitionally relies on the hierarchical ordering while simultaneously undermining that very ordering[287]. The result of taking both notions of downward causation and emergence seriously is that there can be no principled hierarchical causal ordering of "higher-level" macro-processes, and "lower-level" micro-processes. Instead the field of discourse about the relations that define systems is either flattened out, or made into an utterly alien highly multi-dimensional (and unordered) landscape[288,289].

---

legitimate explanations—it renders what is and is not thinkable within its domain. The above point just suggests that it is a feature of the systems approach that weds emergence to a hierarchical structuring.

[287] Viz. without hierarchical ordering, there is no orientation for the word "downward". Similarly, the concept of "cause" is frustrated by this potential multi-directionality. Causation is an ordered concept, and notion of "downward causation" introduces disorder.

[288] Either all relations are on a par (in the sense that they are not hierarchically ordered, in principle), or

To be clear, neither of these two considerations is discussed by Campbell, but is implicit consequences of his taxonomic account. In a sense, they undermine his account *qua* taxonomy, but give credence to his more general intentions to give a process-based account of important qualitative distinctions in the kinds of systems capable of social interaction. With notions of emergence and downward causation in tow, Campbell now has the conceptual resources required to explain social influence on individuals, and individual influence on society at-large. He can give a naturalistic explanation of *what* exactly society *is* in the language of complex systems comprised of constituent processes, and *how* the persons and societies mutually constitute and constrain each other as systems with certain key attributes (e.g. far-from-equilibrium, recursively self-maintenant, self-reflective, social, etc.).

3.4   *Nonlinearity & Process Ontology*

Some quick accounting is now in order. We are amidst an account of Dynamic Systems Theory which began with the structuralist approaches of Durkheim, Mauss and Parsons. This structuralist approach was seen as too static or equilibrious by later theorists who preferred to emphasize dynamics, change, self-organization, recursive self-maintenance, and nonlinearity. In our discussion of recursive self-maintenance, I took an opportunity to introduce some of the Systems Theory terminology by examining Richard Campbell's taxonomy of "interactive ontology". Thus, I have now discussed all of these

---

each relation is *sui generis* and thus cannot be ordered beyond the particular role(s) they play within a system. The effect is that cause as a global concept is undermined, but may operate locally in particular system relations.

[289] On the other hand, both emergence and downward causation can sustain a *mereological* hierarchy, or well-ordering (thanks due to M. H. Bickhard for this insight).

topics of emphasis with the exception of nonlinearity; to which I now turn. After this section I will return to the narrative with which I began this chapter by discussing more recent developments relevant to the relationships between embodied agents and social practices institutions that have shaped Dynamic Systems Theory.

In a certain sense the concept of nonlinearity is very much like or even equivalent to the concept of emergence. The idea of nonlinearity in complexity- and chaos theories derives from the mathematical study of system models. Linear systems are those systems that are decomposable into linear combinations or relations between linear systems. If a system can be modeled by constructing it exclusively out of linear systems and linear functions relating those systems, that system is linear. If a system cannot be modeled in this fashion, it is said to be nonlinear. In this sense a nonlinear system is one that cannot be reduced to the sum of linear parts while linear systems can be so reduced[290].

Linear functions are those that map one domain to another in a way that produces a definite solution. Some examples of linear functions or mapping are rotations, reflections, scalings, inversions, homomorphisms, injections, surjections, identity mappings, definite integrals, first and second order differential equations, and Laplace and Fourier transforms. Nonlinear functions are those that map one domain to another in a way that doesn't produce a definite solution. Some examples of nonlinear functions or mappings are differential equations with higher than second order variables, and many partial differential equations[291].

---

[290] Such characterizations of linearity and nonlinearity can be found in almost any text on the subject. For a thorough treatment see Jordan, D. W.; Smith, P. (2007). *Nonlinear Ordinary Differential Equations* (fourth ed.). Oxford Univeresity Press.
[291] Cf. Gershenfeld, N. (1999), *The Nature of Mathematical Modeling* (1st ed.), New York: Cambridge

Linear systems satisfy two main constraints: one called variously "superposition" or "additivity" and another called either "homogeneity" or "scalar multiplication". Both of these constraints amount jointly to the requirement that linear systems be closed under the basic operations of arithmetic (addition, subtraction, multiplication, and division). Nonlinear systems are by comparison negatively defined such that one or the other of superposition (additivity) or homogeneity (scalar multiplication) does not hold[292].

Linear systems are much easier to model because their computational time complexity is far shorter than nonlinear systems. Certain classes of nonlinear systems can be shown to be computationally intractable (or near-intractable) and thus extremely difficult to accurately model. In the first half of the 20th century, much of the research in modeling nonlinear equations served engineering attempts at approximating nonlinear systems by linear systems. A good analogy is approximating a continuous smooth curve by depicting it with pixels. For certain purposes such approximations were useful, as long as they satisfied the tolerance constraints of the specific, practical engineering task; however such linearized models, rather than revealing the actual principles at work within a nonlinear system, effectively ignored or concealed what actually takes place. The cost was model accuracy, but the benefit was practical tractability. In certain domains like meteorology there were no benefits; no linearized statistical model of weather prediction was accurate to a minimum acceptable level of tolerance (more than five days forecast with any accuracy), and many theorists began to suspect that some classes of nonlinear systems may, in principle, have no acceptable linearizations. Then in

University Press, New York.
[292] Ibid.

1963 meteorologist Edward Lorenz published a paper called "Deterministic Nonperiodic Flow" in the *Journal of the Atmospheric Sciences* in which he proposed the initial foundations of what is now Chaos Theory.

As noted earlier, Chaos Theory is concerned with modeling complex behaviors in relatively simple systems as they arise from the workings of relatively simple rules. The general reason chaos theoreticians cite for this complexity out of simplicity is a system's sensitive dependence on initial conditions (SDIC), sometimes called "path dependence"[293]. The concept of sensitive dependence suggests that small (and in some cases immeasurably minute) differences in conditions can quickly effect major differences in subsequent evolution, even given relatively simple functional rules. This has the significant conceptual effect of allowing deterministic systems comprised of few simple rules to exhibit *in principle* unpredictable behavior. Weather is a prime and paradigmatic example of chaos because the basic functional constituents (wind speeds and direction, atmospheric moisture and temperature) are quite simple and easily modeled separately, but their combination within one weather system exhibits extreme sensitive dependence. Slight differences in the initial state of a weather system quickly develop into major differences in future states. And because of this, predicting weather is quite difficult.

However, predicting weather is also difficult for another reason. Weather systems *aren't simple*. To improve the accuracy of the model more than just a little bit vastly

---

[293] Such systems are also variously called "non-" or "aholonomic" systems, or are said to exhibit "hysteresis". The general concept is that such systems' current and future states depend in some ways on the past or "history" that brought them to that state. Path dependence is also closely related to time-irreversible processes (e.g. thermodynamic tendency toward entropy) and thus processes that contribute to the emergence of qualitatively novel system states, properties, functions, and capabilities.

increases the complexity of the components involved. So, for example, now let's add the

irregular contour of the Earth's surface; a fairly simple addition; one new parameter; but

a major change to the complexity of the system. Sensitivity to the complexity of real-

world systems is the primary concern of Complexity Theory. On the other hand,

however, complexity theory is also interested in the relatively simple behaviors exhibited

by such complex systems. To keep with the example of weather, we'll add one more

component to the system (to make it just a little more accurate) and then look at some of

the more predictable weather patterns we can see. Let's add the fact that the surface of

the Earth isn't static, it changes in ways that also alter, or *perturb,* the conditions that

produce the weather; and it continually does this throughout, while the model is running

(after all, the changes in the shape of the surface of the earth, including people and

animals moving around, and the effects of weather, plants growing, etc. are also complex

and chaotic systems). Even with this now seemingly hopelessly complex weather system,

we are still capable of detecting certain weather patterns. We can identify and predict the

behavior of hurricanes, nor'easters, jet streams, and derechos with surprising accuracy,

given their complexities.

The relationship between chaos and complexity is a hallmark of nonlinear

systems. The equations that Lorenz and other Chaos Theorists use to model chaotic

systems are usually quite simple. For example, Lorenz famous "strange attractor" is given

by three ordinary differential equations with three parameters and three possible system

states, and time[294]. The result is that, as the system is iterated it traces out a path that

---

[294] $dx/_{dt}\ \sigma(y - x),\ dy/_{dt}\ x(\rho - z) - y,\ dz/_{dt}\ xy - \beta z$; where x, y, and z define the system state, t is time, and σ, ρ, and β are the system parameters. See Figure 1 in chapter 4 for a graphical representation of the Lorenz

never actually *repeats* but exhibits seemingly periodic behavior. The Lorenz system is

chaotic; slight variations to the initial position drastically alter the path taken as the

system is iterated; but it also exhibits the kind of regularity in pattern that complexity

theorists study. A chaotic (strange) attractor is a relatively steady pattern of behavior that

certain chaotic systems reliably tend toward. An attractor isn't in, and never reaches, a

state of (static) *equilibrium*; in fact a constitutive characteristic is that attractors are

decidedly *aperiodic* and *far-from-equilibrium* systems. However, attractors do exhibit

stable and regular behavior. Depending on the degree of abstraction and with what

considerations for specific characteristics that one views an attractor the behavior may

seem utterly regular, seemingly random, or surprisingly emergent. When we consider an

attractor as an aperiodic, far-from-equilibrium, chaotic system the fact that there is any

pattern to be recognized *is* surprising. As the system evolves, characteristic patterns begin

to emerge qualitatively at a macro-level; the patterns *emerge* from the micro-level

dynamics of the attractor system. The pattern isn't a component built in to the parameters

of the attractor, but develops out of the interrelations of the equations, the parameters, the

initial conditions, and the constraints that give the possible system states. It is a higher-

order macro-level property irreducible to the lower-order micro-level component

processes.

Thus, nonlinearity can provide some clues for how discernible properties,

characteristics, or patterns can arise at (or emerge from) a macro-level of the system

without applying to any micro-level component properties, characteristics, or patterns; at

Attractor's state space. Lorenz's system of equations is meant to model patterns of thermal convection
in an enclosed torus with one point-source of heat.

least for complex, chaotic systems; and it seems at least reasonable to presume that embodied agents and social institutions and practices are such (complex and chaotic) systems; though I will be discussing this more thoroughly in due course.

Much of the foregoing has taken for granted the ideas of "patterns", "functions", and "systems" and their ilk. I would like at this point to take a step back and more fully elucidate precisely how I understand such notions, and how I do not; and to do this I will have to briefly discuss the differences between the two fundamental metaphysical approaches of substance- and process ontologies.

The decision between substance and process metaphysics is one of the most basic bifurcations in metaphysics. It is the question of which between being and becoming is essential, and which is accidental (in Aristotelian terms); which is fundamental, and which is derivative. Substance ontology commits to the view that being is essential and change (or becoming) is accidental. So, for example, a gold statue is *essentially* gold because the substance of which it is made does not, or cannot change; and its shape, which can be altered, is *accidental* because it can, or does, change. On the other hand, Process ontology take the alternative approach in committing to the view that becoming is essential and stasis (or being) is fleeting and accidental. So, for example, what is essential about a gold statue is whatever it does, and what significant factors it contributes and roles it plays in the goings-on in which it participates. What it happens to be made of is accidental; a part of the factors and roles it can contribute; but if the substance the statue was made of is altered, without changing the processes the statue

161

was involved with, this change would be of little-to-no consequence to the identity of the processual identity[295,296].

More recent substance ontologies have construed the world as structured by three things: thin particulars, properties, and inherence relations[297]. Thin particulars are the propertyless "objects" or substances that exist in the world[298]. Properties are the characteristics that belong to substances. If thin particulars and properties both exist in the natural world, presumably there needs to be something to relate them to each other. This relation is said to be an inherence relation which is the type of relation an attribute has with its subject; in this case properties are the attributes that *inhere in* particulars.

---

[295] One should be careful in characterizing Aristotle's views in particular. He is often taken to espouse a substance metaphysics—but there are infelicities in the use of the word "substance". For Aristotle substance (*ousia*) is the combination of matter (*hyle*) and form (*morphos*); however, later metaphysicians have tended to refer to only matter as "substance", leaving the issue of Aristotelian form out of the picture. Separating matter from form can be seen as one of the major causes for confusion in subsequent substance metaphysics. The Aristotelian picture of metaphysics includes both substance (in the anachronistic sense) and process (as a matter of substantial form in conjunction with the four causes, actuality, and potentiality). Aristotle's is then not straightforwardly a "substance metaphysics" in the way I have been discussing it.
For a sustained discussion of Aristotle's metaphysics as it pertains to the relationship between contemporary distinctions between substance and process, see Louise-Gill, M. (1989) *Aristotle on Substance: The Paradox of Unity*. Princeton University Press.

[296] The history of substance metaphysics is often characterized as a progression toward materialist naturalism which privileges two of Aristotle's "four causes"—material cause and efficient cause—while deprivileging or eliminating final and formal causes. For Aristotle the four causes functioned as modes of explanation—he saw them as four kinds of acceptable ways to understand the workings of the world. Material cause explains something by expressing what it is made of. Efficient cause explains something by expressing the agent that made it do what it is doing. Final causes explain by identifying the ends intended by some action, and formal cause explains something in virtue of its relational configuration. The materialist naturalism that developed out of the Scientific Revolution removed all agency (and normativity) from the natural world, reducing the modes of acceptable explanation to what is often characterized as "efficient cause". However, Aristotle's efficient cause and the "efficient cause" of the Scientific Revolution are considerably different from each other—particularly in how they treat agency. Further, as a substance materialism, instead of treating material cause as a mode of explanation, matter became reified as a condition for natural existence. By comparison, process metaphysics retains the original Aristotelian concepts of "efficient" and "formal" causes while downplaying the explanatory role of "material" causes and reinterpreting "final" causes in terms of function (often biological function—which since Darwin can be expressed in a non-teleological manner).

[297] The terminology of "thin" and "thick" particulars derives from: Armstrong, D. M. (1978): *Nominalism and Realism: Universals and Scientific Realism, vol.1*, Cambridge University Press, Cambridge.

[298] Cf. e.g. Chisholm, R. (1969): "The observability of the self", *Philosophy and Phenomenological Research*, 31, 7–21.

Together they constitute a "thick particular" or *substance*.

By comparison, process ontologies carve the metaphysical landscape in a considerably different way. According to a process metaphysical approach the world consists of modes of becoming and types of occurrences. Modes of becoming are the various, different, qualitative ways in which the world dynamically unfolds through time. Types of occurrences are abstractions about concrete, particular goings-on that allow subjects to perceive patterns in those concrete situations[299].

While the substance approach construes the world in a way such that what something is is what it is made of, essentially separable from temporal changes it continuously undergoes; the process approach alternatively construes the world such that what something is, is what it does. On this construal, temporality and the dynamics of change are essential to what something is, and cannot meaningfully be separated from it. Put slightly differently, according to the substance approach a thing's *quiddity;* its "whatness", its constitution, what makes it what it is; cannot include the elements of time or change, or change through time; while the process approach incorporates dynamicism (temporality and change) as essential to a thing's *quiddity*.

If we attempt to approach the DST-oriented concepts of "pattern", "function", and "system" by taking a substance-metaphysical tack, we will significantly misconstrue what DST has to offer. Since on a substance-metaphysical approach properties inhere in

---

[299] There are many different approaches within process metaphysics. Alfred North Whitehead's Process and Reality is often identified as the first contemporary, sustained articulation and defense of a process metaphysics. Nicholas Rescher, Johanna Seibt, and Mark Bickhard are also strong proponents for process-based approaches to metaphysics. Further, the Continental tradition is replete with diverse philosophies that establish or depend on what are variously called "event ontologies" or "metaphysics of becoming".

particulars, a substantist concept of "pattern" would cast it as a property or treat it as inherent in the object that possesses that pattern. What makes something an instance or exemplification of a pattern would be the fact that the object under consideration *has,* as a part of what it is, that pattern as a property. If the object that is under consideration is a system, then the system exhibits or exemplifies the pattern because it is constituted by that pattern as an attribute. The system itself is not the pattern; the system is a bare particular in which the pattern inheres. But this is incoherent; what is the system *apart from* the pattern(s) that define(s) it? What more is there to defining a system than the patterns that identify that system as *that system*?

This is a specific example of a more general worry with substance ontologies. If everything that exists is substantial, what is the metaphysical status of properties, or inherence relations? Are they also substances? And further, if something's particular array of *properties* is what makes it what it is; what differentiates it from anything else; then what explanatory need could there possibly be for *substance* (or "thin particulars") at all? Certainly the substance; the thin particularity; of the system does nothing, in itself, to explain anything about that system. Substance does not, and cannot factor into explanation and so cannot contribute to understanding. The general worry is that substance ontology decouples metaphysics from epistemology; and when our modes of knowing are principally separated from what we claim exists, empirical efforts to uncover the nature of reality are doomed.

A similar historical dispute also establishes functionalist approaches in the

164

philosophy of mind. Historically, functionalism has been put forth as an alternative to the unpalatable consequences that result from behaviorism and mind-brain identity theories. Behaviorist theories are meant to rely exclusively on materialist substance metaphysics; on which all there is to the meaning of mental life is content expressed by purely physical behavior. There is no need to posit *sui-generis* mental states[300]. Unfortunately, behaviorism largely failed to surmount the very basic methodological problem of varied and inconsistent interpretations of what behavior specific physical actions express. In essence, the problem for behaviorism was that there was no principled way to say that a particular physical occurrence *means* a particular behavior. In every occurrence of behavior there always lurked the problem of determining from underdetermined evidence what that behavior is (or is not)[301].

As a means of remedying this methodological infelicity, mind-brain identity theorists attempted to index the meaningfulness of behavior to specific events or states in the brain. Thus, mind-brain identity theory was also meant to rely exclusively on materialist substance metaphysics, as mental life can be explained completely by identifying the appropriate (and purely physical) brain events or states. Strictly speaking,

---

[300] Two behaviorist theses can be distinguished here—methodological behaviorism, and metaphysical behaviorism. *Methodological behaviorism* is the comparatively weaker claim that whether or not minds exist (over and above the behavior exhibited by agents) the only thing we have epistemic access to is a subject's behavior. *Metaphysical behaviorism* is the stronger claim that outside overt behavior there is simply nothing that can be called a mind—minds either are just behavior, or minds don't exist. Generally, however, both methodological and metaphysical behaviorism can be treated as one thesis of behaviorism.

[301] U.T. **Place** offers this and other criticisms of behaviorism forcefully in "Is consciousness a brain process?" in: *British Journal of Psychology* 47 (1956), pp. 44–50. Other canonical criticisms of behaviorism have been put forth by Noam **Chomsky** (1971) "The Case Against B. F. Skinner," *New York Review of Books*, 30: 18–24.); **Armstrong**, D.M. (1968), *A Materialist Theory of the Mind*, London: Routledge.; **Chisholm**, R. (1957). *Perceiving*. Ithaca: Cornell University Press.; **Geach**, P. (1957). *Mental Acts*. London: RKP.; and **Putnam**, H. (1963). *Analytical Philosophy*: Second Series. Ronald J. Butler (ed.). Blackwell.

both behaviorism and mind-brain identity theory can be construed as *process* rather than

substance-based positions. This is particularly evident in the use of the language of "brain

events" rather than "brain states". While some commentators were careful to be inclusive

of either of these uses of language (in an attempt to preempt foreclosing the possibility

that a process-based approach may prove superior to the received materialist substance

approach) very few genuinely attempted to make the case for a strongly process-based

stance. Ultimately, among other problems (including issues distinguishing types from

tokens, and an incompatibility with multiple realizability discussed in chapter 1), the

mind-brain identity approach faced the insuperable criticism that it was incapable of

explaining the qualitative character of consciousness, *in principle*[302].

Hence, functionalism was developed as an alternative to both of these major

problems with behaviorism and mind-brain identity theory. Functionalism

reconceptualized the mind in totally new terms. Instead of focusing on the substance that

comprised the mind, or the physical medium through which the mind operates,

functionalism focused on the mental operations; or *functions;* themselves, as the subject

of inquiry[303]. Much like behaviorism and mind-brain identity theories, functionalism *can*

be cashed out in strictly physicalist terms; we can think of functions as the purely

---

[302] Granted, this is also a criticism of behaviorism, and as it's designation as the "hard problem" of consciousness suggests, it hasn't yet been adequately addressed by any theories. Process metaphysicians such as Bickhard (2009) have argued that the "hard problem" is made even more difficult by substance metaphysical assumptions that work to stifle progress on the issue—specifically the assumption that all cognition must operate and traffic in representational encodings; which compounds problems.

[303] Early proponents for what is now identified as "functionalism" were **Turing** A.M. (1950). Computing Machinery and Intelligence. Mind 59 (October):433-60.; **Putnam**, H. (1960). "Minds and Machines", and (1967) "The Nature of Mental States" in Mind, Language, and Reality. Cambridge: Cambridge University Press..; **Fodor**, J. (1968). Psychological Explanation. New York: Random House.; and **Block**, N. and Fodor, J. (1972). "What Psychological States Are Not". Philosophical Review, 81: 159–181.

physical causal operations that occur in the brain. However, if we do this, we should

recognize a couple of things: first, such brain operations cannot be brain *states*; they must

be events or processes. Any causally ordered explanation must necessarily have a

temporal dimension to it. This, secondly, means that while physicalist, functionalism

cannot be a *substance-based* physicalism, it must take a process-based approach to the

physical. Thus, functionalism is the first primarily process-based theory of mind. A

consequence of this processual, physicalist functionalism is that both types and tokens

must be thought of as processual (rather than substantial) kinds. Functional tokens, then,

map to occurrent brain events or processes[304], while functional types map to the systemic

functional *roles* that are enacted by particular tokens[305].

In its most general formulation, functionalism states that what the mind *is* is

exclusively a matter of what it *does*. Mental states are defined by their causal relations to

sensory inputs, other mental states, or behavioral outputs. These causal relations are

functional in that they accept input and produce output. In a more abstract sense, the

relations are functional because they enact manipulations on media; they *do something*.

So in the most abstract sense, it should come as no surprise that functionalism is

---

[304] The phrasing I use here betrays a commitment that I don't hold, but which is held by a vast majority of functionalists; a commitment to methodological individualism or what Andy Clark (2008) has called the "BRAINBOUND" dogma – the commitment that all processes relevant to cognition occur in the brain. A sufficiently enactive, ecological, embodied, and extended account like DEA does not, however, take the surface of the brain to mark a *principled* functionally relevant boundary.

[305] I am not here endorsing a functional token identity theory of the kind expressed and advocated by Davidson (Davidson 1970) and subsequent functionalists. Instead, what I have in mind is that functional roles can be expressed at differing levels of abstraction. One could, if one wanted, define a function so narrowly that it could be identified with particular tokens. This does not seem to be at all what Davidson et. al. have in mind; but there may be reasons based in embodiment and ecological accounts of the mind for doing so. At the end of this chapter I discuss the relationship between defining functions and defining systems.

essentially processual[306].

The concept of a *pattern*, on the other hand, is quite complicated. Patterns are detectable regularities. We may ask of a pattern whether it exists (metaphysically, in itself) or whether pattern detection is, at base, more so a function of epistemic modes of perception. "I see a pattern"; do I see *a pattern*, or do *I see* a pattern? An initial observation is that this dynamic; this interplay in the concept of patterns; transcendentally reveals the close relationship between metaphysics and epistemology. If we were earlier concerned by the way in which it seems that substance ontologies separate metaphysics from epistemology, the notion of pattern then appears at first glance to be more grist for the processualist's mill. If the term "pattern" is to mean something to us, it does so only because metaphysics and epistemology are closely related in a mutually constraining dynamic. One's epistemic commitments have metaphysical consequences, and *vice versa*. For something to be a pattern means, simultaneously, that it is detectable; that it has epistemic import; but it also implies that there is *something* (metaphysical) to be detected. Our detections may be fallible, but the fact that we detect patterns at all indicates that if we are able to say anything about metaphysics it must be consistent with our epistemic ability to detect patterns. Likewise, pattern detection is constrained to what is metaphysically possible.

If we accept the definition of patterns as detectable regularities, there is also an

---

[306] However, perhaps because it is conceptually difficult to escape the entrenched substantist conceptual framework, most functionalists (following Putnam, Fodor, Block, et. al.) instead reify functional kinds as a separate kind of substance from materialism. In this guise, functionalism rather recapitulates problems similar to those faced by Cartesian dualism.

added normative dimension. What makes something a *regularity*? Lurking here are very fundamental questions about how norms work. Robert Brandom tracks a distinction close to the one I have in mind when he discusses the differences between what he calls "*regularism*" and "*regulism*". According to Brandom[307], regularism is an approach to norms in behavior that marks their descriptive regularity of occurrence as their distinguishing feature. So, for example, on a regularist interpretation of normative behavior one detects the descriptive fact that people happen to be acting in a certain way; say, wearing black at funerals; and from that makes an ordinary inductive inference generalizing from these regularities to the conclusion that "one wears black at funerals". This doesn't yet explain normative prescriptions (just because I recognize the functioning of a social norm like wearing black to funerals doesn't yet motivate that I should also wear black to funerals), but it does give an explanation of where norms come from; they are derived through an inductive inferential process on the regularities one detects. However we may ask the further question "how does one detect *these* (rather than *those*) regularities? For instance, how is it that the observer in the previous example about wearing black at funerals identifies *this* as a relevant potential regularity? What is the mechanism by which the regularities that comprise the basis for the inductive inference become identified *as regularities*? To be identified as a regularity is for the pattern to already to have (somehow) arisen as a pattern *of something;* it is to already have precipitated from the multiplicity of experiences into a regularity; and for this we still lack an explanation on the regularist account[308]. The regularist appears not to be able

---

[307]*Making it Explicit* (1994), pp. 27-34

[308] There is a striking resemblance here to Wittgenstein's considerations of rule-following in Philosophical

account for the origins of detectable patterns.

We can understand the regularist approach as comparably more epistemically and empirically based than regulism because it emphasizes the detection of patterns. By comparison the regulist approach appeals to the metaphysical realist because it identifies regularities or patterns as something that exist in the world; independent of pattern-detecting perception. By situating the existence of regularities in the world independent of perception, regulists explain prescriptive normativity where regularists could not. Implicitly we may recognize the bootstrapping norm that one should cohere one's epistemic construction of the world to the facts about the world as they are, independent of our perceptions, whenever possible (viz. one should not intentionally delude one's self). Since on the regulist account norms are perception-independent metaphysical facts (the norms are in the world, not just in our heads), and since one should cohere one's understanding with the perception-independent metaphysical facts, then (by *modus ponens*) one should cohere with extant norms whenever possible. This kind of reasoning justifies why, given the existence of the norm that one wears black at funerals, I should also wear black at funerals. Not doing so would be to deny or delude myself about normative reality.

Where regulism falters is precisely where regularism excels: in explaining where norms come from. Instead, regulism takes the existence of norms as a metaphysical primitive; and thus outside the boundary of what can be explained. Unfortunately, this is nothing more than a crafty rhetorical device for shutting down philosophical demands for

---

Investigations, §201: "This was our paradox: no course of action could be determined by a rule, because every course of action could be made out to accord with the rule."

explanation; what Stephen P. Turner calls an "explanatory stop-gap"[309]. Equally problematic for both regularist and regulist approaches is the issue of deriving what the norm is, given the multiplicity of experiences. On the regulist account there may be an objective, metaphysical, fact-of-the-matter as to what the norms are, but that doesn't help in explaining how (by what means) one can go about identifying them. In order to explain regularity, norms, and patterns, all three of these criteria must be satisfied:

1. Explain where regularities / norms / patterns come from.

2. Explain how regularities / norms / patterns become prescriptively binding.

3. Explain the mechanism by which regularities / norms / patterns are detected by the agents for whom they are binding.

Because they are closely related, a failed explanation in one of these three areas undermines the entire explanatory project. Both the regularist and regulist approaches fail, because each fails in at least one of these areas (regularism fails 2 & 3, regulism fails 1 & 3)[310].

---

[309] An "explanatory stop-gap" is a request to cease inquiry because one detects a looming regress and needs to "bottom-out" explanation somewhere. In this case what is being claimed is that regulism—in its attempt to take norms as metaphysical primitives—is merely begging that we not inquire into the etiology of norms. The problem is that when we're faced with the question "where do norms come from?" it doesn't help explain this question by saying that they're metaphysically primitive. It doesn't make the question go away, and it doesn't explain it—so it isn't helpful.

[310] In *Making it Explicit*, Brandom tries to navigate a middle passage between regularism and regulism. He does this by arguing that all conceptual content is discursive, and that normative knowledge is constituted by the set of "materially good" inferences on propositions that define each concept. These inferences are discursive, though implicit, rather than explicit symbolic representational tokens harbored in the brain. Thus, Brandom avoids the charge of regulism by making normativity procedural and implicit, rather than having explicit normative rules. He avoids regularism by giving an account of what makes norms binding – pragmatic and intersubjectively assessable commitment. For Brandom, social practices are then nexūs of implicit, discursive, norms (assessable as commitments with practical relations to contextual circumstances and consequences). So all social practices are, at least in principle, expressible linguistically—they can all be "made explicit". Because of this, Brandom sees no need for non-conceptual content, and as a result secures a representationalist theory of mind (even though he argues against the need for explicit representational discursive tokens). What matters is representability,

We've already seen one alternative approach that, at least partially, gives an account that is meant to simultaneously explain all three areas: Richard Campbell's interactive ontology. Campbell's interactive ontology is explicitly a process-based ontology meant to sketch a taxonomy of the relevant differences in kinds of systems, where the notion of what constitutes a "system" is to be cashed out in exclusively processual terms. We will be returning to this issue shortly, but first I'd like to briefly discuss why we should be impressed by the adequacy of Campbell's theory in explaining pattern detection and normative behavior.

Campbell explicitly identifies where he believes norms first appear in his taxonomy: error-detection. Systems capable of error-detection are capable of a very basic kind of normativity; even if not yet normative *action*. A system's ability to detect error makes it a normative system because in order to detect error the system must also be able to differentiate between actions that succeed and actions that fail. This success or failure is not something that merely recursively self-maintenant systems are capable of differentially responding to. Campbell's example of a (merely) recursively self-maintenant system is a paramecium. The paramecium can detect the differential gradient of sugars in its immediate environment, and use that information to "decide" whether to swim or tumble. But, unfortunately for the paramecium, it cannot differentiate between nutritive sucrose and non-nutritive saccharin. When the paramecium swims up a saccharin gradient it is incapable of detecting its error. The consequence of this inability to detect error can be dire (e.g. the paramecium fails to self-maintain; it dies), but one

---

not actual representation. The only problem here is that Brandom still lacks a naturalistic etiology of norms (or more generally of discursivity)—that is, his view still falters on the joint questions of 1 & 3.

cannot fault the paramecium for failing in its functioning; as it is not the kind of system for which failure in error detection is a fault. What we can fault the paramecium for is in not possessing adequate or appropriate self-maintenance mechanisms for the particular environments in which it finds itself (being evolutionarily "unfit" for navigating saccharin gradients).

A paramecium-like system that *can* detect error would do so, perhaps, by detecting that the saccharin that it just consumed is failing to nourish. However, this is still not a system that we would say is "fully" normative; because a system capable of detecting an error that it cannot act upon cannot be expected to act normatively. For normative *action*, Campbell requires an additional capacity for flexible learning & control. Using the paramecium example, a system that is capable of flexible learning & control would be able to detect that the saccharin wasn't nourishing the system and it would be capable of differentially employing alternative actions in order to change the situation; it would be able to do something other than swim toward the highest sugar-or-saccharin gradient as a response to its determination that this mechanism is failing in its function to help procure nutritive sucrose (e.g. to tumble from saccharin, but swim toward sucrose). A system capable of flexible learning & control can change its actions based on feedback from the environment.

So how does Campbell's account fare in addressing the three issues above? Let's consider them in order. Campbell explains where regularities, norms, and patterns come from by explaining the properties attributable to the kinds of systems capable of differentially responding regularities, norms, and patterns. Thus, for example, patterns

come from the fact that a system is capable of responding to its environment in a way that rely on pattern detection. A paramecium is capable of detecting the pattern of sugar gradient in its environment because it has a certain bodily configuration of mechanisms that sense sugar levels all around its body, can pair-wise compare those levels, and can differentially alter its motility (swim or tumble) as a function of the output of these processes such that the overall action is recognizable as the paramecium swimming toward the highest sugar concentration. "Sugar gradient" is a pattern *for the paramecium* because the paramecium differentially responds to it. Likewise, we can explain where patterns come from for humans by identifying the regularities that play a functional role in the processes of determining action and behavior. Perhaps Jim decides to punch Frank in the face because (among possible other contributing and constraining factors) he is capable of detecting the difference between a real and fake Rolex watch. Perhaps Jim's brother John, on the other hand, is incapable of such a discrimination. Jim employs some pattern recognition capability, while John doesn't. Jim can use his ability to detect this pattern to affect his actions, John cannot. We can treat pattern recognition in this way, on a system-by-system basis. In many cases the system that is capable of differentially acting based on its ability to detect patterns will be an organismal system (like the paramecium, or like Jim), but there is no principled reason it needs to be. In fact, it is important to the Dynamic Embodied Agency account that the relevant systems can change to encompass greater or fewer constitutive components, or (more generally) different components, or can be social systems comprised of (amongst other elements and relations) individual dynamic embodied agents.

Next, let's consider how Campbell's account handles the issue of explaining how norms can become prescriptively binding; how it is that we can hold some (and not other) systems responsible for their actions, say their actions are good or bad, or that they should have or should not have acted the way they did. Since Campbell explains where regularities, norms, and patterns come from by explaining the properties attributable to the kinds of systems capable of differentially responding to regularities, norms, and patterns; this also contributes to his explanation of how norms can be binding. Certainly a norm cannot bind a system incapable of acting in accordance with that norm. We cannot hold a paramecium that is incapable of differentially responding to the nutritive difference between sucrose and saccharin responsible for failing to differentially respond to the nutritive difference between sucrose and saccharin. There must be minimal parity between the capabilities of a system and the kinds of norms that can hold sway[311]. A paramecium that *is* capable of differentially responding to the nutritive difference between sucrose and saccharin (e.g. it tumbles away from non-nutritive saccharin, it swims toward nutritive sucrose) is bound to doing so by its capabilities, and by its being the kind of system that it is; namely a recursively self-maintenant system. Tumbling away from non-nutritive saccharin is part of the paramecium's recursion as a self-maintaining system. Counterfactually, if it, being the kind of system that it is, didn't tumble away from the saccharin, and as a result failed to self-maintain; this is all that is required to say that it failed as the kind of system that it is supposed to be.

---

[311] One can think of the way I talk about the capabilities of a system as analogous to the way in which Brandom discusses implicit inferable normative propositions. A system's capabilities may not be externally 'explicit' but instead inhere in the kinds of affordances and constraints that can arise as significant for the system's possible actions in a given situation.

Here we run into some difficulties. We may ask: what is the difference between a system that is incapable of error detection and flexible learning, and one that *refuses* or *fails* to adequately detect error or learn flexibly? More generally we may ask: what is the difference between a failed or bad flexible learner & controller and something that simply isn't a flexible learner or controllers? Presumably, we are supposed to be able to hold the former normatively accountable (this is, after all, why we get to use normatively charged language like "failed" and "bad"), while the latter cannot or should not be held normatively responsible (after all, we cannot and should not hold something responsible for that which it is incapable; to do so would be the fault of our judgment and not of the organism's). This kind of argument has seen purchase in somewhat recent philosophical debates about biological function[312], and can also be seen in Wittgenstein's arguments about rule-following. If we don't know which rule someone is following, how can we say they are mistaken in applying the rule? In Wittgenstein's words:

> This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And there would be neither accord nor conflict here (PI, §201).

For his part, and the better part of a half-century's worth of subsequent Wittgenstein scholarship, it is still unclear whether Wittgenstein resolves this paradox, or what his resolution is, if he does. It seems that a system can only fail at a function that it has; and that only the system can "know" what that function is (and in many cases even that system may not be capable of such "knowledge"). This, of course, raises the specter of

---

[312] Cf. e.g. **Cummins**, R. (1975). "Functional Analysis." *Journal of Philosophy* 72: 741.765; **Millikan**, R.G. (1989b). "An ambiguity in the notion of function." *Biology and Philosophy* 4: 172-176; and **Buller**, D. (ed.) (1999*) Function, Selection, and Design*. Albany, NY: SUNY Press.

some canonical problems with theories of meaning in the internalism / externalism debate[313]. Something cannot acquire its meaning from internal sources *only* (e.g. meaning cannot be *solely* agent-relative), because then meaning couldn't transcend the boundaries between agents; and nobody could hold anyone else accountable or responsible; or know what anyone else means when communicating; in fact communication becomes impossible for meaning internalists. Alternatively, if meaning is exclusively external then we face Wittgenstein's paradox; that there is no way to (externally) determine the difference between failure to obey one rule, and success in obeying some other rule. For our purposes, the way out of this problem is to know what the relevant functions and capabilities of a system are; and these, for the dynamic systems theorist, are determined through the complex and chaotic dependencies that shape the concrete dynamics of the particular situations in which open, far-from-equilibrium systems interact. We may not be able to identify these functions and capabilities exactly or absolutely; but the focus that complexity theory brings to bear on extremely complex systems that produce perceptibly and approximately regular behavior may aid in our abilities to confidently (if provisionally) decide the problem.

The third issue that needs to be addressed is explaining the mechanisms by which regularities, norms, or patterns are detected by the agents for whom they operate. Campbell's approach puts much of the explanatory weight here on the taxonomic classifications he gives. Basically, the idea is that as long as a system satisfies the inclusion criteria for a specific taxonomic class, there will be naturalistic mechanisms

---

[313] Cf. e.g. Kornblith, Hilary (ed.) (2001) *Epistemology: Internalism and Externalism*, Blackwell Press.

that enact the relevant regularity, norm, or pattern detection functions that allow for its class inclusion. These mechanisms are; it is important to remember; processes "all the way down". By giving both the taxonomic classes and naturalistic explanations for the transitions between classes within his hierarchy, Campbell provides a blueprint (or to keep with a more processual metaphor, a *procedure* or *algorithm*) through which one can construct, for any specific system, a naturalistic explanation of mechanisms that produce the functions and capabilities that determine its inclusion within a taxonomic class. However, it should also be noted that a blueprint (or procedure, or algorithm) for an explanation is not itself an explanation. Further, we should acknowledge that while his taxonomy is meant to apply *generally* to the kinds of systems with which we are interested, it may not always track the most relevant or appropriate capabilities or functions for such systems *in particular*. The degree to which this is problematic will depend on the degree to which one seeks a more general, or more specific explanation (e.g. do I wish to explain how embodied agents enact social practices *in general*, or do I wish to explain why I am dressing in black, *in particular*, here and now?). Explanations are always indexed to the phenomenon that is in need of explaining; just as problems always proscribe and constrain candidate solutions.

Last for this section is to discuss what is meant by "system", and in particular to understand systems with respect to process-based ontologies; that is, *dynamic* systems. The early structuralists like Durkheim, Mauss, and Parsons established a way of talking about systems that is sensitive to organizational and dynamical complexities. The

178

dynamical systems approaches that developed out of these structuralist considerations; in

conjunction with a shift away from substance and toward process metaphysics; has

coalesced a powerful framework through which phenomena can be explained. The basic

conceptual unit of this framework is the "system". At its most basic, a system is a model

of some phenomena structured as a set of elements or components and their functional

relations to each other and to the system as a whole, that together comprise a whole in

which some property or functionality is achieved[314]. Systems can be either

decomposable; which means that the properties or functionalities of the whole can be

expressed by the collection of the components without maintaining all of their functional

relations to each other, or to the whole; undecomposable; which means that the

functionality of the whole depends on the specific interactions of the component parts

with each other and with the whole; or partially [or nearly] decomposable; such that the

whole can be decomposed into some subset of components, but not completely

decomposable into only basic components (e.g. some components are themselves

subsystems comprised of more components, but the functions and properties of such

subsystems are undecomposable). Systems that are undecomposable either possess

properties or perform functions that are emergent; such properties or functions are a

holistic result of the interworkings of the whole[315].

     It is important to understand systems primarily as *epistemic tools or models*; or

---

[314] There are many definitions of "systems", but they all identify systems as models comprised of components. For a sustained discussion on the concept of systems a particularly good resource is the work of Hungarian systems scientist Béla H. Bánáthy. Specifically, cf. e.g. Bánáthy B.H.. *A Taste of Systemics*, The Primer Project, 2007.

[315] Simon, H. A. (1969). "The architecture of complexity". *The Sciences of the Artificial* (pp. 192-229). Cambridge, MA: MIT Press.

conceptual or perceptual frames rather than metaphysical entities. What makes something a system is its definition; and its definition is a function of how its boundary conditions are defined. The boundary conditions that define a system may be *functional* (the system accepts certain input and produces certain output), *temporal* (the system is bounded in time; it begins, does something, then ends; though it needn't be temporally contiguous; it may exist between moments of non-existence), and/or *spatial* (the system is bounded in space; it has a definite extension; though its extension needn't be contiguous; there may be space between its extended parts). Boundary conditions are defined by epistemic agents; they do not exist as metaphysical entities *independent* of the observers who make use of them[316].

Though systems are defined by their boundaries, their boundaries needn't be (functionally, temporally, or spatially) rigid. Systems can be open or closed; they can be nested within, or overlap with other systems; and they can interchange functions, components, or relations with other systems or the environment. An open system is one whose properties or functions are sensitive to conditions (functionally, temporally, or spatially) surrounding but not strictly included within the boundaries of the system. In a certain sense, what makes a system open is its status as continually open to definitional or boundary revision; an open system has the property of being *provisionally-defined*. A closed system, by contrast, is one that is not sensitive to external stimulus. They are self-contained and have no bearing on or relation to anything beyond their boundaries. They are *absolutely-defined*. A system can vary in its degree of openness or closedness; and it

---

[316] The issues here are fraught. Agents may not have any control over the boundaries they perceive—and the [meta]physical world in a real sense contributes to the possibilities that afford and constrain (or bound) perceptual and conceptual possibilities.

may vary as to whether it is open or closed along different dimensional axes (e.g. a system may be temporally open but spatially closed; or functionally open but temporally closed, etc.)

In an abstract sense, all systems are functional because they all exhibit some property or function as a whole, as a result of the interworkings of their components. The term "interworkings" here may or may not imply a temporal relationship, however. Process models are those whose ontologies are comprised only of systems that include both functional and temporal components. Many process ontologies also include systems with spatial components (for example, this may be a requirement of physicalist theories), but this is neither necessary nor sufficient for a model to count as a process-based model. Because they necessarily include a temporal component, the systems studied under the heading "Dynamic Systems Theory" must be process-based models.

Dynamic Systems Theory (DST) is a general framework rather than a particular research project or methodology. It is an approach that can be applied to any discipline or area of research; but is distinguished by its focus on dynamic systems, which are; as we have just seen; based in a process metaphysics. DST is closely related to General Systems Theory (GST), but is distinguished from it in DST's comparative emphases on temporal change, nonlinearity, and far-from-equilibrium open systems. To date, many DST researchers have adopted the related methodologies of chaos and complexity theories to describe and explain the features, components, relations, and patterns in the systems they investigate. The purpose of this project is to bring these resources to bear in philosophical

181

discussions about the relations between individual agents and social practices and institutions. My contention is that the DST framework has much to offer these discussions in terms of explanatory power and possible empirical research opportunities.

The next chapter will utilize the vocabulary and conceptual resources afforded by the foregoing discussion in application toward the notions of Dynamic Embodied Agents (DEAs) and social institutions and practices discussed in the previous two chapters. The goal of the next two chapters, then, is to flesh-out how understanding DEAs and social practices and institutions as interacting dynamic systems helps to explain what social agents, practices, and institutions are, and how they mutually co-constitute DEAs.

Chapter 4 – Persons as Dynamic Systems

The purpose of this chapter is to bring the conceptual resources developed in Dynamic Systems Theory (DST) in contact with the kinds of agents I described in chapters 1 and 2—what I have called Dynamic Embodied Agents (DEAs). In doing so; and because DEAs are ecological, extended, embodied, embedded, and enactive agents; it is important to understand them in concert with the local environment that necessarily factors into their very constitution. In particular, we are interested in a certain kind of DEA: agents that are capable of social interaction. For such agents, the qualities and characteristics of this social interaction are important constitutive factors that shape and define who and what these agents are. However, this gives only one half of the theoretical picture; and for this picture to emerge as one whole coherent theory it must also address what social practices and institutions are from within the same theoretical framework. DEAs and social practices and institutions must lie on the same ontological continuum. Historically, much of the difficulty in explaining social practices has been the result of a discontinuity in metaphysics and epistemology: theories have used one theoretical framework to explain the workings of individual, autonomous agents; and different theoretical frameworks incompatible and discontinuous with the first to explain social practices and institutions. It is a benefit of the theory of Dynamic Embodied Agency that it can explain how individuals and social practices and institutions interact by using one consistent theoretical vocabulary and explanatory framework.

4.1   *Dynamic Embodied Agency (now with more Systems Theory!)*

At the end of chapter 3, I briefly put together an image of the kind of agents that will interest us in this chapter. In the context of that chapter, whose purpose was to establish the conceptual framework of Dynamic Systems Theory, I noted that some systems are complex, chaotic, dynamic, far-from-equilibrium, and open. Using Richard Campbell's taxonomy, we can further narrow our focus to a subclass of such systems that also exhibit the capabilities of recursive self-maintenance, error-detection, flexible learning, self-reflection (and by extension the ability to represent), and the ability to socially interact. Of course, most humans and possibly even many non-human animals fall into this specialized subclass of systems. It will be important to keep in mind that, on the account of Dynamic Embodied Agency that I have in mind, social practices and institutions *also* fall into this subclass. If we take a moment to unpack and apply these qualities to our uninitiated intuitions about social practices and institutions we will see that this is a substantive and potentially contentious claim (especially when it comes to the ability to self-reflect). I intend to defend this claim later in the chapter, but for current purposes it suffices to simply keep this intention in mind.

First I would like to demonstrate that what is usually understood by terms like "human agent" or "person" is adequately expressed by the conjunction of the features just listed. The notion of personhood is itself controversial and it would help to briefly canvass some of views with respect to this controversy in order to better appreciate the contribution of the Dynamic Systems approach to this debate.

One characteristic often thought to constitute personhood is agency. Agency at its

184

most basic is the idea that an autonomous individual is capable of making decisions and choosing to act based on the outcome of their decisions. Different theorists put forth different explanations of how decisions are made, whether the decision-making process is deterministic or indeterministic, how the decision-making process connects up with action processes, and so on. These are important and significant contributions to action theory, but delving too deeply in to these details would take us too far afield. What are relevant to our discussion are the notions of choice, autonomy, and self-movement.

Another common attribute thought constitutive of personhood is self-awareness. Significant to this view is that persons are the kinds of things that can get a hold of who or what they are as a unity and in so-doing have some sense of self. This sense of self is often thought to be a self-*conception*; but some commentators[317] believe that concepts aren't required to have an awareness of one's self. Many phenomenological accounts require self-awareness as a precondition for self-reflection[318]—and, as a precondition for self-reflection, there is a strong historical thread in phenomenology that requires embeddedness and interaction within a social context of other agents as a condition for the possibility of self-awareness[319]. It is a live philosophical question as to whether one's

---

[317] E. g. Many of those following in the tradition of Merleau-Ponty's phenomenological approach, such as Jose Luis Bermudez (1998, 2001, 2003), Owen Flanagan (1992, 1998, 2007), Peter Poellner (2003), Dan Zahavi (1999, 2002, 2005), Shaun Gallagher (2000, 2005, with A. Meltzoff 1996, with J. Shear 1999, with S. Watson 2004), Alva Noë (2004, 2009), and Mark Rowlands (1999, 2006, 2010), Evan Thompson (2003, 2010, with Francisco Varela & Eleanor Rosch 1991), just to name a few. Opponents such as David Armstrong (1971, 1981) , Peter Carruthers (1996, 2000, 2006), Daniel Dennett (1992, 1997, 2007), Jerry Fodor (1975, 1987, 1996), William Lycan (1987, 1996), Zenon Pylyshyn (1984, 1986), David Rosenthal (1986, 1992, 2002, 2005) [among scores of others] follow broadly neo-Kantian and/or Computationalist approaches, arguing variously that the notion of the self is intrinsically conceptual, discursive, thematic, representational, etc.

[318] Notably Jean-Paul Sartre, though this is a theme that is manifest in most phenomenological thought.

[319] I am thinking here, again, of Sartre's discussions in *Being and Nothingness*, Part III, Chapter 1, section IV: The Look, though as Sartre himself recognizes, similar ideas can be found in Hegel, Husserl, and Heidegger. The issue is a little more complicated with respect to Merleau-Ponty—as in *Phenomenology*

sense of self is more or less fundamental than one's social involvement, though there is

some consensus that social interaction is the *sine qua non* in developing a sense of self[320].

A third characteristic commonly thought necessary for personhood is a robust

temporal sense of one's self: the ability to understand one's self as an entity that

progresses through time. Some believe that notions of past and future are fundamental to

the fully developed temporal sense required for personhood; but others have argued that

there are alternative and perhaps more important or fundamental senses of temporality

that do not make use of such notions as past and future[321]. Crucial to the understanding of

temporality as a basis for personhood is often the idea of personal narrativity—of

understanding's one life as an ongoing unfolding of events moving toward its completion

as a temporal whole[322]. Often implicit in narrative conceptions of personhood are

normative and moral directives to live in such a way as to craft an excellent life narrative.

On the other hand, it is not obvious what aesthetic criteria determine the excellence of a

---

of Perception, (p. 96) Merleau-Ponty seems to suggest that one's initial experience of otherness is most likely of one's self as other; and that others (in the usual sense of the term) derive from this self-othering experience.

[320] Though there is much disputed about character of social interaction. The famous disagreement between Sartre's "confrontation" and Heidegger's *Mitsein* is a conspicuous case-in-point. One rather striking account of the formation and development of other-relations is Annette Baier's "Cartesian Persons" (*Philosophia*, 10:(3-4), 1981) in which she theorizes that one's self concept is fundamentally an inversion of more primary second-personal "you" relations (e. g. being taken as a "you" by significant others such as parents, inverted into a taking of others as a "you", and finally of taking one's self as a "you" for others). Similarly, Martin Buber's account of the phenomenal character of interpersonal relations in *I and Thou* is also significant. From an experimental psychology perspective, Meltzoff's& Moore's now de rigeur 1977 "Imitation of Facial and Manual Gestures by Human Neonates"(*Science*, 198, 75-78) suggests ways in which a formative capacity for social interaction may be innate (as exhibited by neonates' abilities to imitate facial and manual expressions and gestures). Further research in so called "mirror neurons" has suggested biological mechanisms whereby others' affective expressions are simulated by one's own limbic system (cf. e. g. Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4: 23-48.).

[321] This will be discussed more thoroughly in the section on temporality, below.

[322] Cf. Harry Frankfurt (1988), Alasdair McIntyre (1989), Charles Taylor (1989), and Marya Schechtman (1996).

narrative. Further, some theorists following Heidegger base personhood on the notion of futural projection[323]—the ability to understand one's self as engaged in a continual process of making or crafting one's life as a whole by taking a stand (outside one's self in the present) on one's future self.

One last common characteristic considered important to many definitions of personhood is moral responsibility. Certain kinds of entities can be held responsible for their decisions and actions; they can be praised or blamed for them; they have rights and duties with respect to the ways in which they interact with other persons. This characteristic is a relatively "high-level" condition for personhood—that is, it requires one or some of the other characteristics as a condition for its possibility. One cannot be held responsible if one is incapable of decision and action—if one isn't already an agent. As with the criterion of narrartivity, however, determining the relevant and binding norms and moral standards by which to judge responsibility is not a trivial problem[324].

There are many other live considerations in determining what constitutes personhood, but these four aspects—agency, self-awareness, temporality, and moral responsibility—give a good sense of the basic contours of the discussion. Appropriately, Charles Taylor combines all of them in his 1983 essay "The Concept of a Person"[325]:

> Where it is more than simply a synonym for 'human being', 'person' figures primarily in moral and legal discourse. A person is a being with a certain moral status, or a bearer of rights. But underlying the moral status, as its condition, are certain capacities. A person is a being who has a sense of

---

[323]*Ek-stasis*.

[324] For excellent and thorough discussion of these matters see Derek Parfit's 1984 *Reasons and Persons* and 2011 *On What Matters*.

[325] Originally from Taylor, C. (1983) "The Concept of a Person". *Social Theory as Practice, The B. N. Ganguli Memorial Lectures* 1981. Delhi, Oxford University Press. More commonly found in Taylor, C. (1985) *Philosophical Papers: Human Agency and Language*. Cambridge University Press.

self, has a notion of the future and the past, can hold values, make choices; in short, can adopt life-plans. At least, a person must be the kind of being who is in principle capable of all this, however damaged these capacities may be in practice.

Running through all this we can identify a necessary (but not sufficient) condition. A person must be a being with his own point of view on things. The life-plan, the choices, the sense of self must be attributable to him as in some sense their point of origin. A person is a being who can be addressed, and who can reply. [326]

Here Taylor sums up what it means to be a person as an entity capable of temporal experience of self, and occupying a subjective perspective. These in turn contribute to the constitution of persons as moral and legal entities. On Taylor's construction of persons, what matters most is one's ability to commit one's self (including especially one's future self) to things; to "own up" to one's life and life choices. I will return to the notion of self-ownership as a determinant of personhood later; but for now I am more interested in showing how the collection of qualities in dynamic systems that I listed above are consistent with these four conditions for personhood. This, in turn, justifies my using the conceptual resources of DST in discussing the kinds of persons in which most social theorists are interested.

Minimally speaking, virtually all biological organisms are complex, chaotic, dynamic, far-from-equilibrium, and open systems. They are complex because they involve "high-bandwidth" interactions among a huge number of interworking and dynamic parts across wide ranging and changing internal and external parameters. They are chaotic because they are sensitively dependent on particular conditions in their bodies and their environments. There are two important related biological concepts to consider

---

[326] Ibid. pg. 97.

in thinking of biological systems as chaotic: *plasticity* and *robustness*. Plasticity is the idea that a biological system has the capacity to change itself in order to adapt to its changing environment. Conversely, robustness is the idea that a biological system is capable of maintaining its features despite its changing environment. Maturana and Varela's idea of *autopoesis* combines both plasticity and robustness into one concept. Autopoesis is the idea that an organism is plastic, and through this plasticity, exhibits its robustness: an organism maintains its phenotypic (outward, general) features *because* it is capable of adapting to its changing environment. Similarly, organisms are both chaotic and complex: an organism exhibits higher-order (more abstract, emergent) characteristics *through* the complex and chaotic workings of and relations between its components[327].

Organisms are dynamic, far-from-equilibrium, and open, for interrelated reasons. They are dynamic because they are open and far-from-equilibrium, though stable—that is, because they are open and far-from-equilibrium, in order to remain stable, they must do so by dynamically interacting and engaging with their environments. Similarly, it is the fact that the biological system is both dynamic and open that keeps it from reaching equilibrium. And no system that is both dynamic and far-from-equilibrium can be closed. In essence, all biological systems engage in a commerce with their local environments that allows them to recursively self-maintain. On the other hand, not all biological systems have the abilities to detect error, learn flexibly, self-reflect (and by extension represent), or socially interact. Further, there is nothing essentially special about humans that makes them exclusively capable of any of these processes, however tracking these

---

[327] Cf. Maturana, H. & Varela, F. (1980) Autopoesis and Cognition: The Realization of the Living. Kluwer Publishing. pp. 9-11.

processes is sufficient for capturing the four criteria many action theorists take to be constitutive of human personhood.

4.2   Agency

Recall that agency is the idea that autonomous individuals are capable of making decisions and acting based on the outcome of their decisions. In order to adequately capture the notion of agency then, the DST approach must establish the equivalent ideas of autonomy, choice, and action. The very act of defining a system simultaneously determines its autonomy with respect to the surrounding environment. The Greek root of the word "autonomy" literally means "self-governing"[328]. By giving the functional, temporal, and spatial boundary conditions for a system, we give the grounds upon which that system is differentiated from all else that is functionally, temporally, and spatially definable. These grounds may be provisional, but if so such provisonality does not *in principle* undermine the functional, temporal, or spatial boundaries that shape the system; it only makes the system *dynamically* defined. Such boundary conditions also act as the "law" of self-governance. Merely by defining a system, that system is made autonomous—as a definition gives the rules of inclusion and exclusion[329]. Often the idea of autonomy is bundled with independence[330]. However, many action and moral theorists

---

[328] From the Greek "αὐτονομία"; αὐτός (autos) meaning "self", and νόμος (nomos) meaning "law".

[329] However, it is often a nontrivial task to define a dynamic system; in large part because of its dynamicity, but also for the Wittgensteinian reason that some systems may not be well-definable (viz. they are not rule-based, rather they bear "family resemblances".

[330] There are also often further distinctions, as with the concepts of "procedural" and "substantive" independence. These do not substantively affect my discussion here; as the point pertains to independence more generally. What I have in mind regarding independence is the ontological idea that a thing is independent insofar as its constitution does not rely on any external relations or influences. In this way, I am not here discussing intensive independence (independence as a matter of degree—the

have argued that the idea of an autonomous, independent agent is incoherent[331]. DST can

help articulate why this is so. No dynamic, far-from-equilibrium, open system can be

independent, because the constraint of independence is equivalent to the requirement that

a system be *closed*. Since all agents necessarily interact with their environments (a

condition for the possibility of *action*), no agent can be a closed system. Thus no agent

can be completely independent. If there is independence in such a system, it must be with

respect to some specific parameter; it cannot be a general characteristic of the system.

DST adequately captures the basic contours of *choice* through the related concepts

of recursive self-maintenance, error-detection, and flexible learning. The qualitative

phenomenal feeling of what it is like to choose can be thought of as emerging from some

lower-level, physical, deterministic processes that, themselves, are the choosing

processes[332]. We can think of error-detection and flexible learning as the processes of

comparing what the system is capable of taking to be the relevant factors for a particular

decision-making scenario and adjudicating as to which action should be taken. As a

deterministic process, error-detection and flexible learning can be understood as the ways

in which a system is differentially sensitive to its surroundings. It detects some of the

---

notion that something can be more or less independent, or more or less dependent)—rather, what I mean by independence is an all-or-nothing proposition. .

[331] I am thinking here specifically of communitarian and feminist criticisms of autonomy as exaggerating human agents' individuality and independence. Alasdair McIntyre (*After Virtue*), Charles Taylor (*Sources of the Self*), and Michael Sandel (*Liberalism and the Limits of Justice*) express precisely these concerns. Excellent feminist critique along these lines can be found in Benjamin, B. *The Bonds of Love: Psychoanalysis, Feminism, and the Problem of Domination*, New York: Pantheon Books, 1988, 183-224. ; and Harding, Sandra and Merrill B. Hintikka, eds. , *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science*, 2 ed. . Dordrecht: Kluwer Academic Publishers, 2003.

[332] For an exposition of arguments along these lines see Gallagher, S. (2006). "Where's the action?: Epiphenomenalism and the problem of free will". In W. Banks, S. Pockett, and S. Gallagher. *Does Consciousness Cause Behavior? An Investigation of the Nature of Intuition* (109-124). Cambridge, MA: MIT Press.

aspects of its surroundings and through some process of differentiation produces behavior that is recognizable as a decision in how to act.

Action is expressed in DST as the continuously updated negative feedback loops, decision processes, and differentially-sensitive responses enacted through a system's sensitivities to and sites of interaction with the local environment. The local environment informs the state of the system; and the system, sensitive to various kinds of changes in its local environment, responds in a way that affects and changes the local environment. The new state of the environment is then detected by the system, and it responds accordingly. All recursively self-maintenant systems involve such low-level feedback loops[333]. The difference between these low-level systems and higher-level systems like those capable of error-detection and flexible learning is that higher-level systems have multiple potential responses to the same environmental condition(s)[334]. This process of selecting amongst potential alternatives is another sense of what is meant by "choice" according to the DST model.

4.3   *Temporality*

Likewise, because dynamical systems depend on a process-based metaphysics,

---

[333] These "loops" shouldn't be seen as periodic or cyclical; but rather as "attractors"; stable but non-repeating patterns of actions that are sentitive to certain environmental parameters. Through these feedback loops, a system continually refines, and is refined by (1) the state of the local environment (2) its own abilities to detect the state of the environment and (3) its ability to respond. An example of the concept is lifting weights. By lifting weights, I am continually changing or updating the amount of weight I am capable of lifting, my relation to the weights that are to be lifted, and my weight-lifting technique. These changes are sensitively dependent on each other—a change in one effects a change in the others.

[334] Thus the difference between "lower" and "higher" level systems is a difference in degree of complexity measured in number of variables, parameters, and relations (degrees of freedom), not a difference in kind.

and this in-turn relies on taking change and temporality as primitive, the idea of temporality as a hallmark of personhood should be easy for DST to accommodate. However, the kinds of ideas often expressed under the banner of temporality when it comes to discussing persons (and especially within Heideggerian scholarship) do not bear immediate resemblance to the ideas of temporality as they are expressed in the more mathematically-oriented fields in which DST is often brought to bear. It is important, then, to expend a bit of effort in getting these two seemingly disparate senses of temporality to meet up.

Throughout DST's development as an offshoot from General Systems Theory in the 1960s through the 1990s, it has always had a distinctively mathematical flavor. Its main area of concentration has been in developing nonlinear differential equations to model complex and chaotic natural behavior in hopes of explaining (often physical, but sometimes social and economic) systems that have traditionally resisted adequate explanation by more linear methods[335]. In breaking with linear methods of analyzing systems, one particular traditional and formative concept in history of mathematics came into question. Traditionally in modeling systems as a time-series, the x-axis of the coordinate plane was assigned to the regular, linear progression of time. This axis has

---

[335] The word "natural" often doesn't help in these discussions. The kinds of behavior that dynamical systems theory has often been employed to help explain may be excluded by some as sufficiently "natural". Some examples of systems that DST assesses are astronomical systems (like the solar system), meteorological systems, systems that have traditionally been treated by classical mechanics (e. g. the motion of billiard balls on a billiards table), chemical or metabolic processes, protein folding, economics (e. g. the stock and currency markets), traffic patterns (e. g. the formation of traffic jams), the collective behavior of ant colonies, and mathematical phenomena. Usually DST will model such phenomena mathematically, abstracting away from the particularities of these real-world systems, and analyze the mathematical characteristics of the model—possibly inferring conclusions from the model to the real-world system. This process of mathematical abstraction helps in identifying analogues and structural similarities between various systems (e. g. how is a traffic jam like an ant colony? If ant colonies can solve traffic problems, can we employ similar strategies?).

traditionally been called the "independent variable", and since time was assumed to elapse in uniform series (one moment after another, progressing unidirectionally from the past to the future, an ever—though regularly—moving present) not much thought had been given to the *variability* of a time-series. Generally when modeling or plotting some phenomenon as a function of time, the uniform linearity of the passage of time is assumed.

But what happens when one is no longer willing to assume that time is linear? In the wake of Einstein's groundbreaking theories of special and general relativity, theoretical physicists began to grapple with the idea of space-time curvature. Arthur Eddington's May 29th, 1919 solar eclipse experiment showed that the propagation of light is affected by gravity. Later, in 1964 Irwin Shapiro demonstrated that light takes longer to travel deep in a gravity well than it does outside one[336]. Paired with the limiting physical constraint of the speed of light as the only possible time-constant, physicists were forced to face the possibility that since the speed of light is variable due to space curvature, time may also be variable (not strictly constant). Einstein referred to this phenomenon as "gravitational time dilation"[337]. On the other hand, it can be shown that the speed of light in a vacuum is constant relative to any local inertial frame of reference[338]. So whether time is constant or variable, from any observer's standpoint it is effectively constant. Nonetheless, the idea of variation in the passage of time found application in DST models

---

[336] A phenomenon predicted by Einstein in 1908.

[337] Einstein, A. (1907) "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen" ("On the relativity principle and the conclusions drawn from it"), *Jahrbuch der Radioaktivität und Elektronik* 4, 411–462.

[338] The point about the variability of the speed of light is confirmed when considering non-inertial frames of reference. Whether a frame of reference is inertial or non-inertial is a good basis for defining locality (or non-locality).

of physical phenomena. Instead of linear time-series, DST researchers began to model time nonlinearly[339]. There are four basic models (really, spatial metaphors) of nonlinear time; what I will call cyclic, dilation, intensive (serial), and meshwork (parallel). However it will help to contrast these models with the "received", linear view(s) of time.

The Modern philosophy and mathematics of Descartes, Galileo, Leibniz, Newton, and Laplace, (among myriad others) firmly entrenched the view that time happens to occur in a linear progression from past to future, but that the laws of physics do not distinguish which is which. This is the idea that time is linear but symmetric. The laws of physics would apply whether time elapsed from past to future (as it happens to) or from future to past (as it happens not to). Additionally, time is taken to be uniform, or homogeneous: all moments of time, considered without respect to any other qualities or quantities, are identical to all other moments of time. There is no way to differentiate one moment, considered in its bare temporality, from another[340]. It doesn't matter whether one attempts to roll a ball down a slope under specific physical conditions one million or one trillion years after the big bang; as long as the conditions are *physically* identical, the ball will roll in an identical fashion[341].

In 1927, Arthur Eddington proposed the idea of time *a*symmetry. Time could still be understood as elapsing in a uniform, linear manner; and time could still be understood

---

[339] Cf. e. g. Tong (1990), Granger and Teräsvirta (1993), Franses and van Dijk (2000), and Kim and Nelson (1999).

[340] This is a consequence of the indistinguishability of indiscernables (or "Leibniz's Law").

[341] In philosophy, the primary touchstone in the analysis of time is J. M. E. McTaggart's 1908 The Unreality of Time. In it, McTaggart assesses three "series" or theories of time. He shows that the first two series, A and B, require the reality of time but that their doing so ultimately makes them incoherent as theoretical position. He then offers his C series which doesn't require time's reality, and he proposes that it functions as an adequate alternative. All three time series assume a linear temporal ordering. For this reason, I do not engage directly with McTaggart's work, or the subsequent scholarship that has since grown around it.

as homogeneous; but now the laws of physics wouldn't work the same in either temporal direction; e.g. holding the natural laws constant, if the universe were "run" in the opposite temporal direction (from future to past) the events that unfolded wouldn't pair identically with the way they do when the universe is run from past to future. The idea of time asymmetry has been termed "the arrow of time", because it runs unidirectionally. Both symmetric and asymmetric conceptions of time can be considered linear, as long as the direction in which time elapses doesn't reverse[342].

The weakest sense of nonlinear time, then, might be the idea of the Big Crunch or Big Bounce. *The Big Crunch* supposes that time is symmetric and that at some point the cosmological expansion of the Big Bang will cease and reverse. It is theorized that at this moment time (which is indexed to space, as space-time) will reverse directions and elapse from future to past. *The Big Bounce* elaborates on this idea, suggesting that eventually the universe will return to the moment of the conditions of the Big Bang, and reverse directions again; and this Bang-and-Crunch process will be repeated infinitely[343]. The Big Bounce is a nearly linear, cyclical notion of time. It can be represented one-dimensionally by tracing a continuous path smoothly oscillating back and forth over a line segment.

Another form of cyclical time can be represented two dimensionally by tracing a continuous path along a circle. Cyclical time of this variety continuously returns to its previous states (continuously traces over its previous path) but not by reversing

---

[342] For an excellent elaboration on the idea and history of the concept of the "arrow of time" see: Price, Huw (1996). *Time's Arrow & Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press.
[343] Cf. Bojowald, Martin (2007). "What happened before the Big Bang?". Nature Physics 3 (8): 523–525.

directions. In his famous "Eternal Return" thought experiment, Friedrich Nietzsche conceived of cyclical time as a way of expressing a moral imperative to commit one's self fully to every choice and moment of one's life as if one had to live that moment again and again for eternity. Cyclical time of this variety is, again, weakly nonlinear (or nearly linear) because it is confined to a strictly periodic temporal trajectory. One can, however, entertain a variation on the concept of cyclical time that is more strongly nonlinear by applying the idea of a chaotic *attractor*. An attractor is an expression of a system in "phase space"—the set of possible "locations" a system can occupy based on the possible values it can take for each of its variables and parameters—as it iteratively progresses or "evolves". Chaotic attractors are semi-periodic, which means their motions are aperiodic but exhibit a noticeably periodic-seeming pattern from which they never deviate. On this view events do not repeat, but they bear noticeable resemblance to past events. With cyclical, and semi-periodic conceptions of time, time is nonlinear and unidirectional, but can be uniform or non-uniform, and homogeneous or non-homogeneous. Periodic time like the Big Bounce is very weakly nonlinear, can be uniform (or non-uniform), and homogeneous (or non-homogeneous), but not unidirectional because it periodically changes temporal direction (this periodic time reversal is the only thing that makes periodic time nonlinear).
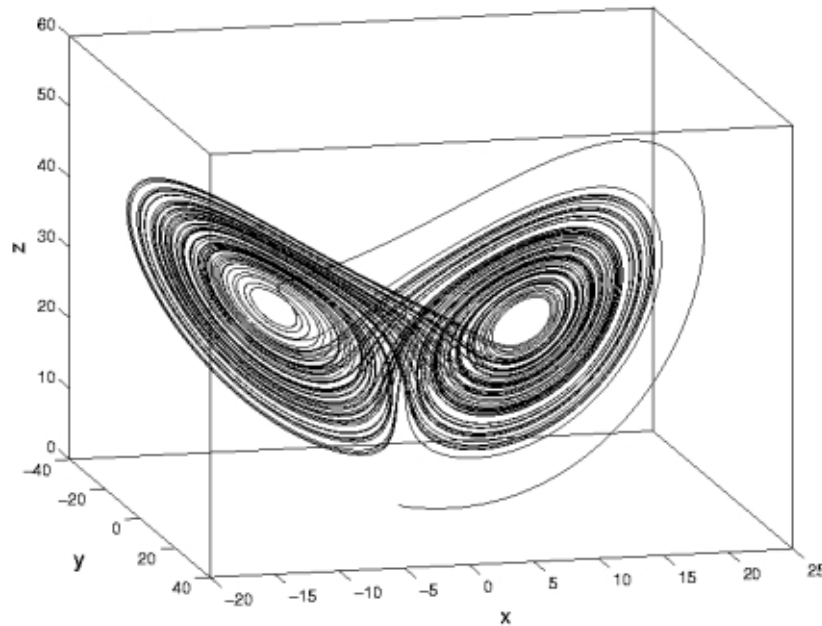
Figure 1: A view of the Lorenz Attractor in 3-D phase space.
Courtesy http://complex.upf.es/~josep/Chaos.html (accessed 12/29/2012).

The second model or metaphor for nonlinear time is *dilation*. The basic spatial

metaphor of time as dilation is that it expands outward in all directions from a center. The

rate of expansion may be uniform or non-uniform, while the quality of temporal moments

must be non-homogeneous because each moment (understood by analogy to the

continuously dilating outer boundary) contains within it all previous moments (e.g. the

qualitative character of each moment is unique because it incorporates into its makeup

previously unincorporated moments). One way to think about time as dilation (or

"dilationary time") is as a continual update, alteration, or revision to the character of

time. On this view, what time is—how time is constituted—itself changes as time

continues to expand or dilate. It is nonlinear because it does not expand out in one

temporal direction. To use the metaphor of a "time line"—a line representing time

extending infinitely to the left and right, with the left representing the past, the right

representing the future, and a point representing the present—part of the idea of

dilationary time is that it expands simultaneously out from the present (center) in *both* past and future directions. The past develops in concert with the future in an essentially dynamic (or processual) and eternal unfolding of the present. The model of time as dilation also has the capacity to accommodate many simultaneous temporal dimensions[344] at once. One way to envision this on the visual metaphor of a dilating horizon is to run multiple time lines through the present center. These multiple time lines represent the unfolding of various distinct events as they occur "on their own time". However, such time lines needn't maintain their boundaries—they may bleed into each other—representing the possibility that the boundaries defining some events aren't so rigid or clear. Just as events bleed into each other, so may time lines on the dilation model.

By comparison the *intensive* model of time is relatively simple. One can take the received symmetric or asymmetric view and alter it slightly by removing the constraints of unidirectionality, uniformity, and homogeneity. The intensive model naturally constructs time as *intensive*—varying in degree. There are a couple of ways that time can vary in degree. Time can elapse "forward" or "backward" along the time line at varying speeds. It can move with a relatively constant "velocity" or it can accelerate or decelerate, or alternate between relatively uniform and non-uniform spans. Time can possess more of some qualitative temporal feature at one moment, and less at another. The basic point of the intensive model is to problematize the assumption of temporal regularity or uniformity.

---

[344] To conceptualize the idea of multiple temporal dimensions, think of various events unfolding on their own timescales, concurrently.

Finally, the *meshwork* model of temporality can be thought of as an elaboration of the intensive model. If the intensive model maintains the basic, linear, serial structure of the received time line, the meshwork model doesn't. As with the dilation model of time, the meshwork model allows for multiple events to unfold on their own time and in accordance to their own rhythms. Such events may interact intensively or relate with each other and in complex, nonlinear ways. One simple way to imagine meshwork time is by considering the particular moment at which you read this sentence. At once, processes are occurring according to their own durations. The universe is expanding on cosmological time. The Earth is unfolding and going to the process of geological change on geological time. The human species is evolving on evolutionary time. Your life is playing out at its own rhythm. We're living in the contemporary era. The U. S. Constitution is developing in concert with the norms upheld by the continually changing set of people that constitute American society. These are all different temporalities unfolding on their own times, but there are also ways in which they can meet up and interact. Temporalities existing in parallel with one another can causally affect each other. These interactions can occasionally be deterministically predictable, happen purely by chance, or be sensitively dependent on the particular conditions in which they occur. In essence, the meshwork approach models temporality as multiple overlapping timescales varying in multiple interacting dimensions of qualities, intensities and durations. It is a highly nonlinear model of temporality[345].

---

[345] These four nonlinear models of time rely on an analogy to developments in the intellectual history of space—particularly in Mathematics from linear Euclidean geometrical spaces to non-Euclidean geometries; in Physics from flat or rigid space and time to curved spacetime; in many engineering fields from linear approximations to nonlinear systems of differential equations. In philosophy, there is no
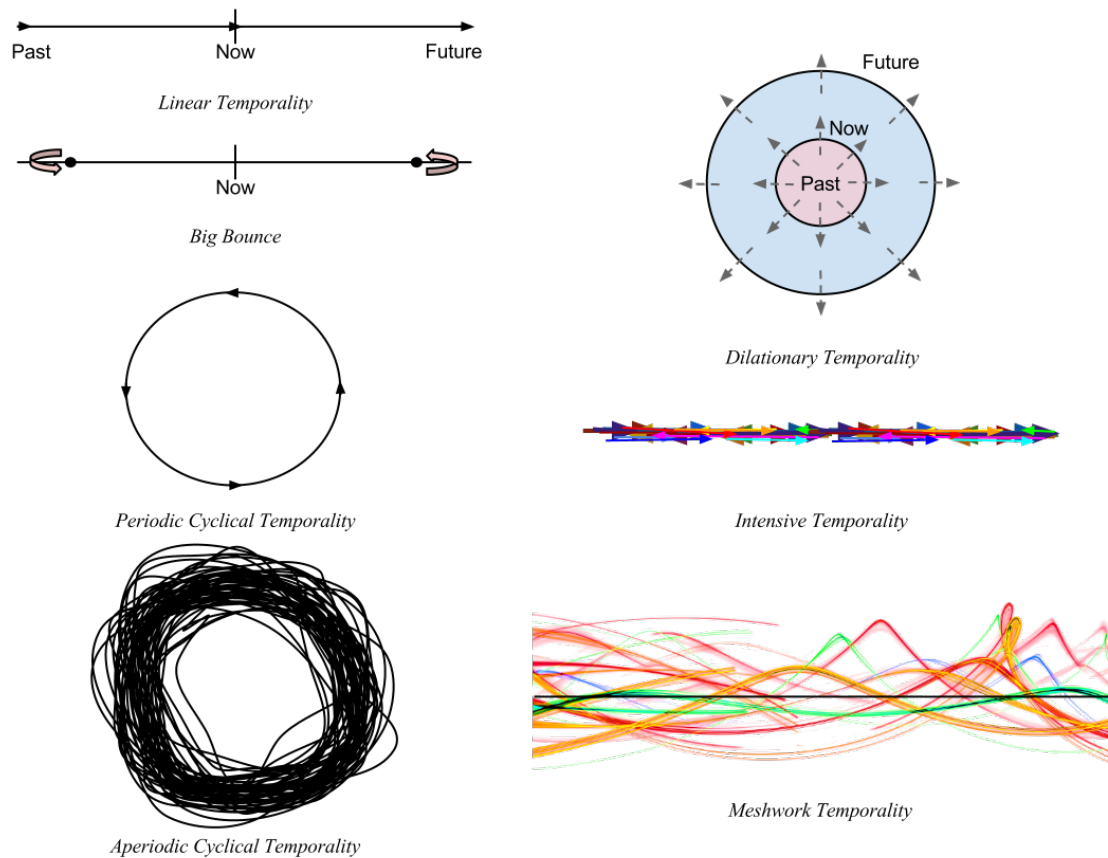
Figure 2: Seven graphical representations of models of temporality.

4.4   Some Quick Accounting

Let us take stock of the argument thus far. We are in the middle of an argument

meant to justify the aptness of DST in expressing four criteria thought to be constitutive

sustained apt analogue from linear to nonlinear conceptions of time. As briefly mentioned in a previous footnote, the philosophy of time has focused almost exclusively on linear ordinal or cardinal orderings. There are scant attempts to conceive of time nonlinearly outside highly abstract mathematical domains. In philosophy one may look to Henri Bergson's *Time and Free Will*, Edmund Husserl's *Phenomenology of Internal Time Consciousness*, Martin Heidegger's *Being and Time*, Gilles Deleuze's *Difference and Repetition*, and Roland Barthes' *The Empire of Signs*. Manuel De Landa draws upon all of these sources the accounts of temporality he outlines in his 1997 *A Thousand Years of Nonlinear History* and 2002 *Intensive Science and Virtual Philosophy*. My exposition of the cyclic model of time draws primarily from a reading of Deleuze's *Nietzsche and Philosophy*. My account of dilationary temporality is influenced in large part by Deleuze's discussion of Bergson in *Bergsonism*, and *A Thousand Plateaus*. The intensive and meshwork concepts of temporality owe their inspiration to the conceptions of temporality De Landa sketches in *Intensive Science and Virtual Philosophy*, and the "bricolagic" temporality of *A Thousand Years of Nonlinear History*, respectively.

of personhood. This, in turn, is intended to justify the appropriateness of developing an account of agency based on DST with the purpose of explaining the relationship between agents as persons and social practices and institutions. We have already looked at how DST handles the personhood criterion of agency and the associated concepts of autonomy, choice, and action. Currently, I have just completed describing five (or so) theories of temporality from a DST perspective—(1) the received linear symmetric/asymmetric views, (2) the periodic/cyclic/semi-periodic attractor views, (3) various versions of dilation, and (4) the intensive and (5) meshwork views. What remains is to show how at least one of these views is adequate in capturing the contours of the temporal criterion for personhood (the criterion that persons must exhibit the ability to understand themselves as an entities that progress through time; or to plan and understand their lives as narratives). This accomplished, I must then show how DST constructs the concepts of moral responsibility and self-awareness—both thought to be conditions for personhood. We will then be in a position to see how Dynamic Embodied Agents, as well as social practices and institutions—each understood as dynamic systems—interact with and relate to each other co-constitutively.

4.5   *The Temporality Criterion for Personhood (Narrative Unity)*

The temporality criterion for personhood is the supposition that in order to count as a person something must be capable of grasping itself as a narrative unity. The term "grasping" is used here intentionally to beg-off commitment to *conceptual* understanding one's self as a narrative unity, at least for the present. Here, "to grasp" doesn't mean to conceive, but rather to have a kind of perceptual awareness, in this case of one's self as a

being whose life takes on a narrative form. At its simplest narrative form, as it applies to one's life, is the temporal sense that one progresses through life starting from a beginning, developing through the course of one's life and ultimately culminating in some meaningful way in one's end[346]. Necessary to the concept of narrative structure is temporal boundedness: one is fundamentally bounded by one's beginning and one's end. One is born, and one dies. Considered temporally, every narrative must have an origin and a terminus which establishes a scaffolding of intelligibility onto which experiences can be grafted. Further, narrative unity binds together a life into a singular whole; it makes the progression from birth to death coherent and meaningful.

According to Charles Taylor, it is narrative unity that signifies the necessary (and perhaps sufficient) condition for inclusion in personhood:

> Agents are beings for whom things matter, who are subjects of significance. This is what gives them a point of view on the world. But what distinguishes *persons* from other agents is not strategic power, that is, the capacity to deal with the same matter of concern more effectively. Once one focuses on the *significance* of things for agents, then what springs to view is that persons have qualitatively different concerns. . . The essence of evaluation no longer consists in assessment in the light of fixed goals, but also and even more in the sensitivity to certain standards, those involved in the peculiarly human *goals*. . . *openness* to certain matters of significance [is] now what is essential to personal agency. (Taylor 104-5, emphasis added)

What Taylor is concerned with here is what he takes to be the "peculiarly human" capacity to have one's life as a unity matter to one's self. He thinks this capacity is *sui*

---

[346] For just one example Jerome Bruner's functionalist approach to narrativity views narrative as the ways in which persons construct reality in order to make sense of their lives. This construction isn't independent, however, of social and cultural influences. Instead, narrative meanings are created and shared by the members who co-participate in a society—that is, narrative meaningfulness is determined socially. This relationship between social interaction and sense-making has a dual-effect: because meaning is made socially through narrative construction, narrative construction is reinforced as that-through-which something can be signified as meaningful—counterpositively: if it isn't expressed in a way that is socially expressible as a narrative, it cannot signify meaningfully. Cf. Bruner, Jerome (1991). "The Narrative Construction of Reality". *Critical Inquiry* 18 (1): 1–21.

generis as compared with other, lower level capacities for things (in general) to be in some way significant to an agent. A person is distinct in his capability to *care* for, and about, his life—how things stand with and are going for himself. Through this attitude of caring, persons recognize and structure their lives around their finitude—their grasping that they are temporally bounded; that they will die.

The notion of personhood that Taylor develops in "The Concept of a Person" is modeled on Martin Heidegger's concept "Dasein"[347]. Heidegger calls this comporting of one's self to one's finitude "Being-toward-death" (*Sein-zum-Tode*), and this attitude of "being-toward-death" works to (re)cast the significance of current concerns by putting them in the context of caring for one's life as a whole—of wanting for one's self for things to come-together into an excellent life narrative. Qualitatively, this is enacted as a "futural projection" of one's self outside the present (*ek-stasis*); a temporal thrusting ahead of one's self into the future. For Heidegger, futural projection creates a tension or struggle (*polemos*) between the way things are in the present (*facticity, Faktizität*) and this ek-static futural projecting of Dasein out onto the wholeness of its life (including one's potential future). This struggle is expressed as existential angst, and as vacillation between inauthentic immersion in occurrent (*ontic*) concerns and an authentic transcendent attention to what he takes to be the more significant (*ontological*) question of the character and quality of one's being in general (*ousia*)[348].

---

[347] I do not wish to delve too deeply into Heideggerian exegesis. For our purposes it is sufficient to understand "Dasein" as "Das Dasein ist ein Seiendes, das nicht nur unter anderem Seienden vorkommt. Es ist vielmehr dadurch ontisch ausgezeichnet, daß es diesem Seienden in seinem Sein um dieses Sein selbst geht. " (Sein und Zeit §4, p. 12) ("that entity which in its Beinghas this very Being as an issue" (Being and Time, trans. Macquerrie & Robinson, p. 68).

[348] Cf. Guignon, C. (2012). "Becoming a person: Hermeneutic phenomenology's contribution", *New Ideas*

If there is anything that is clear about the Heideggerian notion of "futural projection", it is that it relies on a conception of temporality much different than the "received", linear understanding of time[349]. The kind of care for and about the significance of one's life as a whole that interests Taylor (following Heidegger) must instead derive from a nonlinear understanding of temporality. Dasein is a phenomenal locus of experience with the specific capacity to make its own existence an issue for it— that is, Dasein is defined by its ability to attend to its own self-caring. For Heidegger, the temporal moment of the present (now) always contains within it both the past and the future. Temporality structures, and is structured by Dasein's caring—which encompasses the totality of Dasein's being (past, present, and future). Such distinctions as "past", "present", and "future" are then secondary, or derivative as compared to the more fundamental and holistic temporality of the care structure[350].

This Heideggerian notion of temporality operates by way of a dual mechanism that simultaneously brings the past and future into the present while also projecting Dasein out of the present onto its life as a whole[351]. It is both immanent and transcendent: immanent because Dasein is temporally contained or bounded by the finitude entailed within the certainty of death; transcendent because Dasein, in its caring, extends out

---

in Psychology, Volume 30, Issue 1, Pages 97-106.

[349] For instance, Heidegger writes: "Die Zeitung bedeutet kein »Nacheinander« der Ekstasen. Die Zukunft ist nicht später als die Gewesenheit und diese nicht früher als die Gegenwart. Zeitlichkeit zeitigt sich als gewesendegegenwärtigende Zukunft. " (*Sein und Zeit*, §68, p. 350) ("Temporalizing does not signify that ecstases come in a 'succession'. The future is not later than having been, and having-been is not earlier than the Present. Temporality temporalizes itself as a future which makes present in a process of having been. " (*Being and Time* trans. Macquerrie & Robinson, §68, p. 401))

[350] Ibid. Guignon C. (2012).

[351] Here I am attempting to express Heideggerian temporality though the lens of the received linear view of time. If Heidegger is taken seriously, this way of expressing things would be thought to be highly misleading because it reverses the order of ontological primacy.

beyond the bounds of the present. It seems apt to say that for both Heidegger and Taylor,

the meaningfulness or significance of one's life emerges from one's understanding and

owning up to one's life as a whole, and is not reducible to understanding or owning any

component temporal part(s) in isolation.

Similarly, the meaning of a narrative can only be understood within the context of

having the whole narrative in view. One cannot assess the meaningfulness of either a

portion of a narrative or the narrative as a whole, or before that narrative is complete,

because meaning can only be assessed by reference to the narrative as a whole[352].

Narratives can have many structures. They can follow linear or nonlinear plots. They can

be arranged chronologically or non-chronologically. They can exemplify a theme, have a

moral, express truths, and they can reveal the absurd. Narratives resonate with and

become meaningful to us through our relating to or identifying with them in some

(usually analogical) way.

But it is this holistic element of a narrative—the idea of living one's life as if it is

a story, of crafting one's life as if one is the author of a narrative, of acting as if one were

the protagonist—it is this powerful analogy between finding meaning in one's life and

meaning in a narrative that impresses and motivates thinkers like Charles Taylor,

Alasdair McIntyre, and others to suggest that living one's life as a narrative provides the

basis for an ethical life. Usually such narrativists emphasize cohesion, unity, and

---

[352] Viz. Sophocles *Oedipus Rex* l: μηδέ᾽ ὀλβίζείν, ρρὶν ἂν τέρμα τοῦ βίου ρεράση μηδεν ἀλγείνὸν ραθὼν, ("deem no man happy, until he passes the end of his life without suffering grief"); Ovid *Metamorphoses* iii. 135: dicique beatus Ante obitum nemo · · debet, ("nobody should be called blessed before his death"); Aristotle, *Nicomachean Ethics* I. 10: Τί οὖν κωλύει λέγειν εὐδαίμονα τὸν κατ᾽ ἀρετὴν τελείαν ἐνεργοῦντα καὶ τοῖς ἐκτὸς ἀγαθοῖς ἱκανῶς κεχορηγημένον μὴ τὸν τυχόντα χρόνον ἀλλὰ τέλειον βίον; ("he is happy who is active in accordance with complete virtue and is sufficiently equipped with external goods, not for some chance period but throughout a complete life"). Thanks to Alex Levine for this observation.

diachronic continuity in a narrative as "healthy" or positively valanced; while

fragmentation, incoherence, and discontinuity or episodicity are "unhealthy" narratives.

Often, these normative valuations are either implied or asserted as self-evident or their

opposites suggested as absurd, without critical reflection or justification. Some thinkers,

however (Galen Strawson in particular[353]) offer critical assessment of these assumption,

and argue instead on behalf of lives *not* lived as narratives. According to narrativists,

however, such lives would not be the lives of *persons*. It is unclear how much of this

narrativist rejoinder is mere disqualification on terminological grounds and how much

turns on substantive philosophical dispute of the concepts involved.

Here we can see that "narrative" explicitly depends on a *discursive* and

conceptual organization of a story. That is, stories are the kinds of things that can be

told—stories are expressed linguistically, or minimally, if expressed non-linguistically as,

say, through interpretive dance, the meaning or significance of a story derives from the

audience members' translation or interpretation of the dancers' movements into

meaningful discursive or conceptual content[354]. Thus, narrative ability requires that an

agent be capable of conceptual representation or discursivity[355]. So the narrativist

---

[353] Cf. e. g. : Strawson, G. "Against Narrativity" (2004) *Ratio* 17, pp. 428–52. ; and Strawson, G. (2009) *Selves: An Essay in Revisionary Metaphysics*. Oxford University Press, London.

[354] That meaning is interpreter-relative is not a feature exclusive to non-discursive expressions. Here I mean to suggest that all meaning is interpreter-relative. In the case of self-conceptions, one is one's own interpreter; so the ability to interpret discursively must be possessed by the expresser (who is also the interpreter); even in cases of non-discursive self-expression (e.g. gesturing to one's self); because *qua* interpreter such expressions gain meaningful representational purchase discursively. This, however does not mean that interpreters must be indexed to individual humans (or brains) – it would do better to talk of interpretive *systems*, like DEAs.

[355] Once again, Brandom's inferentialism is a good touchstone. According to Brandom, discursivity is required for meaningful expression. Not everything needs to be actually spelled out discursively—there needn't be explicit representational propositional tokens—rather the discursivity that undergirds expressibility must at least be implicit in a way that can be procedurally (or inferentially) made explicit.

conception of personhood that requires self-reflection by conceiving one's life as a whole is necessarily representational[356]; while non-narrativist conceptions of personhood *do not necessitate* representation in the same way (though often still require representation for other reasons).[357]

At any rate, for both Taylor and Heidegger, a crucial aspect of personhood (or *Dasein*) is an attitude or comportment of openness (ἀλήθεια (*aletheia*), *Unverborgenheit*, *Erschlossenheit* variously rendered as "unconcealment" and "disclosedness") to one's personal narrative possibilities. This openness to the world is usually understood as a self-transcendent reliance on a holistic "background" of intelligibility or meaningfulness encountered through one's mundane and practical involvement in their everyday activities. Mapping this onto the conceptual vocabulary of Dynamic Systems Theory, we might say that a person (or *Dasein*) is necessarily an open system. Recall that open systems constitutively depend on their openness as a condition for their systemic identities; open systems are defined through their differential sensitivities to constraints and affordances made available for practical involvement by their local environments.

---

[356] Or minimally, representable. The general point is just that there is no non-discursive concept of narrativity.

[357] The much-debated problem of the relationship between "derived" or "non-derived" content and intentionality are relevant here; however I would like to table discussion of this issue until a more detailed picture of self-recognition emerges. For now, it is sufficient to note that the necessity of others in the intersubjective interactions that condition one's self-concept seems to support a view of derived intentionality; while the novel embodied interactions of particular agents in their specific local environments seems to support a view non-derived or 'original' intentionality. I wish to hold that intentionality and content do transcend the bounds of self (though this may already be implied by the problematization of self-boundaries found in embodied and extended accounts of cognition). Just as functions are functions in virtue of their roles within the wider systemic whole, content and intentionality are meaningful in virtue of the wider (embodied/extended) circumstances in which they mean. On the other hand, such meaning is always immanent or indexed to and cannot transcend such novel and particular agent-oriented circumstances. Open and far-from-equilibrium systems are meaningfully influenced by external factors; but such influence should not be mistaken for transmission or sharing of meaning between epistemic perspectives. Interaction and influence are neither sharing nor transmission.

Insofar as personhood requires narrative unity, we can also say that persons must possess the ability to self-reflect (cf. the taxonomic class of "self-reflection" in Richard Campbell's process-based model for an interactive ontology), which is a characteristic set of interrelated recursive and reflexive relations that system can exhibit. Further, as long as persons must constitutively continually be open to their life-possibilities, this makes them far-from-equilibrium—as equilibrium entails a closing-off of future possibilities (or a concealment of one's ownmost potential for being). Being-toward-death means being far-from-equilibrium[358].

With the idea of *aletheia* (openness or unconcealment) in conjunction with an understanding of one's life as a narrative unity, we can also get a better sense of the kind of nonlinear temporality that Taylor and Heidegger have in mind. As a narrative unity, one's life shouldn't be seen as a haphazard meshwork or hodgepodge of temporal timescales and influences. A life lived as an effort to complete a narrative in a unified way does not chaotically reverse temporal directions—at least not if one is successful at living "authentically". On the other hand, Heidegger certainly realizes that Dasein continually "falls back" into the inauthentic preoccupation by (ontic) concerns. In this sense, Dasein can be understood as alternating (perhaps chaotically) between authentic and inauthentic temporal modes. But what is crucial is that when Dasein falls back into

---

[358] For an open dynamic system equilibrium is death – it is the system ceasing to differentially respond to environmental factors, a closing-off or systemic isolation. Dynamic embodied agency can then be literally understood as definitive of life; as a requirement for being alive. On the other hand, the relationship is asymmetrical. I say that being-toward-death means being far-from-equilibrium, but the converse is not true; being far-from-equilibrium does not necessarily mean being-toward-death. Many systems are far from equilibrium (e.g. a paramecium) but cannot or do not comport themselves in a manner that is being-toward-death. In order to comport oneself in any way, one must be capable of grasping and taking a stand on one's self; one must perceive one's self as a self. Thus, in order to be-toward-death, a system must also minimally satisfy the conditions of recursive self-maintenance (care) and self-reflection.

this inauthentic mode of being—Dasein ceases to express that which constitutes Dasein *qua* Dasein—namely care for its own being, as a whole; allowing its Being to be an issue for it. By not comporting one's self to one's narrative unity, one ceases to fully be a person, according to narrativists like Charles Taylor. Perhaps when we're talking about the kinds of agents that humans are, we mean the kind of thing that can (or does) move between fully being persons and failing to live up to that potential. If so, perhaps the intensity model of temporality may be appropriate after all.

However, it seems that the most appropriate model of nonlinear temporality for capturing the kind of temporality that narrativists like Taylor, McIntyre, and Heidegger have in mind is either "dilationary time" or a well-organized (rather than haphazard) thematically-attuned meshwork temporality. Unfortunately an adequate critical treatment of the comparisons between Heidegger's temporality and these two temporal models would require substantially much more space than is available here. Instead as a proof of concept I will sketch a model that bridges the language of Dynamic Systems Theory, dilationary time, and narrativity. It should be clear by now that, as an approach based in process metaphysics, every aspect of a dynamic system is fundamentally temporal. Each component or relation involved in a dynamic system is a process; and all processes operate diachronically (in, or through, time).

The complexly interrelating processes involved in most biological systems—metabolic pathways, homeostatic functions, motility operations, etc. —occur on varying and different timescales. These processes, and the timescales on which they carry out their functions, are sensitive to and in many cases dependent on other processes and their

210

correlating timescales for their functioning. This means that, in open, far-from-equilibrium systems such as biological organisms both the (identities of the) processes and the timescales on which those processes occur are in continual flux. This fits well with the meshwork conception of temporality, and indeed on some accounts of agency this conception is likely the most appropriate. However, for describing systems that exemplify narrative unity, a dilationary model of time (which can be thought of as consistent with, or a special case of meshwork temporality) may be more appropriate because of its ability to depict both the provisionality and complex interrelations of processes with varying and different timescales for occurrent processes as well as the non-totalizing, open unity that defines narrativity.

A dilationary temporal model expresses the unity of the processes occurring on varying and different timescales in a narratively unified system. It does this by giving an ever-dilating present horizon that expresses the multiple changing timescales of different processes and events significant to the identity of the narratively-structured system. It also provides an orientation (viz. dilation is a dynamic process of outward expansion, so there is the bidirectional orientation of "inward" and "outward") that expresses the general movement from past to future without enforcing a uniformity in that motion. By expressing the present as a dilating boundary or horizon, the dilationary model also captures the provisionality of interpretations of past significances, and accommodates the revision of the significance past occurrences to come into alignment with the unity of the narrative. The shape and character of the inwardly-oriented past is continually reinterpreted by the shape and character of the ever-expanding present. Future prospects

211

are likewise continually reinterpreted in a dynamic process of temporal attunement. One can also understand the acts or moments of interpretation as themselves processes that influence the processes and significances of events, and help shape them into a unified narrative.

One can also use the dilationary model of time to express or explain the differences between systems that are unified narratives and those that are not; as well as the differences between systems that strive for narrative unity versus those that do not[359]. A system organized as having a unified narrative formulates some end that functions as a temporal terminus for that system. Since most narrative systems do not have an exact model of what their particular ends are[360], most systems capable of narrativity are instead characterized by their striving toward *abstract* ends (rather than their attainment of narrative unity[361]). This striving is expressed in the dilationary model of temporality as a *linearization*[362] of many of the processes that factor into whatever is sought as the end for that system. Processes that the system takes to be relevant to its ends are *aligned* with or

---

[359] The former distinction (whether a system is narratively unified or not) is a descriptive difference between the qualities and capabilities of a system. The latter distinction (whether a system strives for narrative unity or not) is a prescriptive difference between the kinds of activities in which such systems engage. It should be noted that only systems capable of narrative unity can strive (or not) for narrative unity; though striving for narrative unity also suggests that the system is hitherto not yet narratively unified.

[360] Viz. most systems aren't capable of predicting particular futures—if systems predict, their predictive success tends to be general or abstract—predicting that it will rain tomorrow is comparatively more general than predicting the particular motions of the molecules of atmospheric water tomorrow. If physical phenomena are genuinely chaotic, then the sensitive dependence on exact conditions should in principle bar such predictability. This in turn would rule out the possibility of genuinely narratively unified systems in the descriptive sense (see previous footnote). All narrative systems, are then systems that strive for narrative unity.

[361] The fact that such narrative ends are *abstract*, and *not concrete* is significant. Far-from-equilibrium, open, sensitively dependent, dynamically-coupled systems cannot possibly narratively entertain their *concrete* ends. However, abstract ends do not map to *particular* concrete conditions (they map to *general* abstract conditions), so one could imagine a set of possible concrete outcomes, the attaining of any one of which would satisfy the abstract conditions of meeting one's ends.

[362] An extrapolation—a "futural projection".

attuned to the more general process of attaining that end—so as to facilitate and reinforce

the attainment of it[363]. Of course, the formulation of this end itself may be sensitive to the

effects of environmental influence, and become subject to alteration. When this happens,

the qualitative effect is that the unifying end that such a system previously took to be

formative of its narrative identity is altered and the system must as a result either redefine

its narrative identity, or abandon narrative unification altogether. The identities of

systems that strive for narrative unity are defined by their ends—but as dynamic systems

whose ends are provisionally stipulated, updated, and altered; the dynamic nature of the

identity of projected ends entails that the narrative identity of such systems must also be
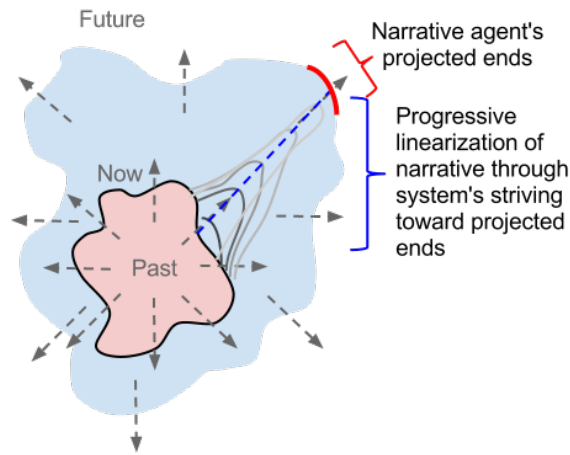
dynamic.

Figure 3: Graphical representation of dilationary model depicting the
linearization or attunement of processes toward attaining an end.

An example will help elucidate. Suppose a narrative system had established as its

end to achieve notoriety as an architect. As a narratively unified system, it would

organize and align its other processes so as to facilitate achieving notoriety as an

---

[363] This also undergirds the normativity of action: an action can be considered good or correct or better if it
contributes to or facilitates the attainment of posited ends; bad or incorrect or worse otherwise.

architect. This may include taking up projects that develop the bodily skills required for

notorious architecting, following a narrative path socially accepted as the way to become

a notorious architect including going to school for architecture, seeking the advice of

expert architects, etc.. Altogether, this narrative system is organized such that its

formulation of this end functions as a guiding norm (either directly or indirectly) for the

rest of the system's processes. Perhaps the projects taken to facilitate this guiding norm

are frustrated to the extent that the system is no longer capable of unifying a narrative

around achieving notoriety as an architect. The system has a couple of general options: it

can replace its formative narrative end with something else, say becoming a skilled

carpenter; it can continue striving "against all odds" at achieving notoriety as an architect,

even though it doesn't seem like a possible option anymore; or it can abandon striving to

fulfill any narrative end whatsoever. If it does the latter, it ceases being a narrative system

(at least by its own lights) and becomes a non-narrative system[364]. If it continues despite

the impossibility of attaining its end, one might call it a deluded narrative system, or a

broken or malfunctioning narrative system. If it changes its end, it is no longer the

narrative system it was—its identity has fundamentally changed, but it remains a more-

or-less functional narrative system (of some or another sort). Its previous attempts at

attaining notoriety as an architect are now re-cast in a new light—taken now as

contributing in some meaningful way toward this new narrative; perhaps as an

---

[364] Whether a system is narrative may be interpreter-relative; that is, one may judge of another that she
sufficiently satisfies the strictures of a life lived narratively, while the other doesn't perceive, plan, or
construct herself in such narrative manner. Does this make her a narrative, or non-narrative system? I
see no reason why such an assessment cannot remain indexed to the perspective of the interpreter. To
herself, she appears as non-narrative; to the second- or third-person observer she appears as narrative.
There is no need to hypothesize an interpreter- or observer-neutral "fact of the matter". Naturally, on the
dilationary temporal model these assessments as to whether one lives a life narratively or not are
provisional and updatable on retrospective reflection for first, second, and third personal accounts.

explanation for how the system was able to come to discover its (current) narrative identity as a skilled carpenter. That is, by taking up a new narrative identity, or by discarding narrativity entirely, the significances of past occurrences become re-signified. However, this process of re-signification also occurs when the narrative identity is thought not to change[365]. Thus, the narrative identity, on the dilationary model, requires a concept of identity that derives from a dynamic process of differentiation instead of stasis. This should not come as a surprise, however, as DST already depends on a process rather than substance metaphysics.

So, as time dilation continues, both the shape of the future (what is yet-outside the broadening boundary of the present)—how the future shows up for, and meaningfully factors into, the system's occurrent processes—and the shape of the past (what is inside the broadening boundary of the present) and its significance(s) for the system's occurrent processing, are continually altered and re-signified with respect to the developing narrative system. This allows us to see dynamic narrative systems[366] as temporally open with respect to both past and future; which means that both can influence present or occurrent decisions, representations, interpretations, etc. while the system navigates its present local environment. Thus, narrative systems should be understood as open, far-

---

[365] It is important to note here that when a narrative identity is thought not to change, it doesn't mean that it isn't changing. There is a reciprocal relationship between a unifying narrative identity in its role as a guiding norm and the concrete processes which are taken to facilitate the attainment of the end(s) that define that norm. So, for example, as the system strives to become an architect of note, its formulation of what an architect of note *is* continually changes and is changed by the experiences thought to facilitate the system's becoming an architect of note. The identity of the narrative system is always in flux, but is still continuously regarded as having "the same" end (viz. "notoriety as an architect"). Thus narrative identities, when understood on the dilationary model of temporality, are constituted dynamically by their difference and change, rather than by their sameness and persistence. For a sustained discussion of a difference-based concept of identity, see the introduction in Gilles Deleuze's *Difference and Repetition*.

[366] Systems dynamically striving for narrative unity.

from-equilibrium, and nonlinear with respect to both spatial and temporal dimensions; and the dilationary model of time adequately captures all of these aspects. Further, these features fulfill the temporal requirements for personhood posed by narrativists[367].

4.6   *Self-Awareness (Self-Reflection, Representation, Social Ability)*

One of the more difficult to pin-down conditions often thought to be at least necessary if not sufficient for personhood is the criterion that persons must be either self-aware or minimally capable of self-awareness. This produces five possibilities of progressive strength. One may be synchronically self-aware, or capable of synchronic self-awareness. One may be diachronically self-aware, or diachronically capable of self-awareness. The strongest requirement is that in order to count as a person one must be both synchronically and diachronically self-aware. This means that a person would need to both be self-aware (here and now), and continuously maintain that self-awareness over time. If the candidate strongly-self-aware person comes and goes; if there is any discontinuity in self-awareness; the candidate would not be considered a person on this strongest requirement. Less stringent are requirements for only one of synchronic or diachronic self-awareness[368]. Weaker still are requirements that one be merely capable of either diachronic or synchronic self-awareness; and the weakest personhood constraint is that one be merely capable of synchronic self awareness[369]. However, there are different

---

[367] I do not mean to suggest here that the dilationary model of temporality is unique in its ability to satisfy these constraints; in fact I would like to suggest (without protracted argument) that a careful formulation of the *meshwork model* should also be able to satisfy them as well.

[368] Though it would seem that to require diachronic continuity in self-awareness would itself entail synchronic self-awareness. The requirement of *only* synchronic self awareness doesn't care whether one has been self-aware in the past, or will be in the future; it is concerned only with here and now.

[369] See Fig. 4, below.

understandings of what awareness and self-awareness, themselves, *are*. For instance, we may ask whether attentive, thematic focus is required for self-awareness[370]. If so, then it can be argued that perhaps many humans—on the stronger self-awareness requirement— often fail to count as persons, because they often are not attentively or thematically focusing on their selves or their awareness of their selves.

Strength		Diachronic	Synchronic	Self-Awareness
1	**is**	x	(x)	continuous
	is capable	(x)	(x)	
2	**is**		x	synchronic
	is capable	x	(x)	(possibly continuous)
3	**is**			capable of
	is capable	x	(x)	both
4	**is**		x	synchronic
	is capable		(x)	(discontinuous)
5	**is**			capable of
	is capable		x	just synchronic

Figure 4: Chart of possible self-awareness requirements for counting as a person.
(Parentheses indicate the requirement is logically entailed by another requirement)

The weaker requirements that take self-awareness as merely sufficient condition for personhood would not, on the other hand, rule out a candidate for personhood merely on the basis that it isn't *currently* (synchronically) self-aware. This weaker requirement only holds that a candidate must be *capable* of, or have the capacity for, self-awareness in order to count as a person[371]. In this sense, most socialized, adult humans meet the weaker requirement for personhood—even when they aren't thematically attending to their self-awareness. You'll notice, however that I still qualified the candidate persons by

---

[370] Daniel Dennett is well-known for advocating this view.

[371] As with the stronger version, one can formulate synchronic and diachronic versions of the weaker requirement. The synchronic version holds that, at some time one possesses the capability or disposition for self-awareness, even if one isn't expressly exhibiting it. Such a formulation, of course, encounters difficult hurdles when it comes to epistemic verification. The diachronic version holds that one has in the past, or may in the future expressly exhibit self-awareness. The latter is what I have in mind for the above.

saying that they are both socialized and adult. Many theorists[372] believe there is an intimate, perhaps even constitutive, connection between the processes of socialization and the development of one's capacity for self-awareness. Most newborns are often thought not to be self-aware, while most adults are[373]. Thus self-awareness is not an innate capacity, but rather something constructed or learned—either as a matter of biological or social development (or both)[374].

Additionally there is a strong thread in Continental phenomenological thought[375] which takes as formative the processes of socialization, and the recognition of one's self as primarily belonging to or with others; of grasping first the personhood of others before one is capable of grasping one's self as a person; and of understanding personhood as dependent on concrete experiences and encounters with particular others—all of which

---

[372] Especially the phenomenologists listed in the first four footnotes in this chapter.

[373] There is some dispute on this issue. Gallagher & Meltzoff (1996) have argued that neonates display a rudimentary form of "body image" which is a form of self-awareness. Others such as Susan Jones (1996), Georgy Gergley (1995, 2004), Anisfeld, Turkewitz, & Rose (2001), Victoria McGreer (2001) have argued in various ways against this claim. See Walsh, Talia (2006): "Do Neonates Display Innate Self-Awareness? Why Neonatal Imitation Fails to Provide Sufficient Grounds for Innate Self- and Other-Awareness", *Philosophical Psychology*, 19:2, 221-238.

[374] It may still turn out that some rudimentary forms of self-awareness are innate, but there are strong reasons to believe that such self-awareness is not of the narrative variety—if for no other reason than that we have strong evidence that newborns must learn language. Alternatively, based on an inference about poverty of stimulus and the systematic complexities of natural language, Noam Chomsky has argued (1965) for the existence of an innate "Language Acquisition Device" (LAD) that provides neonates with the genetic capacity for language-use. If this is so, it is at least possible that the capacity for narrativity is likewise innate. Chomsky has since abandoned his LAD theory in favor of a more robust (though arguably ad hoc) theory of Universal Grammar which suggests that linguistic grammar is generative rather than genetic. Similarly Jerry Fodor has offered his "Language of Thought" hypothesis (LoT, or LOTH) suggesting that all cognition traffics in symbolic representations that are manipulated in accordance with general language-like rules (1983). While there is no widespread agreement as to whether representational thought is innate or if it manifests later through childhood development, the LOTH is often taken to lend support for nativist arguments.

[375] Owing especially to works of Edmund Husserl (*Cartesianische Meditationen und Pariser Vorträge*, 1973 [1931]), Max Scheler (*Wesen und Formen der Sympathie*, 1973 [1912]), Edith Stein (*Zum Problem der Einfühlung*, 1989 [1916]), Martin Heidegger (*Sein und Zeit*, 2008 [1927]), Maurice Merleau-Ponty (*Phénoménologie de la Perception,* 1945), Emmanuel Lévinas (*Le Temps et L'Autre*, 1979), andJean-Paul Sartre (*L'Étre et le Néant*, 1976 [1943]).

requires social interaction and entails that such capacities cannot be strictly innate.

Further, there has been much debate about what precisely constitutes self-awareness. In particular, there is lively discussion as to whether self-awareness requires only "primary" or "first-order" consciousness (consciousness of the current situation in one's local environment, current emotional state, etc.), or whether it requires "secondary" or "higher-order" consciousness (consciousness *that* one is conscious, or consciousness *of* one's consciousness, awareness of one's emotional state)[376]. The general issue is a question as to whether self-awareness must be thought of as *recursively* defined. Some theorists[377] combine secondary consciousness with narrativity, suggesting that the ability to conceive of one's current situation, or one's life as a whole, in terms of a narrative which incorporates (minimally) one's past experiences and emotions or biographical history, is precisely what it means to be self-aware.

Others have related self-awareness to the cognate concept of self-recognition. These "recognition theorists" draw mainly on the Frankfurt School interpretations of the "dialectic" philosophy of G. W. F. Hegel. The basis for "recognition theory" is in explaining how personal and social understandings of personhood or selfhood are developed historically (and dialectically) through intersubjective and social processes of recognition and misrecognition[378]. When one experiences another as sufficiently similar

---

[376] This distinction may be traced to Kant's differentiation between transcendental and empirical modes of apperception. Cf. also Bickhard, M (1993).

[377] Notably Daniel Dennett's "Multiple Drafts Model" (*Consciousness Explained*, 1991, and "The Self as the Center of Narrative Gravity", 1992). His account takes the self to be a "center of narrative gravity" which is a sort of epiphenomenon that results from the generally serial construction and reconstruction of narratives that make sense of occurrent events. For Dennett, self-awareness is just one's sense of that center of narrative gravity—which itself is just an abstract (or epiphenomenal) relation borne of particular moments of experience.

[378] Cf. e. g. Honneth, Axel. (1996) *The Struggle for Recognition: The Moral Grammar of Social Conflicts*.

or dissimilar, the "phenomenal feel" is characteristic of that similarity or dissimilarity, and one's understanding of one's self, by relation to the perceived similarity or dissimilarity of the other, is affected. Similarity and dissimilarity are a difference of degree, as are recognition and misrecognition.

At the extremes one may misrecognize someone as absolutely dissimilar to one's self—this has the effect of "Othering" the other[379]—of experiencing the other as entirely foreign or alien. As a misrecognition, this means that one's phenomenal experience of the other is mistaken; one experiences the other as alien even though the other is more similar to one's self than one perceives. One may also misrecognize another as very similar. This results in taking another to be quite similar when really the other is quite different from one's self. Alternatively, one may *recognize* others as either similar or dissimilar. When another is recognized as dissimilar it is still a manner of "Othering" the other. According to the "recognition theorists" everything that is meaningful to a person's self-understanding is informed by this dialectic social dynamic of recognition, misrecognition, taking as similar, and Othering; and it is this dynamic that defines personhood. Persons are the kinds of things that are defined and shaped by their being caught-up in recognizer-recognizee relationships with others[380]. One important consequence of "recognition theory" is that it construes the definition of personhood as

---

Polity Press. ; and Honneth, Axel. (2007) *Reification: A Recognition-Theoretical View*. Oxford University Press.

[379] In the sense of the "constitutive other"—taking one to be radically different from and utterly alien to one's self. "Othering" is a process of differentiation, of abstracting differences rather than similarities, of constructing something as categorically different or entirely alien. Canonical discussions of "Othering" are found in Levinas, Emmanuel (1974). *Autrement qu'être ou au-delà de l'essence*. ; and Said, Edward W. (1978) *Orientalism*.

[380] Cf. Taylor, Charles. (1994) "The Politics of Recognition". *Multiculturalism: Examining the Politics of Recognition*. Ed. Amy Gutmann. Princeton University Press. pp. 25-73.

indexed to the dynamic processes that define our concrete relationships with others. Further, the Hegelian spirit of this scholarship suggests that such dynamics work at higher levels of abstraction—e. g. at the level social group interactions, or in wider cultural dynamics[381].

More generally the idea of self-recognition involves developing a sense of self and coming to acknowledge that self as in some way being characteristically owned, in contradistinction to what is understood as either unowned or owned by another[382]. One comes to a recognition of one's self by first recognizing others as persons, then redirecting that recognition toward that which can be recognized by others[383]. This is made possible by coming to understand the bidirectional relationship between recognizer and recognizee—that persons occupy the dual roles of recognizing others and being recognized by others. Once one understands that they can be recognized by others, they can simulate putting themselves in the position of the other and use their capabilities as a recognizer on themselves, as simultaneously recognizer and recognizee. This is the recursive and reciprocal process of self-recognition[384]. One becomes a person through realizing this self-recognition. Importantly, to realize such a self-recognition requires

---

[381] For a slightly different approach that reaches remarkably similar conclusions see the preface to Merleau-Ponty's *Phenomenology of Perception*.

[382] Taylor writes: "Being true to myself means being true to my own originality, which is something only I can articulate and discover. In articulating it, I am also defining myself" (Ibid. p. 31)

[383] Cf. e.g. Gopnick, A. (1993) "How we think about our minds: the illusion of first-person intentionality," *Behavioral and Brain Sciences* 16:1

[384] This "dialectical" process is expressed in G. W. F. Hegel's *Phenomenology of Spirit*—in particular the section commonly referred to as the "Master-Slave Dialectic". In this section Hegel shows how being "Othered" (being taken as an inferior and alien other by someone else) results in self-recognition, while failure to do so results in failure to self-recognize. In the Master-Slave Dialectic the slave is "Othered" by the master, and comes to recognize himself while the Master—not having been "Othered" doesn't. Ultimately the slave—having achieved self-recognition—is capable of continued development while the master is 'left behind' as obsolete.

understanding one's self as a self defined by and existing amidst others selves alike in kind[385]. The basic difference between one's awareness or recognition of their self and one's awareness or recognition of others is the phenomenal feel of ownership.

However, one must take care in clearly describing the phenomenal feeling of ownership. One concept of ownership is that of possession. On this concept of ownership one *has*, or is laden with, one's self. Possession as ownership is not a matter of identity—of *being* one's self—rather it is a *distancing* from that identity. The idea of possession both draws objects in to be included as owned, while simultaneously excluding what is owned as object rather than subject—as other rather than self. The possessed self is a detached self; one who owns by distancing and objectifying the self. An alternative concept of ownership is that of identity, or what Heidegger refers to with his use of the word "*Eigenste*" (often translated as being one's "ownmost"). Ownership as identity is a matter of *being* one's self. As identity, there can be no distancing between what is owned and what one is. Thus, on this concept of ownership the self just is (nothing but) whatever

---

[385] It is extremely important here to understand that I have been using terms like "recognize" and "understand" in a more colloquial rather than technical sense here. Others, particularly those in the Pittsburgh School, like Brandom, take the terms and related concepts of recognition and understanding to be fundamentally discursive (so conceptual and representational). I do not mean to imply such things, however. The way I use words like "recognize" and "understand" here are less specific and include the kinds of non-discursive embodied self-perception that Merleau-Ponty describes in a passage on pages 93-95 in *Phenomenology of Perception*, then later in the final chapter of *The Visible and the Invisible* with his notion of "chiasm". Chiasmatic self-recognition is the ability to discover one's self as the kind of thing that both perceives and is perceived by and through using one's body to simultaneously sense and be sensed. His most well-known example includes touching one's left hand with one's right hand, and shifting from one aspect (right hand as touching, left as being touched, or vice versa) to the other—and dwelling in the ambiguity of the sensation. Much like [discursive] recognition of others functions dialectically for Hegel, Husserl, Heidegger, and Sartre in the development of self-recognition, so too for Merleau-Ponty does this "chiasmatic" experience work to develop a non-discursive but bodily-perceptual sense of self-perception. The way I use the term, both discursive (representational) and non-discursive (embodied sensory-perceptual) experiences are forms of recognition.

is phenomenally felt as owned[386]. These two concepts of ownership can be seen in different attitudes toward embodiment. Understood as a possession, one *has* a body and this body is both owned by and distanced from the self. Understood as identity, one *just is* this body (so there can be no distance from the self). The former follows a generally Cartesian conception of self, while the latter expresses an alternative to Cartesianism[387].

Self-awareness in terms of self-recognition is *conceptual, discursive,* and *thematic*[388]. In order to recognize anything *as* something, one needs to draw upon the resources of representation. Thus to recognize one's self *as* being owned is to represent one's self to one's self. One may then ask whether there is any distance between this representation of one's self and one's self; or if one's ownmost self *just is* this representation—in which case the formulation of self is not so much a *re*presentation, but rather a *presentation* of self. However, as a conceptual, discursive, and thematic presentation, the self that is presented must be presented representationally—through the thematic use of language (or some representational language-like "Mentalese"[389]). If this

---

[386] Nietzsche expresses this pointedly in *Ecce Homo* where he, citing Pindar, implores: "*Werde der du bist*" ("Be [become] who you are").

[387] Mark Rowlands explores these differences at length in his 2010 *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*.

[388] These three, in addition to "representational", comprise a cluster of related terms such that each depends on and entails the others. Altogether, they belong to a general view about the structural and functional composition of the mind variously called "Cognitivism", "Classical Cognition", "Computationalism", or more generally "Representationalism". This approach takes cognition as fundamentally the nomological manipulation of discrete, representative, symbol tokens. On this picture, "concepts" are such tokens. The rules for the manipulations that can be performed on these tokens are generally thought to be language-like with the basic unit of semantic cognitive meaningfulness being the propositional sentence—this is what is meant by "discursivity" and what is referred to in discussions of "conceptual content" (by comparison "non-conceptual content" is then the idea that there is something meaningful but is not expressible in propositional or discursive form). To say an experience is "thematic" is to say it can be expressed narratively—and as mentioned in a previous footnote, narrative expressibility depends on discursivity.

[389] "Mentalese" is one of the common names for the domain-general mental language that Jerry Fodor posited with his "Language of Thought" hypothesis.

is so, then selves are thoroughly discursive—there can be no "non-conceptual content".

If, on the other hand, there is distance between the representation of one's self and one's (ownmost) self—there is something more to one's self than what one can express representationally—then this opens up a space in which one's self may outstrip discursivity, and we can then ask in what does such a (non-discursive, non-conceptual, non-thematic) self consist? What am I (*qua* self), over and above however I can represent myself to myself? One avenue may be to acknowledge that self-recognition is fundamentally social—in my ownmost I may be that which I am taken to be by others— that which I am recognized by others *as*. Interestingly, here one's ownmost self is still representational, but in a way that is somehow inaccessible to the subject for whom that self is owned (but accessible to others).

However, there may also be more to self-awareness than self-recognition. Self-recognition may be necessary but insufficient for self-awareness. Whatever else may be involved with self-awareness, we can ask of it whether it is representational (conceptual, discursive, or thematic). So far I have not said much about how sensation, perception, or emotion relate to self-awareness. If there is more to self-awareness than recognition (by one's self, or others) these seem like appropriate areas of phenomenal experience to query.

Sensation and perception go hand-in-hand. At their most general, senses are physiological mechanisms that provide perceptual "data" about the local environment to an organism. There are many ways of sensing. Paramecia sense sugar gradients using apparatus sensitive to specific classes of chemicals in the surroundings. Homing pigeons,

224

it is now thought, are sensitive to fluctuations in the Earth's magnetic field, and relate

those fluctuations to smells (chemical sensitivity) in order to establish unique location

signatures. Humans possess the familiar "five senses" (sight, hearing, taste, smell, touch),

as well as the less familiar senses of equilibroception (sense of equilibrium or balance

and acceleration), thermoception (sensitivity to heat and cold), proprioception (relative

sense of one's body in space), kinesthesia (how that body relates to the space surrounding

it through movement), nociception (sensitivity to pain, or nerve and tissue damage), and

chronoception (sensitivity to the quality of the passage of time[390]). All these senses detect

aspects of the external local environment in a way that can be perceived by the human

organism. In addition to these there are regulatory and homeostatic senses internal to the

organism such as various muscular stretch and chemoreceptors that provide intra-

systemic information to the perceiving organism.

However, one may reasonably wonder how important is the inside/outside

distinction in organismal sensation. Each sense modality has its own perceptual

significance for the organism. Each sense contributes a *sui generis* component to the

overall sensory-perceptual apparatus and as such we might do well to distinguish equally

between the multiplicity of sense modalities rather than categorize them into two

conceptual groupings in accordance with a familiar Cartesian scheme. On the other hand,

the grouping of sense modalities as "inside" and "outside" may itself be the product of

proprioceptive and kinesthetic processes for which such a distinction may be

significant—the fact that a sense modality operates inside or outside may matter

---

[390] That there is such a perception lends support to nonlinear, non-homogeneous models of temporality.

proprioceptively, whether or not it matters conceptually.

Perception is generally understood as the processes through which information from the various sense modalities functioning for an organism become identified and interpreted: the processes by which sensory experience comes to bear significance for the organism. There have been many theories of perception attempting to explain and describe the mechanisms by which perception occurs (often in humans), as well as how perceptual content comes to bear semantic or conceptual content. The differences in these theories shape and constrain the basic phenomenal qualities and character of experiences possible for the organisms to which they pertain. Indeed, how one understands perception bears directly on the resources available for understanding agency and thus personhood. Unfortunately, this chapter cannot sustain a prolonged discussion of the relative differences between theories of perception and how these differences affect theories of mind, cognition, action, and personhood[391].

Self-awareness can be understood as perceiving one's self as a self, or as a person. Here, epistemic constraints shape dialogue on the subject. One may ask for the conditions or criteria that must obtain in order for one's perceptual processes to count as genuinely self-aware. We may ask "what is it like for one to have perceptual self-awareness?". I suspect, however, that because of phenomenological limitations, specifically that one cannot experience another's phenomenological experiences, this question quickly becomes incoherent. The phenomenological character of self-awareness must necessarily be immanent to the phenomenological being for which the question of self-awareness is

---

[391]For a more comprehensive treatment of this topic, see Shaun Gallagher's 2005 *How the Body Shapes the Mind*.

an issue. One cannot meaningfully ask what it is like for *one* (in general) to have

perceptual self-awareness—rather, one can only meaningfully ask what it is like for *me*

(in particular) to have perceptual self-awareness. It is possible that for different subjects,

what it means to have perceptual self-awareness is phenomenally different. However

even such speculation runs up against the same epistemic limitations that quash the above

question. If one is in search of general criteria the best one can hope for is intersubjective

assent about the conditions of perceptual self-awareness; and this requires concrete,

dynamic, embodied interaction between specific agents with unique epistemic

perspectives. However, it also isn't clear that such general criteria would help elucidate

anything about particular instances of perceptual self-awareness. And what else could

self-awareness *be* apart from particular concrete instances of self-awareness? The

problem here is a fundamental disconnect in applicability between general musings about

conditions common to all and only those entities thought to be self-aware (of which there

is no common assent), and the particular concrete epistemic conditions that constitute

instances of self-awareness (which may be token-*sui generis*).

Narratively speaking, we could make a similar point by saying that the character

of self-awareness may be synchronically and diachronically episodic. *Synchronic*

episodicity is the idea that experiences occurring at the same time are distinct from one

another. So, for instance, your current experiences are distinct from my current

experiences[392]. As it pertains to defining perceptual self-awareness, synchronic

episodicity means that at any given time different loci of perception may have different

---

[392] Cf. Strawson, G. "Against Narrativity" (2004) *Ratio* 17, pp. 428–52. ; and Strawson, G. (2009) *Selves: An Essay in Revisionary Metaphysics*. Oxford University Press, London.

definitions or phenomenal feelings of perceptual self-awareness—that you and I may have different phenomenal experiences of what self-awareness "is like". *Diachronic episodicity* is the idea that for some temporal entity moments signify differently across time[393]. So, for instance, the significance of an event (say writing a dissertation) may mean something different now than it did in the past, or something different in the future than it does now; and there may be thematic discontinuity in these significances across time. As it pertains to defining perceptual self-awareness, diachronic episodicity means that for some temporal perceptual entity, what it means to be self-aware, or the phenomenal feeling of self-awareness may change or be different for that perceiving entity at different moments across time.

The narrativity thesis required by some theorists as necessary (or sufficient) for personhood can come in stronger and weaker versions. The weaker version requires only diachronic narrative unity (as opposed to episodicity), but does not demand synchronic narrative unity. This weaker position argues for strongly individualist narrative identity as a requirement for personhood. By not emphasizing synchronic narrative unity, it downplays the importance of social relations and interactions and belonging in social groups for shaping narrative meaningfulness in one's life. This weaker version of narrativity would be included in what Shaun Gallagher has called "philosophically autistic" theories of intersubjectivity—those that do not adequately incorporate social interaction in constituting personhood[394]. The stronger version of the narrativity thesis argues that in order to be fully considered as a person one must strive to unify both

---

[393] Ibid.
[394] Cf. Shaun Gallagher (2009). "Two Problems of Intersubjectivity".

synchronic and diachronic narrativities. This means locating and defining one's self or

one's personhood within social norms, practices, and concrete interactions with others (in

addition to the weaker diachronic narrative requirements). This view ameliorates

Gallagher's charge of philosophical autism by properly including social factors (such as

concrete interpersonal interactions) as essentially constitutive of personhood. Few

narrativists argue for the weaker narrativity thesis[395]—generally those who argue for

narrativity defend this stronger position[396].

Ultimately however, perceptual self-awareness is a matter of phenomenality—

what it is like to have awareness of self. And it is a non-trivial question as to whether

there can be some set of satisfiable, severally necessary, and jointly sufficient conditions

for counting as perceptually self-aware. For this reason, while perceptual self-awareness

may be an important consideration in the calculus of determining inclusion in

personhood, the only meaningful contribution to discourse on the matter must come from

the areas of *conceptual* rather than *non-conceptual* self-awareness. This does not mean

that phenomenal, non-conceptual (or pre-conceptual), perceptual and sensory experiences

necessarily do not or cannot play a role in actually determining personhood—but only

that, by dint of the intractable epistemic constraints that preclude decidability on the

skeptical worry of the comparability of similarities and differences in perception and

phenomenality across subjects, such considerations cannot fruitfully contribute to the

---

[395] For fairly straightforward reasons the weaker, more "individualist" view appeals to libertarians—those who view agents as autonomous individuals who fully own themselves and their actions; and who believe that claims on these 'liberties' by others or by society-at-large are illegitimate. Many Enlightenment thinkers (notably John Locke) espoused broadly libertarian views.

[396] Notably a significant plurality of those in the Phenomenological and Communitarian traditions.

debate about what constitutes persons[397].

There are two ways to approach self-awareness from the perspective of Dynamic Systems Theory. First, we can understand self-awareness as a process by which the various qualitative states of a system are monitored, affected, and/or maintained. Self-awareness, then, would be a more general metaprocess whose domain consists in the synthesized domains of the various occurrent sensory and perceptual (sub)processes. Functionally, self-awareness accomplishes the dual task of bringing together into a common domain sensory and perceptual modalities that are otherwise distinct; as well as reflexively assessing the qualitative status of this more general domain. To be clear, it is not necessary that this more general domain comprised of synthesized sensory and perceptual modalities must be representational, since the assessment of the status of this domain needs only be qualitative. Thus, self-awareness may be enacted as a non-conceptual bodily awareness. However, nothing precludes representation from contributing to or even underwriting the qualitative assessment involved in self-awareness. The DST approach is capable of supporting theories with varying degrees of representation.

The second way that DST approaches self-awareness is by addressing inextricably *phenomenal* character of self-awareness. Self-awareness is, at bottom, a phenomenal experience. There is something it is like to be self-aware—to experience awareness of

---

[397] Another way to formulate this problem is: as long as perception is understood individualistically (or internally or solipsistically), claims about perceptual self-awareness cannot overcome skeptical epistemic concerns about the verifiability of comparisons and contrasts. However, this approach is precisely what is argued against by "Anti-Cartesians" and meaning externalists such as Hilary Putnam, P. F. Strawson, Andy Clark, David Chalmers, Donald Davidson, Susan Hurley, Fred Dretske, John Haugeland, Colin McGinn, Tyler Burge, Ned Block, Graham Harman, Patricia Kitcher, Alva Noë, Mark Rowlands, Shaun Gallagher, and Dan Zahavi amongst a wide array of others.

one's self as a self. How can DST assess or express the essential phenomenality of self-awareness? In short, it can't. But this doesn't put it at any comparative disadvantage with competing theories. If we take phenomenology as epistemologically primitive (as was argued in chapter 1) then it should come as no surprise that the phenomenal character of experience definitive of self-awareness cannot be further explained in terms of any more basic theoretical constructs[398]. Primitives, by definition, are taken as that which is not and cannot be explained further (if they could, they wouldn't be the primitives—rather whatever constitutes the further explanation would be more primitive). So DST cannot do any better than the most popular theory of mind in explaining the phenomenal experience of self-awareness. But it doesn't do any worse on this point either (namely because DST is robust enough to subsume most broadly functionalist theories about cognition). So DST doesn't obviously or immediately solve "The Hard Problem of Consciousness". What is important is that it allows multiple avenues of investigation and provides a general framework[399] through which research questions can be expressed and explored.

As mentioned previously, the systems posited by DST are an expression of the epistemic and conceptual constraints of the positor and do not purport to describe or explain the metaphysical contours of the world *as it really is, in itself*. Explanations are always explanations *for someone*. The world "as it is in itself" does not need anything explained; only that which is epistemically constrained does. DST provides a framework

---

[398] Of course, if one doesn't take phenomenality as primitive, perhaps one could; e.g. if one is a metaphysical realist, and naturalist one might attempt explain phenomenal qualities (qualia) like self-awareness as a qualitative emergent. The difficulty is then in getting such metaphysical commitments to pare with epistemic access. When phenomenality isn't taken as a primitive we face the problem of disjoined metaphysics and epistemology.

[399] Or Kuhnian "paradigm" or Lakatosian "research programme".

through which particular kinds of epistemic interrogators with specific types of epistemic constraints (such as humans with their unique sensory and conceptual apparatūs) can be explanatorily satisfied. Many such investigators are very interested in explaining the conditions for inclusion as persons in a "naturalistic" way; in large part because possessing the social status of a person is significant to many social practices and institutions, and because the naturalistic (or vaguely "scientific") attitude holds a privileged social position with respect to explanation. DST contributes meaningfully to such a project.

For what it is worth (which I do not think is much, if anything at all) I think that as far as the project of determining inclusion in personhood goes, *both* conceptual and non-conceptual, perceptual self-awareness should be included among the *sine qua non* of personhood. I would resist both strong and weak versions of the narrativity argument, however. I believe that neither diachronic nor synchronic narrative unities are necessary (nor sufficient) for inclusion in personhood (viz. there are persons who satisfy neither constraint). Agency on the other hand, is a significant constitutive factor in personhood because without the ability to choose and execute actions, there simply is nothing resembling what anyone means in talking about persons. Alternatively, navigating moral responsibility is precisely the reason for caring about the bounds of personhood so is essential to such discussions. It is this subject to which I next turn.

4.7   *Moral Responsibility*

Moral Responsibility is a surprisingly complex subject when approached from the perspective of naturalism. Some basic questions are:

- How is moral action possible (for naturally evolved entities such as ourselves)?

- How do we as persons develop our sense(s) of moral understanding?

- How do individual and social moral judgments relate?

- How can morality be explained in purely naturalistic terms?

These are four of the many questions that anyone who wishes to naturalize morality (to explain morality using an exclusively naturalistic framework) must address. The purpose of this section is to sketch how DST—as a sufficiently naturalistic framework—can adequately address these four questions.

Moral, or normative, behavior is behavior that can be characterized as good or bad, right or wrong, better or worse, etc.. Earlier I discussed how one DST-oriented position (Richard Campbell's "Process-Based Model for an Interactive Ontology") explained the emergence of normative behavior. Normative behavior is based in a broad sense on what it means for a system to "self-maintain", and is exhibited by any system capable of differentially interacting with its local environment. Any system capable of the most absolutely basic sense of "choice", and which can succeed or fail to self-maintain, acts normatively. In a certain sense, preferring self-maintenance as "good" or "right" or "better", and its alternative (dissipation or dissolution of the system) as "bad" or "wrong" or "worse" is *system-relative*. There is no *generalizable*, fundamental grounding for taking this normative orientation. There are two different senses of system-relativity here, however: relativity with respect to the system for which the norms are binding; and relativity with respect to the epistemic interests of the party investigating the system. The former is necessarily epistemically opaque and its normatively-binding scope is

233

immanently limited to the system itself. With regard to the latter, there may be a psychological explanation that grounds normative orientation. As temporally limited systems—systems that arise, develop, function, then dissipate or die—we tend to make better sense of prolonging functioning and delaying dissipation than of accelerating dissipation. We may then graft our preferences for self-maintenance of ourselves onto our assessments about dynamic systems in general as a way of relating to or making better sense of normativity in systems. Since we are the epistemic interrogators for whom DST is meant to help explain things, it should not be surprising that part of helping explain things involves understanding them in terms that cohere with our own attitudes. If I am a dynamic system, and I believe that my self-maintenance is good and death is bad, it makes sense to understand other dynamic systems by way of the same sort of normative valuation. *We* imagine that, from the paramecium's perspective, its own self-maintenance is good for it, and its own death bad. This helps *us* to explain (to ourselves) why it chooses swim toward, rather than tumble away from the highest concentration of sugar. It is choosing in favor of self-maintenance. But these normative impositions are our own, as epistemic investigators.

This is also how error is possible. Without tethering normativity to a concept like self-maintenance, one encounters the result that error becomes impossible. Without having a touchstone, there can be no basis to assess whether some action is "good" or "right" or "better". For any action one could make such an assessment (or its opposite) and there would be no basis for affirmation or dispute. But, even though the selection of self-maintenance as positive and dissipation as negative is somewhat system-relative, its

establishment as a basis for normativity is crucial—especially in the assessment of error. Only if we know that self-maintenance is "good" can we say that the paramecium's swimming toward a high concentration of non-nutritive saccharin is an error (because it doesn't facilitate self-maintenance, and may accelerate dissipation).

However, this does not yet distinguish between choosing wrongly, and choosing in error[400]. In the case of the former one chooses so as to accelerate dissipation, while in the case of the latter, one chooses in favor of self-maintenance, but is mistaken in that the action chosen fails to facilitate self-maintenance (and may even accelerate dissipation). The difference is one of low-level intention. With the same action or overt behavior one may intend self-harm and succeed, or intend self-maintenance and fail. Functionally, there may be no difference at all between acting in error, and acting wrongly—the relevant functional difference is with whether a certain function was performed sufficiently or not; and the consequences thereof. Functionality needn't involve normativity at all. If it does, normativity is indexed to and dependent upon the identity of the function being executed. The system can succeed or fail (where in some cases success and failure can be understood as a matter of degree, while in others it is a bivalent difference) at some function it is taken to be performing. However, if instead we understand the system to be performing a different function, the success and failure conditions will change accordingly. So for example if we see the paramecium as performing the function of swimming toward the highest concentration of "sugar-or-sugar-like-substances"—we may assess that it succeeds at that function. If instead we see

---

[400] Or between trying successfully to do something stupid and failing to do something smart. (Thanks to Alex Levine for this insight)

the paramecium as performing the alternative function of satisfying its nutrition requirements—we may assess that it fails at this function. Again, success or failure of function is indexed to the identity of that function and the conditions that define its successful or unsuccessful execution.

So we may then ask: what are the relevant functions that correlate with moral action? And what are the characteristics of systems capable of these functions? First, we may notice that moral action can only be accomplished by systems that are capable of moral responsibility—systems that are accountable for the morality of their actions. Here morality is usually taken to have two significant dimensions: responsibilities to one's self, and responsibilities to others. Remember, the primary reason for discussing moral responsibility is because it is taken by many theorists to be constitutive of personhood. So, insofar as moral responsibility is constitutive of personhood; personhood involves obligations both to the self and to others. And by extension, if systems are to be counted as persons they must also be able to bear personal and social responsibility. So the functions relevant to moral action are those functions that facilitate and enact personal and social responsibility. But what are these responsibilities?

Here we may return to issue of recognition. Recognition operates along the same relations that moral responsibility does—it functions reflexively (as self-recognition) and it functions interpersonally (as an asymmetric relation of recognition of and by others). How one reflexively recognizes one's self is informed both by how one is recognized by others and how one recognizes others. How one recognizes others is the mode by which one goes about recognizing one's self; and how one is recognized by others provides a

filter or constraint through which the content of one's self-reflection is informed. Likewise, one's self-recognition informs the way one recognizes others, as well as the way in which one understands how they are recognized by others. This dynamic dialectical process continually operates as a sort of triangulation to identify one's self and one's roles in society[401].

Analogously, moral responsibility functions *reflexively* (as one's obligations to one's self), as well as *interpersonally* (as one's obligations to others in particular, and society at-large). How one reflexively bears their moral obligations to one's self is informed both by how others bear moral obligation to one, and how one bears their moral responsibility to others. How one bears one's moral responsibility to others is the mode by which one goes about bearing one's moral obligation to one's self; and how others bear their moral obligation to one provides a filter that informs the content of one's moral obligation to one's self. Likewise, one's moral obligation to one's self informs the way one bears one's moral obligation to others, as well as the way in which one understand how others bear their moral obligation to them. As with recognition, this dynamic dialectic process continually works as a triangulation to identify one's moral status with one's self, with others, and within society.

We have already discussed one way in which dynamic systems can have a minimal normative obligation to themselves—recursively self-maintenant, far-from-equilibrium, open systems capable of error-detection and flexible learning bear the reflexive obligation to employ their capabilities in interacting with their local

---

[401] Husserl makes use of this kind of account in his phenomenological explanations for how intersubjectivity works. Cf. Husserl, E. (1988 [1931]) *Cartesian Meditations*, trans. D. Cairns, Dordrecht: Kluwer.

surroundings toward their own self-maintenance. It doesn't take much to elaborate this

obligation to a moral obligation that one do what is required in order to take care of one's

self. Of course, the devil is often in the details, and here it is a nontrivial question as to

what consists in "taking care of one's self". The DST approach would suggest that taking

care of one's self will sensitively depend on the particular conditions formative of the

situation in which such "taking care" would be occurring. So giving general, severally

necessary, and jointly sufficient conditions for "taking care of one's self" may turn out to

be a practical impossibility.

Further, the very notion of what one's "self" is, is inextricably tied up with, and

constitutively defined by one's relations with others, and within society [at-large]. Thus,

to "take care of one's self" necessarily includes other particular persons as well as society

at-large. One cannot bear a moral responsibility to one's self without also bearing moral

responsibilities to the others that contribute to defining and identifying both one's "self"

as well as what it means to bear moral responsibility—just as one cannot recognize one's

self without also recognizing others, being recognized by others, and understanding one's

self and others as the kinds of things that can both recognize and be recognized by others

(amongst which one is included). Because the self and moral responsibility are both

essentially social they are both bounded or circumscribed by the particular social

relations that shape them, while simultaneously transgressing that boundary through their

essential reference to the very relations that give them shape.

This means that because the self is defined through social relationships and roles,

the moral obligation to "take care of one's self" extends to taking care of the social

conditions that provide for the relationships and roles that define the self. This is

analogous to Richard Campbell's idea of "flexible learning" which involved the organism

making changes to its external environment as a mode of self-maintenance. As flexible

learners, organisms take differential control not only of their internal conditions in order

to robustly cope with changing external environmental factors; but rather also attempt to

control the external environmental factors so as to lessen the burden of internal

differential adjustment. However, while organisms can either be flexible learners or

not—they can either self-maintain by making changes to their environment, or not—

selves, or persons, *must* extend into the social realm. Campbell writes:

> So crucial and significant for human development is this nurturing and
> induction into sociality that the nice symmetry of the ontological model I have
> been outlining no longer applies. It is not the case that some reflective persons
> come together to form social groups, while others do not. While a few rare
> individuals have chosen to live as hermits, even they could not cut themselves
> off completely from social interaction. Rather, the emergence of reflective
> persons and the emergence of social institutions and certain social groups are
> mutually dependent and interactive. [402]

The point here is just that any claims we make about individual persons, and any qualities

or characteristics that constitute their inclusion in personhood, must essentially affect or

apply to the things we say about the qualities or characteristics pertaining to social

relationships, practices, and institutions; as a sort of "**transitive property of**

personhood".

We began this chapter by identifying four common characteristics thought to be

constitutive of persons—agency, a temporal sense of self (or narrativity), self-awareness,

and moral responsibility. The "transitive property of personhood" suggests that these

---

[402] Campbell (2009), pp. 474-5.

constitutive characteristics of persons must also relate constitutively to concrete social conditions, practices, institutions, etc. (and *vice versa*). It does not mean that social practices and institutions are persons, or that *all* characteristics constitutive of persons bear *the same* or *isomorphic analogical* relations to social practices and institutions. I am not claiming that social practices *must* be agents, narrativistic, self-aware, or possess moral responsibility. Rather I am saying that *because* these are characteristics constitutive of persons, they affect the character of the kinds of social practices and institutions that are possible; and inversely, that the kinds of social practices and institutions available affect the qualities that are constitutive of personhood. So, for example, what it means for a person to be self-aware depends (at least partially) on the social conditions, practices, and institutions through which personal self-awareness can be expressed. Likewise, the kinds of social conditions, practices, and institutions that are possible are shaped (at least partially) by constraints in self-awareness.

This point may be understood more generally as expressing a close co-constitutive relationship between embodiment and social practices and institutions. The bodily sensory modalities that comprise perception are just the different capacities or capabilities for bodily interaction with the local environment[403]. Perception is limited to or constrained directly by these capacities or capabilities. In turn, self-awareness, for instance, is constrained by the limits of perceptibility; as perception is the vehicle through which one recognizes others, and other-recognition is itself a condition for self-recognition (and thus self-awareness). Self-awareness then, constrained by the conditions

---

[403] The point is argued for emphatically about the human capacity for visual perception by Gibson, J. J. (1979). The Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

of perceptibility amongst other constraints, affects the possibilities for social encounters and the types of practices or social institutions in which the agent cant engage. In the reverse direction (the "downward causation" direction) concrete social practices and institutions constrain the kinds of activities that agents actually embody—literally shaping the perceptive body so as to influence and accommodate the transparent engagement and execution of these very social practices and institutions. One example might be the social practice of wearing shoes, boots, and other footwear. Here the practice literally constrains the body; the wearing of shoes shapes feet, and the ways in which feet can interact with the world. By wearing footwear, embodied agents are afforded new possibilities in the kinds of terrain that are traversable; but also constrain the shape and texture of feet that might otherwise have callused. New potentialities emerge for bodies coupled with the social practice of shoe-wearing; but some old potentialities (like running barefoot on rough terrain) are restrained.

4.8   *Looking Ahead*

The purpose of this chapter was to show how the framework afforded by DST is minimally adequate for talk of persons in all the relevant and important aspects of the literature. Naturally, one of the most important aspects of personhood is participation in social roles, relationships, practices, and institutions. So far we have only addressed this aspect of personhood obliquely. The next chapter will address social practices and institutions through the framework of DST in a way consistent and continuous with what has been expressed here. The chapter will end with a concise construction of Dynamic Embodied Agency, drawing upon all of the resources hitherto.

Chapter 5 – Social Practices and Institutions as Dynamic Systems

This chapter's purpose is to bring together the rather motley assortment of resources, concepts, theoretical frameworks, and historical approaches presented thus far into a sketch of an explanatory account that synoptically treats: a metaphysics of agency, a phenomenological epistemology, and a social theory of practices and institutions. The result should bring into view (however blurry) a way of understanding persons and practices as birds of a[n ontological] feather. The culminating point is to show that the conceptual resources of Dynamic Systems Theory prove a potent explanatory basis for thinking of persons and practices in concert – an undertaking not often treated under one explanatory framework.

5.1 Social Practices and Dynamic Embodied Agents Co-Constitute One Another

A Dynamic Embodied Agent is an agent that differentially interacts with the surrounding environment by employing its body in the appropriate ways. What makes a bodily employment appropriate or inappropriate is the degree to which the activities undertaken in particular instances contribute to or detract from the maintenance of the embodied system. As an embodied system, a DEA is sensitive to various aspects of the

surrounding environment. These sensitivities can be thought of abstractly; they needn't exclusively involve sensory apparatuses, they can also be cognitive apparatuses, social apparatuses, and technological or infrastructural apparatuses[404] – anything that can be deployed in a way that constitutively factors into the "calculus" of recursive self-maintenance. As far-from-equilibrium, open, recursively self-maintenant systems capable of error-detection and flexible learning, Dynamic Embodied Agents develop and learn modes and strategies for differential and highly interactive coping with particular situations through creative processes of detecting and recognizing patterns that can contribute to recursive self-maintenant action. These processes may be instantiated in any number of ways, but a helpful analogy is the evolutionary process of natural selection.

At its most basic, natural selection is a set of processes whereby favorable characteristics are selectively retained through the gradual and interrelated processes of reproduction (sexual reproduction, for metazoans), genetic mutation and recombination, and a host of other "epigenetic" factors. What counts as "favorable" is, like self-maintenance, a matter of propagation. That which is capable of persisting over time is favorable; while that which is not, is not[405]. Favorable characteristics are those that contribute to a species' continued existence as determined through the specific contingent processes of selection at work on the particular units of selection (whether they are phenotypic or genotypic individuals; whether they're organisms or genes).

Likewise, we may consider the patterns that are recognized by recursively self-

---

[404] Walter Ong has termed such apparatuses "technologies". Cf. Ong, W. (1982). *Orality and Literacy: The Technologizing of the Word*. Routledge, New York.
[405] Though it is important to acknowledge that even if a characteristic isn't favorable—so long as its lack of favorability doesn't actively detract from propagation in the appropriate ways—such a characteristic may still be retained.

maintenant systems to be analogous to the genetic mutations or recombinations that alter

the selection units. An alteration that produces a pattern "favorable" to self-maintenance

is selectively retained, while patterns detrimental to self-maintenance are gradually culled

out. Those patterns that do not greatly adversely affect self-maintenance may be retained,

or not. Thus – as with the processes of natural selection – it is possible for a recursively

self-maintenant system to recognize and operate using patterns that do not actively

contribute to self-maintenance. In fact, the retention of maintenance-neutral patterns may

itself help contribute to creative processes of recombination that may increase system

plasticity with respect to self-maintenance[406].

Nothing important hinges on the correctness of this model of pattern recognition

and retention. Different recursively self-maintenant systems will have different modal

types of interfaces with their local environments; just as different species have different

potential modal interactions with their environments. For example, in order to find their

ways around in the world bats echolocate, homing pigeons are sensitive to combined

magnetic field and aroma profiles, ants scaffold the world with chemical paths, etc..

These various modal interactions constrain and afford different available possible

patterns for their respective (organismal) systems to recognize and utilize. What is

important here is that the Dynamical Systems model provides a framework through

which the mechanisms of pattern recognition and differential actions of embodied

systems (whatever they may be) can be expressed and explained.

Over time, DEAs hone bodily habits through the highly-interactive world-

---

[406] That is, the retention of maintenance-neutral patterns may itself be a favorable pattern with respect to self-maintenance. The evolutionary analogue to this hypothesis is that increased species diversity increases species robustness.

involving processes they develop through their self-maintenant copings. These habits needn't optimize, nor need they even contribute to self-maintenance. Their only necessary constraint is that they not overtly cause system dissipation[407]. The embodied skills and habits developed by DEAs are behavioral attractors—relatively stable, though aperiodic, nonlinear patterns of activity resistant to perturbation. These behavioral attractors are constrained by the "parameters" of available affordances and "variables" of contingent and dynamic situational factors. Again, for DEAs these situational factors and affordances crucially include *social relations*[408].

How a DEA develops or constructs the self that functions normatively as the basis for *self*-maintenance is defined in part by its perception of itself in relation to general and particular others – that is, its *social* perception. These general relations are exhibited through the phenomenal sense of one's "belonging" or inclusion as a member in a social group. The particular relations are enacted through concrete agent interactions— specifically those whose situations dictate multiple agents acting in concert with each other to accomplish a shared outcome[409]. Thus, the habits and skills developed by DEAs are simultaneously oriented both "inward" toward the self, and doubly "outward" toward general and particular others. Embodied habits and skills are therefore ineliminably social.

Further, since DEAs have myriad open and dynamic relations to general and

---

[407] Even self-destructive behavior can be habitualized so long as the system remains intact (and far-from-equilibrium). If we consider living organismal systems the point is just that nothing that kills the organism can be habitualized. Anything that sends far-from-equilibrium open system into equilibrium cannot be habitualized.

[408] In part because DEAs are constitutively identified through their relations with others.

[409] The *outcome* is shared, but its significance need not be the same for the agents involved in the joint activity.

particular others, the embodied habits and skills that they develop can be causally

influenced and shaped through these very relations. A DEA's social relations constitute a

genuine source of perceptual constraints and affordances. Consequently, social practices

can factor causally into a DEAs development of embodied skills and habits[410].

But what exactly are social practices? In Ch. 2 I discussed this issue with some

seriousness, comparing Turner's, Gallagher's and Schatzki's theories. I concluded that

social practices are regularities of behavior detected and detectable by DEAs[411]; and that

DEAs normatively "attune" themselves to the practices and styles of practice exhibited

by particular others in concrete encounters. This may happen at the pre-noetic, non-

discursive "level" of body-schema, or at the thematic, or discursive "level" of body-

image. DEAs may be aware and attentively attuned to the practices and styles they

encounter, or pre-reflectively unaware of those practices and styles, yet still perceptually

attuned to them. This process of "attunement" works both to triangulate a sense of self

through relating with others, as well as to emulate skills and habits that may be of

potential use in self-maintenance. Thus, "attunement" is a positively reinforcing (or

"rectifying") attractor whereby DEAs are continuously socialized and sociality is

continuously constituted.

---

[410] Recall, Turner's conclusion in *The Social Theory of Practices* that embodied habits and skills should replace the "stop-gap" incoherent notion of causally efficacious social practices. My contention is that social practices are enacted through embodied habits and skills, but further that because DEAs "attune" themselves to social norms and practices these habits and skills are causally affected by social relations to and with general and specific others. This may be consistent with Turner's position, but it is unclear from the texts (*The Social Theory of Practices*, and *Brains/Practices/Relativism*).

[411] Perhaps non-consciously, at the level of *body-schema*.

5.2   *Social Practices and Institutions are DEAs*

It should be clear by now that Dynamic Embodied Agents and social practices can and should both be considered with recourse to the conceptual framework availed by dynamical systems theory. In addition, I hope to have shown in the previous chapter that DEAs are consistent with most full-bodied construals of *persons*. Understanding both DEAs (or persons) and social practices on the same ontological continuum[412] is helpful for two reasons: it establishes a continuous explanatory link between a naturalistic construction of DEAs[413] (so, persons) and social practices; which in-turn helps us to understand each of these domains better. However, one might be skeptical about my claim that social practices and institutions are *themselves* Dynamic Embodied Agents. Are social practices – like waving goodbye, thanking someone for a gift, and cooking food – really far-from-equilibrium, open, recursively self-maintenant embodied systems capable of differentially responding to available constraints, detecting error, and flexible learning? In this very paragraph it also seems I have even conflated DEAs with persons – does this mean that I'm claiming that social practices are also persons?

First it is important to distinguish DEAs from persons. Not all DEAs are persons, but all persons are DEAs: dynamic embodied agency is necessary but not sufficient for personhood. In addition to definitive characteristics of dynamic embodied agency, persons also possess the distinguishing capabilities of conceiving themselves temporally, perhaps thematically structured as an ongoing narrative. Persons are self-aware, perhaps

---

[412]As far-from-equilibrium, open, recursively self-maintenant systems capable of error detection and flexible learning.

[413] See my discussion of this in considering Richard Campbell's process-based ontology in chapter 3.

in many conceptual and non-conceptual ways. Persons are capable of and socially

beholden to morally responsible action. Some of these characteristics *may* also apply to

certain social practices and institutions. If *all* of them do, then it would not be hyperbolic

to ascribe personhood to those practices and institutions. However, to claim that all social

practices and institutions exhibit these characteristics would be highly misleading.

Consider the social institution that is an American university. Richard Campbell

helpfully writes:

> As I used to tell my students, it is the university which admitted them as
> students, and which might eventually confer on them a degree. As an
> individual, I cannot perform those actions, not because someone else performs
> them, but because no individual person does. Even when I was the one whose
> role it was to sign the relevant documents, that was not something I did as an
> individual person. I was exercising a role, a function that only makes sense,
> and only has validity, because it derives from the structure and dynamics of
> the institution.[414]

It is the university itself that possesses the capacity to confer degrees upon students, not

any individual. The university is mereologically comprised of functional roles (such as

"Dean of the College of Arts and Sciences", "Registrar", "Bursar", and many others); as

well as the persons who instantiate those roles (e.g. Eric Eisenberg *in his capacity as*

Dean of the College of Arts and Sciences, or Eric Eisenberg *in his capacity as* professor

in the Department of Communications); the physical resources of the university

(campuses, classrooms, labs, whiteboards, lab equipment, the paper on which degrees are

printed, etc.); and also the established social and cultural practices, institutions, and

conventions (e.g. the classroom setting, the role of professor, the student-teacher

relationship, norms of professionalism, the "school year", etc.). Without all of these

---

[414] Campbell, Richard. (2009) "A Process-Based Model for an Interactive Ontology," *Synthese*: 166. p. 475.

interworking, overlapping parts causally impinging on each other in a relatively stable (though far-from-equilibrium[415]) way, the university would cease to function as a university. Its identity as a university depends on the well-functioning of its parts; but it is not merely the sum of those well-functioning parts, it possesses and bestows its own capabilities over and above what those parts can contribute individually. As a result of the particular, dynamic, complex set of relations enacted by the functional parts that comprise a university, it is capable of *emergent* action (in this case, conferring a degree). This is a genuine case of agency. This agency is *dynamic* and *embodied* because it is a far-from-equilibrium, open, recursively self-maintenant system whose actions are contingent upon the constraints and affordances in the local physical, social, and cultural environment affecting the execution of its processes.

We may ask, however, whether the university system is capable of error-detection and flexible learning. Because it incorporates mechanisms of governance among the dynamic processes of systemic control; and because the processes of deliberations address and handle chaotic contingencies that arise from both within the systems workings (e.g. the faculty union attempting to renegotiate their contracts) and outside influences (e.g. the values of faculty 403Cs plummeting); the university system involves the appropriate kinds of feedback mechanisms for both error-detection and flexible learning. The university is able to detect internal and external conditions and differentially respond to them in order to self-maintain. A university is a Dynamic Embodied Agent. But is it a *person*?

---

[415] In order to operate a university depends on a continual in- and out-flow of new resources, including persons, physical resources, dynamic social relations, etc.

This is often a difficult question to assess. Certainly universities are taken by both individuals and other social institutions (e.g. the U.S. legal system) to possess moral culpability and agency. Universities can act, and can act either in accordance with or in violation of societal norms. So the personhood conditions of agency and moral responsibility do not cause any pause. Universities can also unquestionably possess thematic narratives in the sense that their histories and mission statements establish a context in light of which the progress of the University can be measured. But who does the measuring? In considering the temporality and narrativity of persons, narrative unity was assessed variously by the self for which that unity was in question, as well as intersubjective others. Can the university self-reflectively measure its own narrative progress? In order to do so, it seems, the university must possess some form of self-awareness; and this is where *our* assessment of the university's personhood becomes strained[416].

It is important to consider the context of the question being asked. We are now wondering whether we should consider universities to be self-aware. We are asking this question from our own perspectives as DEAs, but more specifically as epistemically constrained persons. When we ask of ourselves whether we are self-aware, we can be minimally satisfied by a *cogito*-like response: If I can ask whether I'm self-aware, I can be certain that I am. However we do not have the same kind of epistemic access to others – including social institutions in which we may (or may not) directly participate (e.g. by fulfilling some functional role). The more general "problem of other minds" seems to

---

[416] For a similar observation: Cf. Dewey, J. (1926). "The Historic Background of Corporate Legal Personality" *Yale Law Journal* 35:655.

arise: How can I determine whether *any* others are self-aware? What are the satisfaction criteria or possible evidentiary bases?

The phenomenological approach has worked to dissolve the epistemic *gulf* between one's self and others. One's sense of self is developed from a more primordial and epistemically direct interaction with others. What it means to be self-aware is a matter of having developed the appropriate social relationships with general and particular others. It is a matter of being able to locate one's self socially. Can a university locate itself socially? It is difficult to say without the appropriate epistemic access. But we can turn the question around – in one's phenomenological assessment of another's personhood is the university relevantly different than another individual person on this score? Can I *take* the university *as* another subject—as another *person*? If so, then we should be at least as comfortable with claiming that the university is a person as we are in claiming that of each other. More likely (as I alluded to in chapter 2) our assessment of subjectivity or personhood comes in degrees – I assess others as more or less persons as a result of my ability to identify and relate to them socially.

Because my self-perception is the product of my concrete relations with particular others, *which always occurs in contexts of already established and enacted social practices and institutions*; a large part of this assessment is shaped by my self-perception which itself derives from the embodied significance of social practices. Thus, the question becomes whether the person, social practice, or institution being assessed is sufficiently similar[417] to those persons, practices, and institutions that have been

---

[417] The sufficiency of similarity can be qualitatively assessed by the affective phenomenological character of experience. Something is sufficiently similar if it produces the feeling in the assessor of being "at

formative in my development. If so, once again we should be comfortable in saying that

social institutions can be persons. If not, not.

home" with it. Again, this is likely a feeling that admits of degrees.

252

References

Adams, F., & Aizawa, K. (2001). "The Bounds of Cognition," *Philosophical Psychology*, 14(1): 43–64.

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Malden, MA: Blackwell Pub.

Anaxagoras, & Sider, D. (1981). *The fragments of Anaxagoras*. Meisenheim am Glan: Hain.

Anisfeld, M., Turkewitz, G., & Rose, S. A. (2001). "No compelling evidence that newborns imitateoral gestures". *Infancy*, 2, 111–122.

Aristotle, & McKeon, R. (1941). *The basic works of Aristotle*. New York: Random House.

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge & K. Paul.

Armstrong, D. M. (1971). "Meaning and Communication". *Philosophical Review* 80 (4):427-447.

Armstrong, D. M. (1978): *Nominalism and Realism: Universals and Scientific Realism, vol.1*, Cambridge University Press, Cambridge.

Armstrong, D. M. (1981). "The Causal Theory of the Mind". In *The Nature of Mind and Other Essays*. Cornell University Press.

Baier, A. C. (1981). "Cartesian Persons". *Philosophia* 10 (3-4):169-188

Baker, L. R. (1987). *Saving belief: A critique of physicalism*. Princeton, N.J: Princeton University Press.

Bánáthy B. H.. (2007) *A Taste of Systemics*, The Primer Project.

Bánáthy, B (1996) *Designing Social Systems in a Changing World*. Plenum, New York.

Barthes, R. (1982). *The empire of signs* (R. Howard, Trans.). New York: Hill and Wang.

Bateson, G. (1973). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. St Albans, Australia: Paladin.

Benjamin, J. (1988). *The bonds of love: Psychoanalysis, feminism, and the problem of domination*. New York: Pantheon Books.

Bergson, H., & Pogson, F. L. (1910). *Time and free will: An essay on the immediate data of consciousness*. London: S. Sonnenschein & co., lim.

Bermúdez, J. L. (1998). *The paradox of self-consciousness*. Cambridge, Mass: MIT Press.

Bermúdez, J. L. (2001). "Nonconceptual Self-Consciousness and Cognitive Science". *Synthese* 129 (1):129 - 149.

Bermúdez, J. L. (2003). *Thinking without words*. Oxford: Oxford University Press.

Bertalanffy, L. . (1968). *General system theory: Foundations, development, applications*. New York: G. Braziller.

Bickhard, M. (1993). Representational Content in Humans and Machines. *Journal of Experimental and Theoretical Artificial Intelligence*, 5, 285-333.

Bickhard, M. (2009). "The Interactivist Model". *Synthese* 166 (3):547 - 591.

Bishop, R. C. (2009). "Chaos", *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition),* Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/chaos/>.

Block, N. and Fodor, J. (1972). "What Psychological States Are Not". *Philosophical Review*, 81: 159–181.

Bojowald, M. (2007). "What happened before the Big Bang?". *Nature Physics* 3 (8): 523–525

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, Mass: Harvard University Press.

Brentano, F., In Kraus, O., & In McAlister, L. L. (1973). *Psychology from an empirical standpoint*. London: Routledge and Kegan Paul.

Bruner, J. (1991). "The Narrative Construction of Reality". *Critical Inquiry* 18 (1): 1–21.

Buber, M., & Kaufmann, W. A. (trans.) (1970). *I and Thou*. New York: Charles Scribner's Sons.

Buller, D. (ed.) (1999) *Function, Selection, and Design*. Albany, NY: SUNY Press.

Burge, Tyler (1979). "Individualism and the Mental," in French, Uehling, and Wettstein (eds.) *Midwest Studies in Philosophy*, IV, Minneapolis: University of Minnesota Press, pp. 73–121.

Campbell, D. T. (1974). Evolutionary Epistemology. In P. A. Schilpp (Ed.) *The Philosophy of Karl Popper*. (413-463). LaSalle, IL: Open Court.

Campbell, D. T. (1990). Levels of Organization, Downward Causation, and the Selection-Theory Approach to Evolutionary Epistemology. In Greenberg, G., & Tobach, E. (Eds.) *Theories of the Evolution of Knowing*. (1-17). Hillsdale, NJ: Erlbaum.

Campbell, R. (2009) "A Process-Based Model for Interactive Ontology", *Synthese*, 166: 3, pp. 453-477.

Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.

Carruthers P, Smith P. K. (1996) *Theories of theories of mind*. Cambridge University Press, Cambridge, UK.

Carruthers, P. (1996). *Language, thought and consciousness: An essay in philosophical psychology*. Cambridge [England: Cambridge University Press.

Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.

Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford: Clarendon Press.

Chalmers, D. (1995). "Facing Up to the Problem of Consciousness". *Journal of Consciousness Studies*. 2 (3):200-19.

Chalmers, D. (2003). "Consciousness and its Place in Nature". In Stephen P. Stich & Ted A. Warfield (eds.), *Blackwell Guide to the Philosophy of Mind*. Blackwell.

Chalmers, David. "Consciousness and its Place in Nature" in Stitch & Warfield (ed.'s) *Blackwell Guide to the Philosophy of Mind.* (2002)

Chisholm, R. (1969): "The observability of the self", *Philosophy and Phenomenological Research*, 31, 7–21.

Chisholm, R. M. (1957). *Perceiving: A philosophical study*. Ithaca: Cornell University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.

Chomsky, N. (1971). "The Case Against B. F. Skinner," *New York Review of Books*, 30: 18–24.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Church. A. (1936) "A Note on the Entscheidungsproblem". *Journal of Symbolic Logic*, 1, 40-41.

Church. A. (1936). "An Unsolvable Problem of Elementary Number Theory". *American Journal of Mathematics*, 58, 345-363.

Churchland, P. (1989) "On the Nature of Theories: A Neurocomputational Perspective, in Haugeland, J. (1997). *Mind design II: Philosophy, psychology, artificial intelligence*. Cambridge, Mass: MIT Press.

Churchland, P. M. (1979). *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press.

Churchland, P. M. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*. Cambridge, MA: MIT Press.

Clark, A. & Chalmers, D. (1998). "The Extended Mind". *Analysis* 58 (1):7-19.

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

Cosmides, L. & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality*. Cambridge, MA: The MIT Press.

Cummins, R. (1975). "Functional Analysis." *Journal of Philosophy* 72: 741.765

De Jaegher, H. & Di Paolo, E. (2007). "Participatory Sense-Making". *Phenomenology and the Cognitive Sciences* 6 (4):485-507.

De Landa, M. (1997). *A thousand years of nonlinear history*. New York, NY: Zone Books.

De Landa, M. (2002). *Intensive science and virtual philosophy*. London: Continuum.

Deleuze, G. (1983). *Nietzsche and philosophy*. New York: Columbia University Press.

Deleuze, G. (1988). *Bergsonism*. New York: Zone Books.

Deleuze, G. (1994). *Difference and repetition*. New York: Columbia University Press.

Deleuze, G., & Guattari, F. (1987). *A thousand plateaus: Capitalism and schizophrenia*. Minneapolis: University of Minnesota Press.

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

Dennett, D. C. (1988). "Evolution, Error and Intentionality". In: Y. Wilks and D. Partridge, (eds) *Sourcebook on the Foundations of Artificial Intelligence,* New Mexico University Press

Dennett, D. C. (1988). "Quining Qualia". In: Marcel, A and Bisiach, E (eds), *Consciousness in Contemporary Science*, 42-77. New York, Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Dennett, D. C. (1992). "The Self as a Center of Narrative Gravity". In Frank S. Kessel, P. M. Cole & D. L. Johnson (eds.), *Self and Consciousness: Multiple Perspectives*. Lawrence Erlbaum.

Dennett, D. C. (1997). "Consciousness in Human and Robot Minds". In M. Ito, Y. Miyashita & Edmund T. Rolls (eds.), *Cognition, Computation and Consciousness*. Oxford University Press.

Dennett, D. C. et. al. (eds) (2007). *Neuroscience and philosophy: Brain, mind, and language*. New York: Columbia University Press.

Derrida, J. (1976). *Of grammatology*. Baltimore: Johns Hopkins University Press.

Descartes, R., & Cottingham, J. (trans.) (1986). *Meditations on first philosophy: With selections from the Objections and Replies*. Cambridge.

Descartes, R., Kennington, R., & Kraus, P. (trans.) (2012). *Discourse on Method*. Newburyport: Focus Publishing/R. Pullins Co.

Descartes, R., Miller, V. R., & Miller, R. P. (1983). *Principles of philosophy*. Dordrecht, Holland: Reidel.

Dretske, F. (1988). Explaining *Behavior: Reasons in a World of Causes*. MIT Press.

Dreyfus, H. L. (1991). *Being-in-the-world: A commentary on Heidegger's Being and time, division I*. Cambridge, Mass: MIT Press.

Einstein, A. (1907) "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen" (On the relativity principle and the conclusions drawn from it), *Jahrbuch der Radioaktivität und Elektronik* 4, 411–462.

Feyerabend, P. (1975). *Against method: Outline of an anarchistic theory of knowledge*. London: NLB.

Flanagan, O. J. (1992). *Consciousness reconsidered*. Cambridge, Mass: MIT Press.

Flanagan, O. J. (1998). *Self expressions: Mind, morals, and the meaning of life*. New York: Oxford University Press.

Flanagan, O. J. (2007). *The really hard problem: Meaning in a material world.* Cambridge, Mass: MIT Press.

Fodor, J. A. & Pylyshyn, Z. W. (1988). "Connectionism and Cognitive Architecture: A Critical Analysis" in S. Pinker and J. Mehler, eds., *Connections and Symbols*. Cambridge, Massachusetts: MIT Press.

Fodor, J. A. (1968). Psychological Explanation. New York: Random House.

Fodor, J. A. (1974). "Special Sciences, or Disunity of Science as a Working Hypothesis". *Synthese* 28 (2):97-115.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Fodor, J. A. (1980), "Methodological Solipsism Considered as a Research Strategy in Cognitive Science," *Behavioral and Brain Sciences,* 3: 63-73.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, Mass: MIT Press.

Fodor, J. A. (1987). *Psychosemantics*, Cambridge MA: MIT Press.

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

Fodor, J. A. (1996). *Concepts: Where cognitive science went wrong*. Oxford: Clarendon Press.

Frankfurt, H. G. (1988). *The importance of what we care about: Philosophical essays*. Cambridge, England: Cambridge University Press.

Franses, P. H., & Dijk, D. (2000). *Nonlinear time series models in empirical finance*. Cambridge: Cambridge University Press.

Frege, G. 1974, *The Foundations of Arithmetic*, J. L. Austin (trans.), Oxford: Basil Blackwell.

Gallagher, S. (2000)." Philosophical Conceptions of the Self". *Trends in Cognitive Sciences* 4 (1):14-21

Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Clarendon Press.

Gallagher, S. (2006). "Where's the action?: Epiphenomenalism and the problem of free will". In W. Banks, S. Pockett, and S. Gallagher. *Does Consciousness Cause Behavior? An Investigation of the Nature of Intuition* (109-124). Cambridge, MA: MIT Press.

Gallagher, S. (2009) "Two Problems of Intersubjectivity", *Journal of Consciousness Studies,* 16, No. 6-8.

Gallagher, S., & Crisafi, A. (2009). Mental Institutions. *Topoi* 28 (1):45-51.

Gallagher, S., & Meltzoff, A. N. (1996). "The Earliest Sense of Self and Others: Merleau-Ponty and Recent Developmental Studies". *Philosophical Psychology* 9 (2):211-33.

Gallagher, S., & Shear, J. (1999). *Models of the self*. Thorverton, UK: Imprint Academic.

Gallagher, S., & Watson, S. (2004). *Ipseity and alterity: Interdisciplinary approaches to intersubjectivity*. Mont-Saint-Aignan: Publ. de l'Université de Rouen.

Gallagher, S., & Zahavi, D. (2008). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. London: Routledge.

Gallese, V. (2001) "The 'shared manifold' hypothesis: from mirror neurons to empathy" *Journal of Consciousness Studies*, 8.

Gallese, V. (2005). "Embodied simulation: from neurons to phenomenal experience". *Phenomenology and the Cognitive Sciences*, 4: 23-48

Gallese, V. (2005). 'Being like me': Self-other identity, mirror neurons, and empathy. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science*, Vol 1 (pp. 101-118). Cambridge, MA, US: MIT Press.;

Gamut, L. T. F. (1991). *Logic, language, and meaning*. Chicago: University of Chicago Press.

Geach, P. T. (1957). *Mental acts: Their content and their objects*. London: Routledge & Paul.

Gergely, G. (1995). "The role of parental mirroring affects in early psychic structuration". Paper presented at the IPA's 5th Conference on Psychoanalytic Research, London, England.

Gergely, G. (2004). "The development of understanding self and agency". In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 26–46). Malden, MA: Blackwell.

Gershenfeld, N. (1999*), The Nature of Mathematical Modeling* (1st ed.), New York: Cambridge University Press, New York.

Gibson, J. J. (1972). "A Theory of Direct Visual Perception". In J. Royce, W. Rozenboom (Eds.). *The Psychology of Knowing*. New York: Gordon & Breach.

Gibson, J. J. (1977). "The Theory of Affordances" (pp. 67-82). In R. Shaw & J. Bransford (Eds.). *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Lawrence Erlbaum.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Goldstein, E. B. (2001). *Blackwell handbook of perception*. Oxford, UK: Blackwell.

Gopnick, A. (1993) "How we think about our minds: the illusion of first-person intentionality," *Behavioral and Brain Sciences* 16:1

Grammont, F., Legrand, D., & Livet, P. (2010). *Naturalizing intention in action*. Cambridge, Mass: MIT Press.

Granger, C. W. J., & Teräsvirta, T. (1993). *Modelling nonlinear economic relationships*. Oxford England: Oxford University Press.

Guignon, C. (2012). "Becoming a person: Hermeneutic phenomenology's contribution"*, New Ideas in Psychology*, Volume 30, Issue 1, Pages 97-106.

Harding, S. G., & Hintikka, M. B. (1983/2003). *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology, and philosophy of science 2nd. Ed.*. Dordrecht, Holland: D. Reidel.

Haugeland, J. (1985) *Artificial Intelligence: The Very Idea*. Cambridge. MIT Press.

Haugeland, J. (1993). "Mind Embodied and Embedded". In Yu-Houng H. Houng & J. Ho (eds.), *Mind and Cognition: 1993 International Symposium*. Academica Sinica.

Haugeland, J. (1998) "The Mind: Embodied and Embedded". *Having thought: Essays in the metaphysics of mind.* Cambridge, Mass: Harvard University Press.

Hegel, G. W. F., Miller, A. V., & Findlay, J. N. (1977). *Phenomenology of spirit*. Oxford: Clarendon Press.

Heidegger, M. (1967). *Sein und Zeit*. Tübingen: M. Niemeyer.

Heidegger, M., Macquerrie, & Robinson (trans.) (1962). *Being and time*. New York: Harper.

Hellman, G. & Thomson, F. (1975). "Physicalism: Ontology, Determination and Reduction". *Journal of Philosophy* 72 (October):551-64

Honneth, A. (1996) *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Polity Press.

Honneth, A. (2007) *Reification: A Recognition-Theoretical View*. Oxford University Press.

Hooker, C. A. (2011). *Philosophy of complex systems*. Oxford, U.K: North Holland.

Hull, D.L. 1970. "Systemic Dynamic Social Theory." *Sociological Quarterly*, Vol. 11, Issue 3, pp. 351–363.

Hume, D., Selby-Bigge, L. A., & Nidditch, P. H. (1978). *A treatise of human nature*. Oxford: Clarendon Press..

Hurley, S., & In Chater, N. (2005). *Perspectives on Imitation: Vol. 1: From Neuroscience to Social Science - Mechanisms of Imitation and Imitation in Animals*. (MIT CogNet.) Cambridge: MIT Press.

Husserl, E. (1966). *Cartesian meditations: An introduction to phenomenology*. The Hague: M. Nijhoff.

Husserl, E. (1988 [1931]) *Cartesian Meditations*, trans. D. Cairns, Dordrecht: Kluwer

Husserl, E., & Gibson, W. R. B. (1931). *Ideas: General introduction to pure phenomenology*. London: G. Allen & Unwin.

Husserl. E. (1973) *Cartesianische Meditationen und Pariser Vorträge,* Husserliana I (Den Haag: M. Nijhoff).

Husserl, E., & Brough, J. B. (1991). *On the phenomenology of the consciousness of internal time* (1893-1917). Dordrecht: Kluwer Academic Publishers.

Iacoboni, M. (2009) "Imitation, Empathy, and Mirror Neurons", *The Annual Review of Psychology*, 60. 653-670.

Iacoboni, M. (2009). *Mirroring people: The science of empathy and how we connect with others*. New York, N.Y: Picador.

Iacoboni, M., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, 3, e79.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (n.d.). "Grasping the Intentions of Others with One's Own Mirror Neuron System". *Public Library of Science*.

Jones, S.S. (1996). "Imitation or exploration? Young infants' matching of adults' oral gestures". *Child Development*, 67, 1952–1969.

Jordan, D. W.; Smith, P. (2007*). Nonlinear Ordinary Differential Equations* (fourth ed.). Oxford Univeresity Press.

Kampis, G. (1991). *Self-modifying systems in biology and cognitive science: A new framework for dynamics, information, and complexity*. Oxford: Pergamon Press.

Kampis, G. (1999). "The Natural History of Agents", in: Gulyás, L., Tatai, G., Váncza, J. (ed.): *Agents Everywhere*, Springer, Budapest, pp. 24-48.

Kant, I., ed: Guyer, P. A. (2000). *Critique of the power of judgment*. Cambridge, UK: Cambridge University Press.

Kant, I., ed's: Guyer, Wood, P. A. (1999). *Critique of Pure Reason*. Cambridge.

Kim, C.-J., & Nelson, C. R. (1999). *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications*. Cambridge, Mass: MIT Press.

Kim, J. (2006). "Emergence: Core ideas and issues," *Synthese*, 151(3): 347–354.

Kornblith, H. (ed.) (2001). *Epistemology: Internalism and Externalism*, Blackwell Press.

Kripke, S. (1972). *Naming and Necessity*, Oxford: Blackwell.

Kripke, S. (1979). "A puzzle about belief," in *Meaning and Use*, edited by A. Margalit. Dordrecht and Boston: Reidel.

Kripke, S. (1982). *Wittgenstein on Rules and Private Language*, Cambridge: Harvard University Press.

Kripke, S. (1982). *Wittgenstein on rules and private language: An elementary exposition*. Cambridge, Mass: Harvard University Press.

Kuhn, T. S. (1996). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lakatos, I., & In Musgrave, A. (1970). *Criticism and the growth of knowledge*. Cambridge.: University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

Laszlo, E. (1996). *The Systems View of the World*. Hampton Press, NJ

Legrand, D. & Grammont F. (2005). A Matter of Facts. *Phenomenology and the Cognitive Sciences* 4 (3):249-257.

Lévinas, E. (1974). *Autrement qu'être, ou, Au-delà de l'essence*. La Haye: M. Nijhoff.

Lévinas, E. (1983). *Le temps et l'autre*. Paris: PUF.

Lettvin et al. (1959). "What the Frog's Eye Tells the Frog's Brain", *in Proceedings of the Institute of Radio Engineers*. pp.1940-51.

Livet, P. (2002). "Emotions, Revision, and the Explanation of Actions". *European Review of Philosophy* 5:93-108.

Livet, P. (2006). "Identities, Capabilities and Revisions". *Journal of Economic Methodology* 13 (3):327-348.

Locke, J. (1959). *An essay concerning human understanding*. New York: Dover Publications.

Louise-Gill, M. (1989) *Aristotle on Substance: The Paradox of Unity*. Princeton University Press.

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20, 2, 130-141.

Lycan, W. G. (1987). *Consciousness*. Cambridge, Mass: MIT Press.

Lycan, W. G. (1996). *Consciousness and experience*. Cambridge, Mass: MIT Press.

Lynch, W. T. (2005) "The Ghost of Wittgenstein: Forms of Life, Scientific Method, and Cultural Critique". *Philosophy of the Social Sciences*, 35:139. p. 146.

MacIntyre, A. C. (1984/1989). *After virtue: A study in moral theory*. Notre Dame, Ind: University of Notre Dame Press.

Malcolm, N. (1995) *Wittgensteinian Themes*.Cornell University Press

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht, Holland: D. Reidel Pub. Co.

Mauss, M. (2001). *The gift: The form and reason for exchange in archaic societies*. London: Routledge.

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition : Volume 1: Foundations*. Cambridge: MIT Press.

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition : Volume 2: Psychological and biological models*. Cambridge: MIT Press.

McGeer, V. (2001). "Psycho-practice, psycho-theory and the contrastive case of autism: How practices of mind become second-nature". *Journal of Consciousness Studies*, 8, 109–132.

McIlroy, D. (1969). "Mass produced software components". *Software Engineering: Report of a conference sponsored by the NATO Science Committee*, Garmisch, Germany, 7-11 Oct. 1968. Scientific Affairs Division, NATO. p. 79.

Meltzoff, A. N. and Moore, M. K. (1977). "Imitation of Facial and Manual Gestures by Human Neonates", *Science*, 198, 75-78.

Menary, R. (2010). *The extended mind*. Cambridge, Mass: MIT Press.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Merleau-Ponty, M., & Landes, D. A. (2012). *Phenomenology of perception*. Abingdon, Oxon: Routledge.

Merleau-Ponty, M., & Lefort, C. (1964). *Le visible et l'invisible*. Paris: Gallimard.

Merleau-Ponty, M., & Smith, C. (1962). *Phenomenology of perception*. New York: Humanities Press.

Millikan, R. G. (1986). "Thoughts Without Laws: Cognitive Science Without Content," *Philosophical Review*, XCV, pp.47-80.

Millikan, R. G. (1989). "An ambiguity in the notion of function." *Biology and Philosophy* 4: 172-176

Minsky, M. L. (1968). *Semantic information processing*. Cambridge, Mass: MIT Press.

Minsky, M. L. (1974). *A framework for representing knowledge*. Cambridge: Massachusetts Institute of Technology A.I. Laboratory.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N.J: Prentice-Hall.

Newell, A., & Simon, H. A. (1976), "Computer Science as Empirical Inquiry: Symbols and Search", *Communications of the ACM*, 19

Noë, A. (2004). *Action in perception*. Cambridge, Mass: MIT Press.

Noë, A. (2009). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness*. New York: Hill and Wang.

Norman, D. A. (1999). "Affordances, Conventions and Design." *Interactions* 6(3):38-43, May 1999, ACM Press.

Ong, W. (1982). *Orality and Literacy: The Technologizing of the Word*. Routledge, New York.

Overgaard, S. & Zahavi, D. (2009) "Understanding (Other) Minds: Wittgenstein's Phenomenological Contribution" In E. Zamuner and D. K. Levy (eds.), *Wittgenstein's Enduring Arguments*. London: Routledge, pp. 60-86.

Parfit, D. (1984). *Reasons and persons*. Oxford [Oxfordshire: Clarendon Press.

Parfit, D., & Scheffler, S. (2011). *On what matters*. Oxford: Oxford University Press.

Parsons T., Bales, R. (1955). *Family Socialization and Interaction Process*. Routledge, London.

Parsons, T. (1964). *Social structure and personality*. New York: Free Press of Glencoe.

Parsons, T. (1991). *The social system*. London: Routledge.

Parsons, T., & Shils, A., (eds) (1976) *Toward a General Theory of Action*, Harvard University Press, Cambridge.

Petit, J-L. (1999). "Constitution by Movement: Husserl in Light of Recent Neurobiological Findings". In *Naturalizing Phenomenology*. Stanford: Stanford University Press.

Petit, J-L. (1999). *Naturalizing Phenomenology*. Stanford: Stanford University Press.

Pinker, S., & Mehler, J. (1988). *Connections and symbols*. Cambridge, Mass: MIT Press.

Place, U.T. (1956). "Is Consciousness a Brain Process?", *British Journal of Psychology*, 47, pp. 44–50.

Plato, & Fowler, H. N. (1990). *Euthyphro, Apology, Crito, Phaedo, Phaedrus*. Cambridge, Mass: Harvard University Press.

Poellner, P. (2003). "Non-conceptual content, experience and the self," *Journal of Consciousness Studies*, 10 (2): 32–57.

Port, R. F. & van Gelder, T. (Eds.), (1996). *Mind as motion: Explorations in the dynamics of cognition* (pp. 1-43). Cambridge, MA: MIT.

Price, H. (1996). *Time's Arrow & Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press.

Prinz, W. (2005). Experimental approaches to action. In J. Roessler and N. Eilan (Eds.), *Agency and self-awareness* (pp. 165-187). New-York, Oxford University press.

Putnam, H. (1960). "Minds and Machines". In Sidney Hook (ed.), *Dimensions of Mind*. New York University Press.

Putnam, H. (1963*). Analytical Philosophy: Second Series*. Ronald J. Butler (ed.). Blackwell.

Putnam, H. (1967). "The Nature of Mental States". In W.H. Capitan & D.D. Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh University Press.

Putnam, H. (1973). "Meaning and Reference," *Journal of Philosophy* 70, 699-711.

Putnam, H. (1975). "The Meaning of Meaning," *Philosophical Papers, Vol. II : Mind, Language, and Reality*, Cambridge: Cambridge University Press.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, Mass: MIT Press.

Pylyshyn, Z. W., Demopoulos, W., & University of Western Ontario. (1986). *Meaning and cognitive structure: Issues in the computational theory of mind*. Norwood, N.J: Ablex Pub. Corp.

Quine, W. V. O. (1948). "On What There Is". *Review of Metaphysics* 2 (5):21-36.

Quine, W. V. O. (1953). "Two Dogmas of Empiricism". In Darragh Byrne & Max Kölbel (eds.), *From a Logical Point of View*. New York: Harper Torchbooks.

Quine, W. V. O. (1960). *Word and object*. Cambridge: Technology Press of the Massachusetts Institute of Technology.

Rescher, N. (1996). *Process metaphysics: An introduction to process philosophy*. Albany: State University of New York Press.

Rescher, N. (2000). *Process philosophy: A survey of basic issues*. Pittsburgh: University of Pittsburgh Press.

Rosch, E. (1975), "Cognitive representation of semantic categories," *Journal of Experimental Psychology* 104(3): 192-233.

Rosenthal, D. M. (1986). "Two Concepts of Consciousness". *Philosophical Studies* 49 (May):329-59.

Rosenthal, D. M. (1992). "Time and Consciousness". *Behavioral And Brain Sciences* 15 (2):220-221.

Rosenthal, D. M. (2002). "The Higher-Order Model of Consciousness". In Rita Carter (ed.), *Consciousness*. Weidenfeld & Nicolson.

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Oxford University Press.

Rowlands, M. (1999). *The body in mind: Understanding cognitive processes*. Cambridge, U.K: Cambridge University Press.

Rowlands, M. (2006). *Body language: Representation in action*. Cambridge, Mass: MIT Press.

Rowlands, M. (2010). *The new science of the mind: From extended mind to embodied phenomenology*. Cambridge, Mass: MIT Press.

Rumelhart, D. (1989) "The Architecture of Mind: A Connectionist Approach", in Haugeland, J. (1997). *Mind design II: Philosophy, psychology, artificial intelligence*. Cambridge, Mass: MIT Press.

Rupert, R. (2004). "Some Problems for the Thesis of Extended Cognition." *Journal of Philosophy*. 101:389 – 428

Russell, B. (1912/1988). The problems of philosophy. Buffalo, N.Y: Prometheus Books.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson's University Library.

Said, E. W. (1978). *Orientalism*. New York: Pantheon Books.

Salmon, W. (1977). "An at-at theory of causal influence". *Philosophy of Science*. 44:2 215-224.

Sandel, M. J. (1982). *Liberalism and the limits of justice*. Cambridge [Cambridgeshire: Cambridge University Press.

Sartre, J.-P. (1966). *L'être et le néant: Essai d'ontologie phénoménologique*. Paris: Gallimard.

Sartre, J.-P., Barnes, H. E. (trans.) (1992). *Being and nothingness: A phenomenological essay on ontology*. New York: Washington Square Press.

Schatzki, T. R. (1996). *Social practices: A Wittgensteinian approach to human activity and the social*. New York: Cambridge University Press.

Schatzki, T. R. (2010). *The timespace of human activity: On performance, society, and history as indeterminate teleological events*. Lanham, Md: Lexington Books.

Schechtman, M. (1996). *The constitution of selves*. Ithaca, NY: Cornell University Press.

Scheler, M., & Frings, M. S. (1973). *Wesen und Formen der Sympathie: Die deutsche Philosophie der Gegenwart*. Bern: Francke.

Searle, J. R. (1983). *Intentionality, an essay in the philosophy of mind*. Cambridge, Cambridge University Press.

Seibt, J. (2003). *Process theories: Cross-disciplinary studies in dynamic categories*. Dordrecht: Kluwer Academic.

Simon, H. A. (1957). *Models of man: Social and rational; mathematical essays on rational human behavior in society setting*. New York: Wiley.

Simon, H. A. (1969). "The architecture of complexity." *The Sciences of the Artificial* (pp. 192-229). Cambridge, MA: MIT Press.

Smart, J. J. C. (1961). "Further Remarks on Sensations and Brain Processes". *Philosophical Review* 70 (July):406-407.

Smolensky, P (1988). "On the Proper Treatment of Connectionism". *Behavioral and Brain Sciences* 11:1-23.

Smolensky, R. (1989) "Connectionist Modeling: Neural Computation / Mental Connections", in Haugeland, J. (1997). *Mind design II: Philosophy, psychology, artificial intelligence*. Cambridge, Mass: MIT Press.

Stein, Edith. (1983) *On the Problem of Empathy*, trans. Waltraut Stein. ICS Publications.

Strawson, G. (2004/2008) "Against Narrativity". *Ratio* 17, pp. 428–52.

Strawson, G. (2009) *Selves: An Essay in Revisionary Metaphysics*. Oxford University Press, London.

Taylor, C. (1983) "The Concept of a Person". *Social Theory as Practice, The B. N. Ganguli Memorial Lectures 1981*. Delhi, Oxford University Press.

Taylor, C. (1985) *Philosophical Papers: Human Agency and Language*. Cambridge University Press.

Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Cambridge, Mass: Harvard University Press.

269

Taylor, C. (1994) "The Politics of Recognition". *Multiculturalism: Examining the Politics of Recognition*. Ed. Amy Gutmann. Princeton University Press. pp. 25-73.

Taylor, C. (1995) *Philosophical arguments*. Cambridge, MA, Harvard Univ. Press.

Thompkins, E.F. (1990) "A Farewell to Forms of Life". *Philosophy*, Vol. 65, No. 252.

Thompson, E. & Stapleton, M. (2009). "Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories". *Topoi* 28 (1):23-30.

Thompson, E. (2003) "The Problem of Consciousness: New Essays in Phenomenological Philosophy of Mind". *Canadian Journal of Philosophy*, Supplementary Volume 29. University of Alberta Press

Thompson, E. (2007/2010). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, Mass: Belknap Press of Harvard University Press.

Tong, H. (1990). *Nonlinear time series: A dynamical system approach*. Oxford: Clarendon Press.

Towarnicki, F. & Beaufret, J. (1993). *A la rencontre de Heidegger: Souvenirs d'un messager de la Forêt-Noire*. Paris: Gallimard.

Turing, A. M. (1950). "Computing machinery and intelligence". *Mind : a Quarterly Review of Psychology and Philosophy*, 236, 433.

Turing, A.M. (1936). "On Computable Numbers, with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society, series 2*, 42 (1936-37), 230-265.

Turner, S. P. (1994). *The social theory of practices: Tradition, tacit knowledge, and presuppositions*. Chicago: University of Chicago Press.

Turner, S. P. (2002). *Brains/practices/relativism: Social theory after cognitive science.* Chicago: University of Chicago Press.

Varela, F. (1997). "Patterns of Life: Intertwining identity and cognition", *Brain Cognition* , 34:72-87.

Varela, Thompson, Rosch, *The Embodied Mind*: *Cognitive Science and the Human Experience*. MIT Press. 1991.

Welsh, T. (2006): "Do Neonates Display Innate Self-Awareness? Why Neonatal Imitation Fails to Provide Sufficient Grounds for Innate Self- and Other-Awareness", *Philosophical Psychology*, 19:2, 221-238.

Whitehead, A. N., Griffin, D. R., & Sherburne, D. W. (1978). *Process and reality: An essay in cosmology*. New York: Free Press.

Wittgenstein, L., & Anscombe, G. E. M. (2001). *Philosophical investigations*: *The German text, with a revised English translation*. Oxford: Blackwell.

Wittgenstein, L., Anscombe, G. E. M., & Wright, G. H. (1972). *On certainty*. New York: Harper & Row.

Wittgenstein, L., In Anscombe, G. E. M., & In Wright, G. H. (1967). *Zettel*. Berkeley: University of California Press.

Wittgenstein, L., Wright, G. H., Nyman, H., & Luckhardt, C. G. (1980). *Remarks on the philosophy of psychology: Vol. 2*. Oxford: Blackwell.

Wittgenstein, L., Wright, G. H., Nyman, H., & Luckhardt, C. G. (1980) *Remarks on the Philosophy of Psychology, vol. 2*, Oxford: Blackwell.

Worrall, J. (1989). "Structural realism: The best of both worlds?" *Dialectica*, 43: 99–124. Reprinted in D. Papineau, ed., *The Philosophy of Science*, pp. 139–165. Oxford: Oxford University Press

Zahavi, D. (1999). *Self-Awareness and Alterity: A Phenomenological Investigation*, Evanston: Northwestern University Press.

Zahavi, D. (2001). "Beyond Empathy: Phenomenological Approaches to Intersubjectivity". *Journal of Consciousness Studies* 8 (5-7):151-167.

Zahavi, D. (2002). "First-Person Thoughts and Embodied Self-Awareness: Some Reflections on the Relation Between Recent Analytic Philosophy and Cognitive Science". *Phenomenology and the Cognitive Sciences* 1 (1):7-26.

Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the first-person perspective*, Cambridge, MA: The MIT Press.

Zahavi, D. (2012). "Levels of empathy." *SPHS Lecture in Social Phenomenology*. Rochester, USA.