

3-12-2015

The Encultured Mind: From Cognitive Science to Social Epistemology

David Alexander Eck

University of South Florida, mrdavideck@yahoo.com

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>



Part of the [Philosophy Commons](#)

Scholar Commons Citation

Eck, David Alexander, "The Encultured Mind: From Cognitive Science to Social Epistemology" (2015). *USF Tampa Graduate Theses and Dissertations*.
<https://digitalcommons.usf.edu/etd/5472>

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

The Encultured Mind: From Cognitive Science to Social Epistemology

by

David Eck

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

Co-Major Professor: Alexander Levine, Ph.D.
Co-Major Professor: Stephen Turner, Ph.D.
Charles Guignon, Ph.D.
Joanne Waugh, Ph.D.
William Goodwin, Ph.D.

Date of Approval:
March 12, 2015

Keywords: enactivism, participatory sense-making, embodied knowledge, tacit knowledge,
testimonial knowledge

Copyright © 2015, David Eck

Dedication

To Lauren, whose understanding and support lie quietly between the lines of the many pages that follow.

Acknowledgments

It is hard to imagine the existence of the following project—never mind its completion—without the singular help of Alex Levine. His scholarship is as remarkable as it is unassuming, only surpassed by his concern for others and readiness to help. I must express immense thanks for the assistance of Stephen Turner, whose intellectual candor and acuity repeatedly spurred the dissertation forward. I am also especially thankful to Charles Guignon—for his kindness and philosophical subtlety; to Joanne Waugh—for her enlivening conversations, not least those concerning Plato; and to William Goodwin—for his steadfast questioning, which has added valuable perspectives to the project. My committee members have been invaluable mentors—in philosophy and much else. All of the errors in the following—I *assure* you—are mine.

I would like to also thank the USF philosophy community as a whole, from the teachers and staff to my fellow graduate students. A special thanks to Roger Ariew, whose philosophical sensibilities and historical acumen have greatly improved my own. My fellow graduate students have helped me at every turn, most of all Andrew Winters who shepherded me through the entire writing process. I would like to also stress Brian Dunst's contributions, whose own dissertation dispelled the impasse that I found myself at. Nate Draluck and Jared Kinggard, both of whom are veritable oracles, have been instrumental; "It's just one big paper." To name only some of the many others who have supported me: Phil Scuderi, Steve Burgess, Megan Altman, Adam Buben, Casey Rentmeester, Pete Olen, Michael Thompson, Michelle Merritt, Ali Elamin, David Cheely, Christine Wieseler, Sarah Wieten, Dahlia Guzman, Anthony Fernandez, and Carter Hardy. Also,

I want to separately thank those who began graduate school with me and made the transition worthwhile: Brian Mintey, Sacha Greer, Mark Castricone, Emre Keskin, and Greg McCreery. (I hope you didn't expect my acknowledgements to be brief—the following dissertation claims that philosophers have systematically marginalized the social dimensions of knowledge.)

The philosophy department at Lehigh University set me up to succeed. Gordon Bearn taught me more than could be put into words and much else that can. Rick Matthews turned his classes into living Socratic dialogues or group therapy sessions, it was never clear which but it was always transformative. Roslyn Weiss opened up to me the wonders of ancient philosophy. Mark Bickhard taught me the productivity of error. I would be remiss to not also thank Hannah Behrmann, Alex Ganim, John Pettegrew, Seth Moglen, Erin Seeba, Donna Wagner, Olga Stewart, Fareed Awan, Molly Luft, Emily Wettstein, Joey, Jess, Pat, and Colin.

My insatiable curiosity, questioning disposition, and love of conversation is due in no small part to a rich network of hometown friends. Thank you Cutter, Jesse, Eric, Dylan, Paul, JP and many others. For food, wine, literature, and stonelaying, I am eternally indebted to the entire Miele-Wood clan, especially Cutter, Erin, Gloria, Jim, Liz, Bob, Dante, and Marcia.

As with most good stories, mine begins with a singular family experience. My mother's love was only matched by her unflinching example of independence. My older brother Jake's forced impressment into helping with his newspaper route instilled in me the determination and space for contemplation necessary to be a philosopher. Amber, Brandon, Adriana, and Rachel, among other things, remind me of the adolescent injustices I inflicted upon them and the need to compromise without compromising oneself. Thanks also go to Grandmère, Erica, Ralph, the Lee family; and to my father. Chad and Linda have been great friends. And a final thanks to Lauren, who has buoyed me, Jon and Jackie, and the Wagner, Anderson, and Troup families.

Table of Contents

Abstract	iii
Chapter 1: The Testimony View	1
1.0 Introduction	1
1.1 Epistemology's Litmus Test	3
1.2 The Nature and Domain(s) of Testimony	10
1.2.1 Elizabeth Fricker's Partitioning of Testimony's Domain	10
1.2.2 Jennifer Lackey on the Nature of Testimony	13
1.2.3 John Greco's "Friendly" Epistemic Environments	18
1.3 The Transmission Thesis versus The Transmission View	22
1.4 Alvin Goldman's Reliabilism	25
1.5 Goldman's Epistemics Project: Some Major Issues	30
1.5.1 Epistemics: An Overview	30
1.5.2 Epistemics and the Generality Problem	32
1.5.3 Epistemics' Conception of Intelligence and the Basing Relation Problem	34
1.5.4 Epistemics' Conception of Intelligence and the Pursuit of Knowledge	37
1.6 Truth-Oriented Social Epistemology	41
1.6.1 Testimony within Goldman's Social Epistemics	42
1.6.2 Epistemic Paternalism and Epistemics' Two Loci of Control	46
1.6.3 Concluding Thoughts on Truth-Oriented Social Epistemology and Epistemics	50
1.7 Kusch's Communitarian Critique of the Testimony Literature	54
Chapter 2: The Translation View	61
2.0 Introduction	61
2.1 Polanyi on Tacit Knowledge: Imitative Learning within Traditions	63
2.2 Dreyfus and Searle: Tacit Knowledge as a Transferable Objective Precondition	71
2.3 Substitutes for the Normative	85
2.4 The Cognitive Basis for the Translation View: Connectionism and Mirror Neurons	93
2.5 Social Theory: Individualist or Collectivist?	108
Chapter 3: Coordinated Interaction	112

3.0 Introduction.....	112
3.1 From Autopoietic Machines to Sense-Making: The Enactivist Approach to Cognition	114
3.2 Teleology, Adaptivity, and Other Forbidden Things.....	126
3.3 An Enactivist Approach to the Social: Participatory Sense-Making.....	134
3.4 Extended Functionalism?.....	146
3.5 Structural Normativity?	149
Chapter 4: Coordination in Epistemic Communities	161
4.0 Introduction.....	161
4.1 Relating Embodiment and Social Epistemology: Two Ways of Regulating Error.....	163
4.2 Kuhn as a Bridge to Contemporary Social Epistemology	166
4.3 D’Agostino’s Federal Model of Enquiry	172
4.4 Two Disembodied Social Epistemologies	182
4.5 Cogency and the Immanent Normativity of Epistemic Communities	191
Chapter 5: The Encultured Mind	199
5.0 Introduction.....	199
5.1 The Extended Mind, Functionalism, and Their Vanishing Cognitive Subject.....	201
5.2 Uncovering the Adaptationist and Manipulationist Core of the Extended Mind.....	209
5.3 Socializing EM: Insights, Advances, and Some Intractable Difficulties.....	218
5.4 Hutto’s Radical Enactivism: Eliminating Basic Mental Content But At What Cost?.....	229
5.5 Enculturation via Cognitive Practices: Menary’s Erasure of Epistemic Agency	235
5.6 The Encultured Mind: An Enactive Account of Complementarity	241
Conclusion: Epistemic Enquiry, Encultured Enquirers, and their Future.....	249
6.1 From the Problem of Justification to Enculturation: How Much Has Changed?	249
6.2 The Information Age and Not-Just-Information Problems.....	259
References.....	265

Abstract

The last quarter-century has witnessed monumental advances in studies of the social dimensions of knowledge. In this dissertation, I chart a number of those advances and some remaining difficulties, drawing upon work in social philosophy, cognitive science, epistemology, and the philosophy of science. It has been common, within each field, to initially approach the social dimensions of knowledge as simply another resource to be utilized or controlled: in the philosophy of science, for instance, Kuhn (1977) focuses on how a diversity of individuals foster scientific progress by balancing the tension between innovative and conservative research impulses; in epistemology, Coady (1992) focuses on how other people, via their testimony, are a ubiquitous and indispensable source of beliefs; in social epistemology, Fuller ([1988] 2002) focuses on how to organize scientific research in order to most efficiently realize a set of political values; and in cognitive science, there has been a widespread focus on how individuals' predict or explain others' behavior (Churchland [1984] 2013). From this utility perspective, other people—insofar as they are epistemically significant—appear as resources or tools to be managed and exploited. I call this view—in which people's epistemic significance are only of instrumental value—manipulationism. The following dissertation is dedicated to identifying manipulationism, tracing its manifestations in the aforementioned fields, and explaining how to move beyond it.

That manipulationism was not cause for immediate alarm is attributable in large part to

two factors: adaptationism and the distinction between knowledge and its practical application. Regarding the former, most theorists who even deign to consider the evolutionary origins of knowledge are stuck in the modern Darwinian synthesis, treating random variation and natural selection as the *sole* consequential evolutionary processes. It is only but one more step from this adaptationism to an egoism in which individuals' epistemic aims are grounded in the drive to survive. Regarding the second factor, distinguishing between knowledge and its application suggests that the epistemic dimension of social interaction is isolable. From this perspective, people *might be* more than just valuable tools, but not when considered strictly in terms of their epistemic value. A large portion of the following dissertation concerns how the knowledge-and-its-application distinction has been undermined, and the issues with manipulationism that arise therefrom. The distinction's shortcomings have been brought to the fore by the palpable failures of attempts to treat other people as just instrumental epistemic resources. To assess these failures and to chart a path beyond them, I consider three broad approaches to analyzing the epistemic significance of social interaction: the testimony view, the translation view, and the transformation view.

The testimony view is an outgrowth of analytic epistemology and the attendant justified true belief account of knowledge. On the testimony view, other people are essentially efficient belief transmitters. All of the transmitted beliefs could, in principle, be acquired by one's own senses, but owing to one's physical finitude, it is necessary to rely on others' firsthand experiences. Ideally, as in the case of courtroom testimony, the other person reports their sensory experience without covertly introducing any information absent from their original experience. As a result, the central epistemic question, on this view, concerns the problem of justification: can beliefs gained through testimony, under any circumstances, be considered knowledge despite

the recipient apparently lacking proper justification. As an alternative to the testimony view, I draw upon the translation and transformation views for understanding the epistemic significance of social interaction, both of these views shifting the epistemic question away from the traditional one of justification and towards the nature of embodied skills.

The translation view is tied to the notion of tacit knowledge, as first developed in the philosophy of science. Becoming a scientist requires the direct tutelage of experienced professionals, suggesting that being a scientist involves inarticulable skills (i.e., tacit knowledge). Accordingly, the label “*translation view*” refers to the inherent need for agents to translate the behavior and words of others into their own unique competencies. Social interaction is thus important precisely because of the ineliminable differences between individuals, not because of what they share. The definitive tenet of the translation view is that diversity amongst individuals is pervasive, persistent, and profoundly shapes social interaction. While the emphasis on the profound differences between individuals is salutary and serves as a crucial standard for social theory, the translation view fails to overcome manipulationism. Other people, on this view, are epistemically significant only insofar as they are a means for passing along tacit skills. To overcome the manipulationist presupposition, I assemble a transformation view of cognition over the course of Chs. 3 – 5.

One of the principal aims of the transformation view is to bridge the gap between cognitive science and social epistemology. There has been a troubling lack of exchange between, on the one hand, cognitive scientists who examine the thinking subject as an isolated individual and, on the other, social epistemologists who limit themselves to population-level descriptions and prescriptions. The key tenet of the transformation view, as the name suggests, is that cognition—especially cognition within a social context—is a transformative process. To put the

same point polemically, the transformation view objects to the existence of an innate and *immutable* cognitive core. I begin assembling the transformation view in Ch. 3, using the enactivist movement within cognitive science, then shift to the population-level perspective of neo-Kuhnian social epistemology in Ch. 4, and, finally, return to cognitive science in Ch. 5 in order to assess the Extended Mind debate. Incorporating both first-personal and population-level perspectives is essential to modeling enculturation without reifying one aspect of the dynamic relationship between self and others.

The main impetus behind enactivism is to provide a naturalized account of cognition that is non-exceptionalist and non-saltationist. Aligned with the so-called Extended Synthesis in contemporary biology, enactivism challenges egoistic pictures of agency, in which the agent chooses to interact with its environment in order to satisfy self-interested desires. The enactivist subject, by contrast, is a self-maintenant network of processes that lacks a fixed boundary between self and environment. By presenting an alternative to egoistic models of agency, enactivism points the way to overcoming manipulationism. Yet enactivism's implications for social theory have only recently begun to be articulated with the concept of participatory sense-making. Applied to two-person interactions, participatory sense-making shows how epistemically productive social coordination does not presuppose individual cognitive mechanisms. But extending participatory sense-making to larger-scale and longer-term social contexts has proven difficult: most efforts have reified coordination patterns and consequently obscured the role of idiosyncratic individuals. To provide a more compelling extension of participatory sense-making, I draw upon accounts of epistemic communities found in neo-Kuhnian social epistemology.

Social epistemology's rich accounts of epistemic communities identify a number of

coordination mechanisms. One mechanism that I discuss in Ch. 4 is *cogent* argumentation, a coordination mechanism that exemplifies the further insights made possible by linking social epistemology to embodied and enactive accounts of cognition. According to the concept of cogency, argumentation is a context-specific process that involves an indefinite range of possible factors. When viewed in terms of embodied cognition, the immanent nature of argumentation reflects the fundamental nature of knowledge rather than being merely a practical exigency. My synthesis of social epistemology and enactivism, however, runs contrary to most theorists who have tackled the problem of how to relate the predominantly individualist methodology of cognitive science to population-level accounts of epistemic communities. I review a sample of such competing theories in Ch. 5, two of which are also enactivist.

One indication of the general progress made in the last quarter-century in cognitive science has been a noticeable shift in polemical targets. The trenchant attacks leveled against cognitivism by the so-called “four E” cognitive theories (i.e., embodied, embedded, extended, and enactive approaches) have ceded way in large part to 4E theories critiquing each other. In Ch. 5, I argue that many 4E theories have failed to overcome manipulationism. Using the extended approach as an exemplar, I trace manipulationism to unsatisfactory accounts of first-personal agency. On the extended account, the cognitive subject quite literally disappears, lost in a jumble of multiply realizable processes. In this context, the manipulationist presupposition is necessary in order to glue the processes into something resembling an individual agent—that is, the drive to survive is the multiply realizable end upon which all epistemic goals derive. The disappearing cognitive subject issue also manifests itself in retrospective enactivist accounts.

Some varieties of enactivism model the epistemic agent in strictly retrospective terms. This strictly retrospective orientation is a product of modeling enculturation as a process wherein

an individual agent internalizes objective communal norms. Although the model falls under the transformation view—since it treats socialization as a transformative process—it does so at the cost of erasing any meaningful notion of first-personal agency. Positing a shared set of objective communal norms obscures the idiosyncratic differences between agents, differences that—as suggested by the translation view—play a productive epistemic role.

In contrast to retrospective interpretations of enactivism, I present a future-oriented epistemic agent. With the shift in temporal orientation, the agent's history—including its evolutionary, ontogenetic, and social aspects—no longer appears as a determinate causal chain but rather as an open-ended constraint. Participating in joint epistemic ventures, the encultured agent enacts her history in the continuous process of creating a new one. In what follows, neither agent nor community possess analytic priority; agent and community are rather, I suggest, in a mutually transformative relationship.

Chapter 1: The Testimony View

1.0 Introduction

Imagine a world in which most of the knowledge you possess originates from the verbal reports of other people. These reports are usually based on the first-hand observations of the reporter, him- or herself, and the reporter's primary intention is to inform you of his or her observations. Now imagine another world descended from the first. In this second world, most of your knowledge still depends on others' reports, only now others' reports also come in the form of encoded symbols inscribed on relatively stable physical media. As in the first world, the primary intention of reporters is an informative one. The use of physical media in conjunction with verbal reports facilitates longer chains of reports: people often report to you another person's observations, the third person having inscribed his or her first-hand observation on a physical medium.

Finally, imagine a third world that is like the second, except that encoded inscriptions on physical media are vastly more common. Other people still verbally report their own first-hand observations as well as the first-hand observations of a third person. But it is routine, perhaps even more common, for acquaintances to report the observations of people whom they have never met, who often are themselves reporting another person's observations, some of whom lived hundreds or even thousands of years ago. Furthermore, a large portion—perhaps the majority, or even vast majority—of one's knowledge originates from physical inscriptions rather than verbal reports. According to the epistemological concept of testimony, this imagined third

world represents the epistemically significant aspects of social interaction in the present-day world, while the first two worlds outline its epistemically salient precursors. Given this general picture of epistemic interaction, testimony theorists see their chief epistemological task as explaining in what sense and to what extent (if any) are our testimonial beliefs justified.

In the present chapter, I argue that the concept of testimony and the social epistemology founded upon it are fundamentally asocial. The concept of testimony leads to a clinical, perhaps even pathological approach to the epistemic significance of social interaction. Central to my claim is what I describe as testimony's "transmission view," which underlies the various approaches to testimony that I examine. The transmission view treats social interaction as essentially nothing more than the transfer of discrete units of information. In characterizing the transmission view, I identify two troubling effects it has on social epistemology. First, social interaction—and, by extension, other people as well—are understood in strictly instrumental or manipulationist terms. Second, epistemic agency appears to be inherently passive. In the following, I focus on identifying and tracing the transmission view within the testimony literature, while in subsequent chapters I will offer alternative models of social interaction.

The chapter contains two major parts. In the first half (§1.1-1.3), I examine the testimony literature in general. This general account begins with C.A.J. Coady's seminal 1992 work *Testimony: A Philosophic Study* and traces some of the major outlines of the literature that his work has inspired. I give extensive attention to Jonathan Adler's "core case" approach to testimony because I think it epitomizes the asocial tendencies endemic to the literature. To show the pervasiveness of this asocial tendency I examine three different critical strands that attempt to address the issue, though all ultimately fail to do so. In the second half of the chapter (§1.4-1.6), I turn to Alvin Goldman's epistemics project. On the basis of reliabilism, Goldman avoids many

of the issues that dominate and unduly limit the larger testimony literature. The epistemics project focuses on ways of more efficiently acquiring true beliefs, shifting away from the issue of justification. This shift is helpful in drawing more attention to the transmission dynamics of testimony.

In the final section, I discuss Martin Kusch's communitarian epistemology. Kusch offers one of the most fundamental critiques of the testimony literature, arguing that it has been dominated by the "individualist testimony view." My critique of testimony's transmission view shares much in common with and is partially indebted to Kusch's diagnosis. But unlike Kusch, I find the concept of testimony fundamentally flawed. Kusch's attempt to redirect the testimony literature is understandable given its role in establishing social epistemology as a vital field of study. In subsequent chapters, however, I argue that social epistemology needs to abandon the concept of testimony.

1.1 Epistemology's Litmus Test

The testimony literature, more than any other discussion, is responsible for showing that the epistemic significance of social interaction is more than an afterthought. Frederick Schmitt, in the introduction to *Socializing Epistemology*, considers testimony the "most fundamental test of epistemological individualism" (Schmitt 1994). Even though Schmitt does not think there has been a decisive argument against "weak individualism," social epistemology appears as a vast, ripe, and largely unexplored domain. Grappling with testimony, he thinks, is one of the keys to unlocking the domain. Yet the testimony literature has been subject to a notably narrow dialectic since its contemporary inception with C.A.J. Coady's *Testimony: A Philosophic Study*.

The fundamental nature of testimony has been largely taken for granted. Coady (1992), for example, focuses not on characterizing the nature of testimony but rather arguing for how to

conceptualize testimony's justificatory grounds. Adopting Coady's focus and terminology, the resulting testimony literature has been dominated by the reductionist – anti-reductionist/fundamentalist debate. Reductionists hold that testimonial beliefs ought to be justified by appeal to non-testimonial sources, while anti-reductionists argue that testimony is itself a fundamental source that provides at least *prima facie* warrant.

In the spirit of the reductionist – anti-reductionist debate, Alvin Goldman (2010) casually remarks that the “problem of testimony is a problem about justification.” And Jonathan Adler's (2010) entry in the *Stanford Encyclopedia of Philosophy* on the topic of testimony is not about testimony *per se* but instead the “The Epistemological Problems of Testimony,” with the problem of justification chief among them. Testimony itself (i.e., the act of testifying) appears to be as theoretically uninteresting as social epistemology previously had been. This rhetorical trajectory, however, is not simply a function of Coady's work but, more importantly, reflects the underlying notion of testimony at work.

Testimony is understood to be ubiquitous. Its ubiquity marks testimony's general significance in that it shows our unavoidable dependence on it. In arguing against the reductionist position, for example, Coady (1992) is at pains to establish our dependence on testimony by describing its variety and pervasiveness: from the mundane, such as knowing our date of birth, to more far reaching and complicated matters, such as knowing about historical events and serving as a prerequisite for scientific collaboration. While testimony's ubiquity establishes our dependence, it also implies that the essential act of testifying is relatively trivial. The common element to all of the examples is the generic notion of passing along information, what Elizabeth Fricker (2006) terms as a “telling” in the broadest sense. The only theoretically interesting aspect of the act of testifying itself consists in clearly distinguishing such “tellings”

from other non-epistemic communications. Axel Gelfert (2006) offers a very different motivation for the epistemic significance of testimony but is committed to same underlying notion of testifying as a mundane act of telling.

Gelfert uses the unlikely figure of Immanuel Kant to argue for the epistemic significance of testimony. Kant's argument differs markedly from the contemporary literature because in it our dependence on testimony follows from ethical considerations rather than empirical ones. Gelfert interprets Kant's lectures on logic as presenting our trust in others' testimony—so long as the testimony concerns empirical as opposed to moral matters—as a presumptive imperfect duty. The duty follows from the thought that if we want to be believed by others, then we have a duty in turn to believe others (Gelfert 2006). Despite its distinctive ethical dimensions, Kant's argument, like the contemporary literature on testimony, is primarily concerned with establishing our reliance on testimony rather than examining the act of testifying itself.

Having established people's dependence on testimony, the subsequent question is what implications this has for our understanding of knowledge. Both reductionists and anti-reductionists agree that people rely on testimony, but disagree over what epistemic status to accord to testimonial beliefs. More specifically, if knowledge involves some element of justification, then how are testimonial beliefs justified. Reductionists attempt to *reduce* testimonial beliefs to individual cognitive resources, as opposed to fundamentalists like Coady who think that such a task is hopeless and thus infer that there must be an additional *social* source of justification. It is worth returning to Schmitt's remark, noted above, that testimony is the fundamental test of epistemological individualism. The test is specifically over the question of whether knowledge can be justified solely on the basis of individual cognitive capacities.

Adler labels testimony's justificatory question as the "Vulnerability Problem." He poses the problem as follows: "Given that speakers of a language sometimes assert falsehoods and fail to be sincere, under what conditions, if any, is someone's word alone sufficient to justify the beliefs a hearer acquires from those assertions?" (Adler 2010). Dishonesty alone makes people epistemically vulnerable if they do not and, in many if not most everyday situations, cannot verify a speaker's claims. Additional factors, such as the possibility of honest errors, only further highlight the vulnerability of endorsing testimonial beliefs. The Vulnerability Problem—the very idea that individuals are epistemically vulnerable due to their reliance on other people—is what makes testimony the fundamental test of epistemological individualism. For the Vulnerability Problem to be a genuine test of epistemological individualism, however, the Problem must be more than a marginal worry.

In order to show that the Vulnerability Problem is a widespread concern, Adler (2010) appeals to three general features of our "conversational practice." The first corresponds to the main fundamentalist point from above, namely that we are profoundly dependent on testimony. This is the point of showing that reliance on testimony is necessary in mundane matters as well as in rigorous epistemic endeavors, such as scientific research. Adler's second feature is people's tendency to accept "ordinary informative testimony." The idea here is that at least with respect to everyday topics—such as the time of day, the weather forecast, the name of a nation's president, and so on—people usually accept answers to such questions without a second thought. The final feature of our conversational practice is closely related to the first but deserves more extensive treatment.

The third feature of our conversational practice concerns people's ability to test their trust in a speaker. Within the constraints of normal testimonial exchange, it is infeasible for hearers to

“seriously” check or confirm a speaker’s reliability or sincerity. The reliability issue results in part from our far-reaching dependence on testimony, as established in the first feature. Coady’s critique of Hume as the representative of reductionism carefully unpacks this point: the scope of testimonial beliefs dwarfs any particular individual’s first-hand experiences. Thus, the individual’s experiential base is far too narrow to support an inferential induction to the reliability of speakers in general (Coady 1992). And, if it is impossible to infer the reliability of speakers in general, practical constraints will prevent people from inferring the reliability of all or even many particular speakers. Thus, these three features of ordinary conversational practice show that the Vulnerability Problem pervades much of what we consider ourselves as knowing.

It’s worth questioning, however, whether it is misleading to treat all of the various instances of testimony in terms of a single general “conversational practice.” The danger is to abstract away from epistemically significant aspects of social interaction. This danger is even more evident when Adler attempts to more clearly delineate the Vulnerability Problem by identifying a “class of core cases.”

Adler’s class of core cases is defined by six conditions (2010). First, a piece of testimony can only be a single sentence, eliminating the possibility of the statement being supported by cohering with other assertions. Second, there can only be one speaker, eliminating the possibility of support via corroboration with other speakers. Third, the social context must be one in which the norm of truthfulness holds and the purpose is to inform, ensuring that there is a strictly epistemic dimension to the exchange. Fourth, the testimony must sustain the corresponding belief in the hearer, ensuring that the testimony is doing the work and not some separate source. Fifth, the speaker should not be recognized as an expert, since expertise could provide an additional source of support. And, finally, the hearer must have no special knowledge of the

speaker, ensuring that such personal knowledge does not provide independent support for the testimony. The six conditions aim to distill the essential element of testimony as one person passing along a belief to another with the receiver accepting the belief as such. But the usefulness of this theoretical idealization, even more than the “conversational practice” idealization, is unclear.

The core of testimonial acts, according to Adler’s idealization, is *one* statement passing in *one* direction. The worry, at this point, is that the class of core cases seems to have lost touch with the reality of testimony. Given the latter condition, the core of testimony is not even an exchange but rather a one-directional transmission. It is worth doubting whether all six conditions could *ever* be simultaneously fulfilled. One of the basic issues is that people do not stay strangers for long, which directly problematizes the sixth condition of the core case idealization and, more generally, the aim of isolating the transmission and reception of a discrete belief from any other epistemic considerations.

By making the class of core cases the definitive core of testimony, our understanding of what is meant to be a pervasive part of our everyday lives is based on a questionable idealization. I think Adler’s approach is symptomatic of what Hanne De Jaegher and Ezequiel Di Paolo (2007), when discussing classical approaches in the philosophy of mind, pejoratively refer to as the “Rear Window Approach.”¹ I will return to this thought in Ch. 3 when discussing the concept of participatory sense-making, but, for the moment, I only want to gain more perspective on the underlying notion of testimony and the literature that it has inspired.

¹ It should be noted that the “Rear Window” criticism does not imply that testimony cannot be put to valuable uses. Perhaps its most valuable application has been in Virtue Epistemology, by thinkers such as Miranda Fricker, John Greco, and Ernest Sosa.

Testimony, as a topic of philosophic inquiry, has offered an uneven perspective on the social dimensions of epistemology. On the one hand, the contemporary literature initiated by Coady (1992) has thoroughly established our epistemic reliance on other people, a reliance that had been obscured by traditional individualist epistemology. In this respect, the topic of testimony has been a major corrective to traditional epistemology.

But, on the other hand, particular acts of testifying appear to be relatively trivial and, for the most part, theoretically uninteresting. This tendency is common to diverse approaches to testimony, from Coady's original treatment to reductionist responses as well as other approaches such as Gelfert (2006). The major theoretical question has been what implications testimony has on our concept of justification, as evidenced by the dominance of the reductionist – anti-reductionist debate. Justification seems to both sides to be the sticking point: the ubiquity of testimonial beliefs makes it impractical if not entirely infeasible to verify them. Given this situation, it is unclear in what sense testimonial beliefs could be adequately justified in order to be considered knowledge. And dismissing testimonial beliefs as less than knowledge appears to discount far too much of what people *know*; in contrast to Quine's ontological preferences, not everyone likes desert landscapes. One potential casualty of this debate, I have suggested, has been an adequate treatment of testimony as it occurs in our everyday lives.

Given the size of the testimony literature, it should be no surprise that these issues have been at least alluded to. Not everyone, for example, has been satisfied with the reductionist – anti-reductionist debate. And the basis for identifying a class of core cases—which presupposes that there is some homogenous essence common to all instances of testimony—has also been called into question. These critical strands shed some light on the basic concept of testimony *per se*, which as previously mentioned tends to be taken for granted in the general literature. In the

next section, I examine three critical strands within the testimony literature and claim that they are insufficient for adequately addressing the present concerns.

1.2 The Nature and Domain(s)² of Testimony

1.2.1 Elizabeth Fricker's Partitioning of Testimony's Domain

Adler's explication of the Vulnerability Problem implies that testimony consists of one unified domain. For if testimony does not consist of one unified or homogenous domain, then the use of core cases to explain and address the Vulnerability Problem would be inappropriate. Testimony's domain has been an ongoing (though mostly marginal) concern of the contemporary testimony literature, beginning with Elizabeth Fricker's (1995) extensive critical review of Coady (1992).

Fricker's most general criticism of Coady is for treating testimony as "one unitary category" (1995). In place of one category, Fricker makes two crosscutting distinctions. The domain of testimony is partitioned based on, first, "kinds of tellings" and, second, the "phases in the career of a recipient of testimony." Different "kinds" of tellings refers to differences in subject matter, while the "phases" of a receiver's career simply refers to the recipient's maturity level. Both of these are very coarse-grained concepts, as Fricker is only interested in two sub-categories within each. With respect to maturity level, there are two stages—a developmental phase defined by simple trust in other people and a mature phase. Likewise with kinds of tellings, Fricker is not interested in particular subject matters but rather whether a given subject matter lends itself to competency. As a result, Fricker is only interested in two types of subject matter—

² In keeping with the testimony literature, I use "domain" in the non-technical sense of scope or range. There are times at which a technical sense of domain seems to be implied in the literature (e.g., claims regarding a finite number of types of testimony), but there has not been a rigorous treatment of such implications.

those that should and should not be expected to support competent testimony. Coady's unified domain is thus crisscrossed by four sub-categories.

By partitioning testimony's domain, Fricker attempts to reconcile the fact that, on the one hand, some testimonial beliefs seem to be appropriately accepted on trust alone with the belief, on the other hand, that some testimonial beliefs ought to have justificatory support in order to be deemed knowledge. The second distinction regarding maturity level has the most obvious implications for this question. In the developmental phase, cognitive subjects abide by the Presumptive Right Principle (PR) in which they are entitled to believe what they are told *as such*. Among the purported benefits of the PR principle within the developmental stage is enabling the acquisition of language. Thus, the developmental stage's PR principle is not merely a temporary dispensation but also serves some positive epistemic ends. The first distinction from above helps to explain how the PR principle may appear as though it still holds in the mature phase when it actually does not.

Even for mature cognitive subjects, there are some subject matters that do not deserve close scrutiny. Topics that readily support a speaker's competency—such as the speaker's name, where she lives, anything that is readily accessible to her perceptual capacities, and so on—ought to be trusted more than complex or contentious matters, such as political issues. In cases of simple subject matters, mature cognitive subjects often *rightfully* accept the testimony without any attempt to independently verify it. Even though the exchange thus looks similar to the PR principle, Fricker claims that mature subjects monitor the speaker for problem signs. This is already implied by the need to discriminate between which subject matters support competent testimony, but even with regard to these subject matters it is also necessary to monitor the speaker's sincerity. Monitoring is meant to be a background mental activity that only switches

into conscious scrutiny when a problematic sign appears. Hence, a mature subject's quick acceptance of a testimonial belief may appear compliant with the PR principle when, in fact, it is more epistemically robust having passed unconscious monitoring for sincerity and competence.

Fricker's interest in partitioning the domain of testimony relates back to the reductionist – anti-reductionist debate initiated by Coady. One result of the partition is distinguishing between two possible versions of reductionism—local and global. Coady (1992) seems to only recognize the possibility of global reductionism, which entails establishing the reliability of testimony in general. Fricker, by contrast, endorses local reductionism: monitoring testimony for signs of dishonesty and incompetence only justifies particular testimonial acts. Fricker's primary interest in addressing the justificatory question explains the limitations of her partition. The four-category partition of testimony's domain, while an improvement on Coady's monolithic understanding, is too coarse-grained to address the concerns raised in the previous section, namely whether the testimony literature abstracts out epistemically significant aspects of social interaction. This is evident when comparing Fricker's partition with Adler's core case approach.

Despite their differences, Fricker's account is compatible with the spirit of Adler's core case approach to the Vulnerability Problem. Reconciling the two only requires indexing Adler's core cases to one of Fricker's four partitions, namely the domain of mature cognitive subjects who are dealing with a subject matter that lends itself to competency. Fricker, after all, is most interested in this domain anyway, as it is within this domain that background monitoring of testimony occurs. Adler's fifth condition on core cases—that the speaker not be seen as an expert—even hints at Fricker's subject matter requirement. Hence, although differing on the details of how to conceptualize testimony's domain, both Adler and Fricker are invested in

identifying the general factors that justify particular testimonial beliefs. And, for both Adler and Fricker, this justification depends on largely unconscious background activities and beliefs.

1.2.2 Jennifer Lackey on the Nature of Testimony

Jennifer Lackey poses a different challenge for Adler's core case approach by more closely examining the nature of testimony itself. In a pair of articles, Lackey argues for a dualism within the epistemology of testimony as opposed to treating it as one homogeneous whole. Dualism in this context refers to synthesizing elements from both reductionism and anti-reductionism into a hybrid theory. Constructing the hybrid theory of testimony is what leads Lackey to reassess the nature of testimony itself.

Examining the fundamental nature of testimony *per se* is unusual in the testimony literature. Most discussion of its nature is limited to distinguishing testimony from other non-epistemic aspects of social interaction. Lackey (2006b) is led to more closely examining the topic because she thinks that the reductionist – anti-reductionist debate collapses two distinct aspects of testimony. These two aspects—hearer testimony and speaker testimony—are hinted at and alluded to in others' attempts to define nature of testimony.

Lackey considers three definitions of testimony in support of her argument for dualism. First, Coady's (1992) definition of testimony is extremely restrictive: the speaker must intend to communicate a piece of information and the recipient must be interested in the information. The definition discounts a great deal of what seem to be instances of testimony. Lackey refers to this as the Narrow View of Testimony, which Peter Graham improves upon by dropping Coady's listener requirement while keeping the speaker's intention requirement. For if someone intends to inform another person, then this is an act of testifying regardless of whether the audience cares

to be informed. Lackey labels Graham's modification the Moderate View of Testimony.³ Finally, Elizabeth Fricker dispenses with speaker intentions entirely, arguing that any expression of thought constitutes an instance of testimony. Fricker's definition—which Lackey labels the Broad View of Testimony—has the benefit of accounting for instances in which a listener may make use of a speaker's thoughts despite the speaker lacking any intention to inform.

While highlighting Coady and Graham's overly narrow construal of testimony, Fricker's Broad View is subject to its own problem. Fricker's Broad View cannot properly exclude instances in which no one—neither the speaker nor listener—takes an expressed thought to be epistemically significant. Relevant examples include conversational fillers and words of encouragement: if neither the speaker nor the recipient finds an expression epistemically significant, then Lackey believes that it does not count it as testimony.

All three definitions contribute to Lackey's concept of dualism. Coady's definition even includes both speaker and hearer testimony but treats each as a necessary rather than sufficient condition for testimony, while Graham and Fricker each focus on only one of the halves. Lackey's dualism forms the basis for her intervention in the reductionist – anti-reductionist debate, as explicated in Lackey (2006a).

As evidenced from the title "It Takes Two to Tango," Lackey (2006a) thinks that the testimony literature has failed to appreciate the epistemic importance of there being multiple parties to a social interaction. For Lackey there are two distinct and intrinsically important roles in a testimonial exchange—the speaker and hearer—regardless of how many individuals are actually involved. The reductionist and anti-reductionist camps gravitate towards opposing aspects of testimony. Reductionism requires that the hearer find positive reasons for a testimonial

³ Although Graham's account is "moderate" in the sense of how many acts are counted as testimonial, his larger Information-theoretical framework has dramatic implications. As I discuss later, this framework is perhaps the starkest exemplar of testimony's "Transmission View."

belief, the positive reasons reducing the beliefs to non-testimonial epistemic grounds. By exclusively focusing on the hearer's need to provide positive reasons, the reductionist neglects speaker testimony. Conversely, anti-reductionism emphasizes the role of defeaters or what Lackey terms "negative reasons": a testimonial belief is justified so long as there are no negative reasons, such as indications of dishonesty or incompetence. Negative reasons shift the focus from the hearer to the speaker: the presence or absence of defeaters tells us nothing substantive about the hearer and instead indicates the reliability of the speaker. Lackey's dualism claims that both positive and negative reasons are necessary for justification.

Given the scrutiny that the reductionist project has received, dualism's positive reason requirement is the most contentious issue. Lackey thus brackets the issue of how to identify a reliable speaker in order to defend the need for the hearer to provide some positive justificatory reasons. The argumentative strategy, here, is two-pronged. First, the positive reasons requirement does not entail reducing testimonial beliefs to non-testimonial sources. Second, recognizing defeaters/negative reasons as a separate standard alleviates some of the burden on the positive reasons requirement.

The two-pronged strategy means that positive reasons are necessary though insufficient for justifying a testimonial belief. This position is not open to reductionists since their positive reasons are meant to accomplish a reduction and so would necessarily imply justification. On Lackey's view, the positive reason requirement alone does not imply that a person is rationally justified in his or her belief, only that he or she is "*not irrational*" in holding the belief. The counterexample to anti-reductionism—the alien diary example—is meant to show the need for this sort of positive reason. In the example, a person sees an alien drop something in the woods. Upon retrieving the item, it bears all the markings of a diary written in the English language;

there are no defeaters that suggest that the statements cannot be read as a diary. Yet Lackey thinks that it would be epistemically unjustified to endorse the testimonial beliefs that appear within the apparent diary. It is not rational to accept the beliefs given the potential differences between alien and humans; hence, the hearer in this case lacks positive reasons despite there being no defeaters. For Lackey, the basic issue with the alien diary is that we do not know whether the alien culture includes a testimonial practice analogous to our own.

Before considering the lessons that Lackey draws from the alien diary example, it is necessary to first flag some basic concerns with the example. Most immediate, it beggars the imagination that someone would be able to identify an alien without knowing *anything* else about it. More generally, the example is another highly questionable idealization and, as with Adler's class of core testimony cases, it lacks any clear connection to the lives of actual epistemic agents. As a result, the alien diary scenario does not support any substantial conclusions regarding the nature of testimony. Even the less far-fetched version of this example in which an amnesiac person has lost all her beliefs about people's testimonial practices is similarly flawed. Contrary to Lackey's presentation, Martin Kusch (2012) points out that these sorts of atypical cases do not actually support obvious and stable intuitions regarding justification, or anything else for that matter.⁴ The dubiousness of Lackey's examples calls into question her notion of positive reasons, which in turn questions whether her account ultimately differs from reductionism. This concern becomes even more prominent given Lackey's discussion of particular positive reasons.

⁴ Another deeply troubling feature of the positive reason examples is the implication that all individuals possess determinate beliefs about *a* testimonial practice that all of humanity participates in. I think this implication is an artifact of a retrospective reconstruction on the part of theorists rather than actual beliefs held (whether consciously or unconsciously) by all human epistemic agents. In this respect, Lackey's examples typify the "Rear Window Approach" that De Jaegher and Di Paolo (2007) criticize.

On the basis of the alien diary example, Lackey recognizes three classes of positive reasons. The three classes are not meant as an exhaustive list but rather only to clarify the concept of positive reasons. Context is the first class: without knowing anything about the alien's culture, it is impossible to properly contextualize the putative diary. Context is simply meant to indicate whether a critical epistemic attitude is appropriate. The *National Geographic Magazine*, for example, deserves a less critical attitude than the *National Enquirer*. The second class discriminates between types of reports: some reports (e.g., mundane perceptual observations) are more trustworthy than others (e.g., a reported UFO sighting). The third class is the individuation of epistemically reliable speakers, although individuation does not consist of evaluating speakers in terms of their personal histories. Instead, individuation means typing a particular speaker in terms of epistemically salient factors: for example, an accountant is a reliable source of information about taxes, campaigning politicians are unreliable sources concerning their opponents' qualifications, and so on.

Lackey's examples of positive reasons closely resembles Fricker's reductionist position. Most obvious, the second class closely resembles Fricker's subject matter requirement. More generally, whether it is a matter of context, kinds of reports, or "individuating" reliable speakers, the positive reason requirement serves the same function as Fricker's unconscious monitoring test. Failing Fricker's monitoring test and Lackey's positive reasons standard have the same result, namely conscious deliberation and thereby an attempt at independent verification. Lackey's framework does improve upon Fricker's: there are more fine-grained considerations than Fricker's four categories, and Lackey is not committed to the questionable psychological claim regarding unconscious monitoring. But Lackey offers only a modest improvement—

mostly a conceptual clarification—that fails to address the larger issue of obscuring the social reality of testimony. This is especially notable since Lackey touches directly on the issue.

The conflation of hearer and speaker testimonies is symptomatic of the over-abstraction pervasive within the testimony literature. Lackey has to remind her interlocutors that there are two distinct parties involved in a testimonial exchange (i.e., “it takes two to tango”). Yet even Lackey’s dualism downplays the importance of social engagement. The speaker and hearer both need to make a substantive contribution to the testimonial exchange, but the other person appears as a general type or cluster of types rather than a unique individual. Our epistemic interest in the other person is, strictly speaking, instrumental and clinical (i.e., manipulationist): we identify the epistemically relevant types that apply to the other person in order to judge their trustworthiness and competence. Social engagement is epistemically interesting in only two respects: how it affects the efficiency of transmission and how it affects our ability/opportunities to detect signs of dishonesty or incompetence. Hence, despite raising concerns about collapsing the speaker and hearer aspects of testimony, Lackey’s dualism still marginalizes the social reality of epistemic interaction, especially the importance of engagement.

1.2.3 John Greco’s “Friendly” Epistemic Environments

John Greco (2012) poses a direct challenge for Adler’s core case approach. He argues that not all cases of testimonial knowledge are epistemically homogenous. Unlike Fricker and Lackey, Greco is not content with status-independent criteria for identifying trustworthiness and competence. Thus, Greco’s proposal is more than a partitioning of testimony’s domain, as with Fricker, and also more than a generic division of labor, as with Lackey’s distinction between speaker and hearer testimonies. The initial impulse for Greco’s proposal stems from trying to account for testimonial exchanges involving children.

On the one hand, it seems obvious that children gain knowledge from testimonial exchanges with their parents but, on the other hand, children appear to fail any listener requirements. For example, a parent may have passed along the belief that the earth is a sphere, which the child now *knows*; but the parent could have just as easily told the child that the earth is flat, which the child would then believe despite its being false. Given the apparent gullibility of children, it is unclear how the former case can still be counted as a case of testimonial knowledge. This issue, which was initially broached by Sanford Goldberg, arises in a number of other situations.

Greco provides two additional situations that also involve “epistemic caretakers.” Students similarly accept their teacher’s testimony and patients their doctor’s testimony. In all three cases, the recipients—children/students/patients—appear to fail any possible hearer reliability requirement since they would just as easily believe the opposite of what the parent/teacher/doctor testified to. Despite seemingly blind acceptance, Greco thinks that these are still cases of testimonial knowledge. The acceptance is justified because as epistemic caretakers, the testifiers are sources of testimonial knowledge in a way that people more generally are not.

Epistemic caretakers inhabit special social roles. Greco does not detail how exactly these roles originate and are recognized or maintained but, instead, appeals to the sense that social contexts or “environments” can have relevant epistemic differences. Epistemic caretakers correspond to “friendly” social environments in that the caretaker is justifiably expected to provide reliable testimony. In these friendly social environments, the hearer reliability requirements are lower; in fact, the only mistake that a hearer can make in such a situation would

seem to be one of inattention or misperception. Barring such errors, the caretaker will transmit testimonial knowledge to his or her recipient.

Greco's social environment proposal calls into question the fundamental nature of testimony. In comparison, Lackey reassesses the fundamental nature of testimony but only finds the need to clarify a generic division of labor between speaker and hearer: speaker and hearer are determined by the activity of the individuals involved, not by a social status. By taking the social environment into consideration, Greco's proposal offers a more complex perspective on testimonial exchanges.

Greco's proposal also reconceptualizes testimony's domain. Unlike Fricker, Greco does not attempt to partition the entirety of testimony's domain but, instead, circumscribes cases that involve the transmission of testimonial knowledge. This departs from the prevailing view in which all cases of testimonial knowledge involve transmission. The issue with the prevailing view is that there are cases in which a recipient may end up more informed than the speaker. To illustrate why knowledge transmission is *only* one type of testimonial exchange, Greco presents an identical twin example.

In the example, a reporter misidentifies an identical twin. The speaker S calls the receiver R and notifies R that Jim has entered the building, but Jim happens to be standing in front of R. Consequently, R knows that Jim's identical twin is in the building despite S's misidentification and possible continued ignorance. In such a case—which even if relatively uncommon is far from extraordinary—it would be inaccurate to say that knowledge was transmitted from the speaker S to the receiver R. Yet the testimony did result in testimonial knowledge since R's knowledge of the whereabouts of Jim's twin depends on S's testimony. On this issue, Greco's critical ambitions are much more modest.

Even though not all cases of testimonial knowledge involve transmission, Greco thinks that the transmission cases are what make testimony epistemically important. He has in mind especially the caretaker examples, which are one type of transmission. Parents, teachers, and the like help to explain the inexhaustibility of testimonial knowledge: within such friendly environments, people can very efficiently gain knowledge because of the relaxation of hearer requirements.

Although factoring in social environments is an important departure from the prevailing testimony literature, Greco's proposal is still socially impoverished. As with Lackey, there is little concern for the particular details of how individuals engage each other. Social environments are only epistemically relevant insofar as they affect justificatory standards. And like Fricker's partitions as well as Lackey's defeaters and positive reasons, all of the various social environments only provide two epistemically salient outcomes: either the social environment is friendly and so lax hearer requirements are justified, or the social environment is not and so lax hearer requirements are unjustified. Thus, while Greco is right that social environments directly factor into judgments of testimonial knowledge, his focus on the question of justification leads to an overly abstract treatment of social environments.⁵

The shortcomings of Greco's proposal are only amplified by the fact that he glosses over the details of the origination, recognition, and maintenance of social roles. It is also unclear whether some of the caretaker examples stand up to closer scrutiny. The doctor-patient relationship, for example, is not *simply* a "friendly" environment, as evidenced by issues such as

⁵ Greco's work in Virtue Epistemology, including its intersection with testimonial knowledge, does not stick as closely to the issue of justification. Furthermore, Greco (2007) argues that it is possible for multiple individuals to deserve credit for cooperative epistemic achievements. While the discussion of credit is an improvement on justification, especially with its closer association to practical activity, the discussion's main aim is to solve the Gettier problem.

prescription kickbacks and insurance company mandates. Given the many well-publicized issues with doctors and their fallibility, it is unjustifiable to accept one doctor's prognosis, at least when it concerns a serious medical condition. Evaluating an actual epistemic interaction—such as that between a particular doctor and patient—thus requires a much more sophisticated account of social environment than offered by Greco.

An alternative lesson to draw from my doctor-patient example would be to reconceptualize the epistemic agent's responsibilities. Alvin Goldman's proposal for reliabilism provides such an alternative, shifting the agent's responsibility from directly verifying a particular belief (e.g., a doctor's prognosis) to evaluating the broader belief-forming processes that are the causal underpinning of the belief (e.g., the doctor being part of an HMO). Goldman (2010) recognizes two basic ways of extending traditional individualist epistemology: either focus on the goal of acquiring true beliefs or focus on the goal of having justified beliefs. In terms of the present example, focusing on the goal of acquiring true as opposed to justified beliefs implies worrying about the general causal processes that underlie one's beliefs (e.g., does the doctor's affiliation make him an unreliable source of information) rather than directly justifying the particular belief in question. While Goldman recognizes the need to have an account of justification, he does not think it should drive our enquiries in social epistemology. Importantly, Goldman's focus on the goal of acquiring true beliefs draws attention to the transmission view that underlies the concept of testimony.

1.3 The Transmission *Thesis* versus The Transmission *View*

Before turning to Alvin Goldman, it is worth examining the scant critical attention that has been given to the topic of transmission within the testimony literature. Much of the critical discussion is due to Jennifer Lackey's efforts, who in a series of articles (1999, 2003, 2006c, 2008)

questions what she refers to as the Transmission of Epistemic Properties (TEP) thesis.⁶ Lackey's approach here closely resembles her diagnosis and proposed solution concerning the nature of testimony noted above. I think that in both cases Lackey alludes to a fundamental issue with the testimony literature but only offers a modest conceptual revision to address it. Therefore, in the next section I contrast Lackey's critique of the "transmission thesis" (i.e., TEP) with the broader "transmission view" that Lackey and her interlocutors are committed to.

In the introduction to *The Epistemology of Testimony*, Lackey summarizes her work on the transmission thesis (TEP). The main reason why the topic of transmission receives little critical attention within the testimony literature is that both non-reductionists and reductionists alike are committed to TEP. The guiding idea behind TEP is that a speaker transfers a belief and the belief's epistemic properties to a recipient in a testimonial exchange. Lackey breaks TEP down into two aspects—a necessity and a sufficiency thesis—using the metaphor of a bucket brigade to explain each.

Lackey's critique of TEP's necessity thesis (TEP-N) leads her to conclude that testimony can *generate* knowledge as opposed to merely *transmitting* it. Because this is Lackey's most dramatic conclusion, I am going to restrict my attention to TEP-N. The necessity thesis entails that a speaker S *must* have a justified belief in her testimony in order for the recipient R to have a justified *testimonial* belief. The bucket brigade analogy is meant to express this intuition as follows: a person S cannot pass along a full bucket of water to another person R, unless S had a full bucket to begin with. In the analogy, the full bucket symbolizes justified belief/knowledge,

⁶ Peter Graham (2000a, 2000b) also criticizes the TEP thesis, although using a different label. Graham's critique is offered from an Information-theoretic point of view, claiming that information rather than knowledge is transmitted via testimony. The Information-theoretic view exemplifies the Transmission View that I discuss in the next section and, in that respect, the fundamentally asocial character of testimony. I later discuss Graham's view, but I use Lackey's critique here because it complements her dualist account of testimony and thus is part of one of the most important attempts to capture the *social* element of testimony.

the passing along of water symbolizes the act of testifying, and R's resulting bucket of water represents the status of the recipient's testimonial belief (i.e., if full, then R possesses justified belief/knowledge). According to the metaphor, it is impossible for R to end up with a full bucket of water if its contents depend solely on S's contribution and S did not have a full bucket to begin with.⁷

A counterexample to TEP-N consists of a testimonial exchange in which the testifier lacks knowledge but the recipient gains knowledge based on the testimony.⁸ There are two general ways for this to occur: the speaker S may fail either the condition of belief or justification while the recipient R satisfies both (Lackey 1999). In either scenario, S would not possess knowledge regarding her testimonial belief *p* even though it is possible for R to satisfy both conditions and thus to know *p*. I will focus on the latter possibility—in which the speaker fails to satisfy the justification condition—since Greco's twin example exhibits such a scenario.

As noted above, in Greco's identical twin example, a receiver R gains knowledge despite the speaker S's false testimony. It is not necessary for S to have knowledge in order for his or her testimony to support testimonial knowledge in R. All that is needed is for S's testimony to have a non-accidental/truth-conducive relationship to a corresponding fact: in Greco's example, S's testimony need only be the product of a reliable perceptual capacity that recognizes a person who appears identical to Jim; given this condition, R knows it is Jim's identical twin. There are many other possible examples of testimonial exchanges in which the recipient gains knowledge despite the reporter's lack of it. One virtue of the identical twin example is that it is clear that knowledge

⁷ Since Lackey is attacking TEP-N and the Bucket Brigade analogy is meant to represent TEP-N, it should be no surprise that the analogy is deeply flawed. Issues with analogy include: assuming that (1) everyone has the same sized bucket, (2) a person cannot receive buckets from more than one source, and (3) a person can only pass along a full bucket. Lackey's own alternative modestly challenges (3), though does not clearly address the first and second assumptions.

⁸ Greco's identical twin example, which closely resembles the following, is partly indebted to Lackey.

generation cases are far from rare. Thus, it appears as though Lackey has identified a major lacuna within the testimony literature. But Greco's understanding of these cases suggests that the implications of Lackey's critique are much more modest.

Lackey's generation examples are not what make testimony the so-called *litmus test* for epistemology. In the given generation cases, the hearer already possesses knowledge directly relevant to the testimonial belief, as opposed to Coady's ubiquity examples in which the hearer is profoundly dependent on the speaker's testimony (e.g., knowledge of our date of birth). Hence, even though the transmission cases are not the only type of testimonial exchange, Greco (2012) concludes that they represent perhaps the most important class of cases for testimony. Lackey's critique thus only offers a conceptual clarification of testimonial exchanges, showing that there are at least two basic types of exchange. But the critique of the transmission thesis raises more fundamental questions regarding the premises of the testimony literature—which Lackey does not discuss—bringing us to the transmission view.

1.4 Alvin Goldman's Reliabilism

Lackey and her interlocutors are committed to the transmission view of social interaction. The transmission view consists of two basic claims:

- (1) the transfer of discrete units of information exhausts the epistemically significant aspects of social interaction,
- (2) the transfer of discrete units of information is isolable from all other aspects of social interaction.

In terms of the bucket brigade metaphor, the transmission view corresponds to the notion that epistemic interaction is founded on the use of a common currency, which in the metaphor is water. The epistemic equivalent of the metaphorical water is propositional beliefs or just bare information. The nature of the currency—whether beliefs or bare information—depends on the

particular account of testimony, but essential to the all of the accounts is that epistemic agents share and circulate the same currency.

In Lackey's generation cases, assertions of belief are the fundamental currency. The transmission of an assertion is the precondition for the recipient to satisfy additional conditions (e.g., justifying the belief via a defeater-defeater) that then "generates" knowledge. Lackey thinks that all cases of testimony, regardless of how the requirements for knowledge are satisfied, are founded on the transmission of an assertion. Thus, Lackey's distinction between the transmission and generation cases is more concerned with how knowledge is justified than with what exactly is exchanged and, in this respect, still reflects the testimony literature's myopic focus on justification. As noted earlier, Peter Graham's Information-theoretic framework exemplifies the transmission view underlying the testimony literature.

Graham (2000a, 2000b, 2000c) applies Fred Dretske's Information-theoretic account of perceptual knowledge to the concept of testimony. Graham's central move is to challenge the same two knowledge transmission theses as Lackey, only he models testimony in terms of "conveying information." Conveying information results from "a law-like correlation or counterfactual dependence between a signal ... and another event, condition, or state of affairs" (Graham 2000b). In the context of testimony, the informational "signal" corresponds to the reporter while the resulting "condition" is the recipient's cognitive state of knowledge. Graham thus sees the foundation of testimony—at least in cases of perceptual knowledge—as a reporter acting as a truth-conducting signal for the recipient; the informational signal is the shared currency. This exemplifies testimony's transmission view by implying that the reporter is nothing more than an informational conduit, qualitatively no different from the signals produced

by all physical entities. In this respect, the Information-theoretic view not only exemplifies but represents the logical culmination of modeling epistemic interaction on testimony.

Yet Graham does not fully reduce testimony to the information-theoretic framework. I have already alluded to one of his qualifications, namely he limits his application to cases of perceptual knowledge. A second and more important restriction is that the conveyance of information is a necessary though *not* sufficient condition for knowledge. More specifically, Graham (2000b) thinks that justified belief is still an “important” component of knowledge. The Information-theoretic framework is thus only meant to be a *part* of the epistemology of testimony, namely the part that describes the underlying dynamics of at least some cases of testimony.⁹ In the present section, I detail an account closely related to the Graham-Dretske Information-theoretic view, namely Alvin Goldman’s reliabilism.

Goldman’s reliabilism, like the Information-theoretic model, suggests that more attention in social epistemology should be given to the aspect of transmission and less to the conscious justification of knowledge. The Information-theoretic view is reliabilist in its emphasis on truth-conduciveness, though Goldman is not committed to the concept of an informational signal. While Goldman’s account is thus not as stark an exemplar of the transmission view, Goldman offers a more comprehensive framework for testimony that is helpful for spelling out the implications of testimony’s transmission view.

Goldman’s reliabilism is aligned with his “epistemics” project.¹⁰ The general aim of *epistemics* is to correct traditional epistemology’s disregard for humans’ actual cognitive abilities

⁹ Graham’s ultimately qualified conclusion is a microcosm of a deep ambivalence within the testimony literature that, on the one hand, aims to explain the intrinsic epistemic importance of social activity but, on the other hand, is pervaded by conceptions of naturalism that erode the intrinsic significance of any social activity.

¹⁰ Goldman (1993) relates reliabilism to Virtue Epistemology; in short, reliable processes are intellectually virtuous ones. The connection offers reliabilism more terms for epistemic evaluation without changing its fundamental commitments, the fundamental commitments being my present focus.

(Goldman 1978). By correcting this lacuna, epistemics can provide practical recommendations for acquiring knowledge. Epistemics is thus not meant to be a complete epistemology, but only to cover the normative dimensions of traditional epistemology, prescribing ways to more efficiently acquire truths. Reliabilism serves as a jumping off point from traditional epistemology to Goldman's epistemics project, introducing a gap between the goals of acquiring true beliefs and of justifying one's beliefs. By recognizing the former—acquiring true beliefs—as the primary epistemic goal, epistemics entails a reprioritization of how to pursue knowledge.

In the following three sections dedicated to Alvin Goldman, I first discuss reliabilism and then his epistemics project as a whole. In the third and final section, I examine the social epistemology that follows from these commitments and relate Goldman's social epistemology back to the larger testimony literature.

Reliabilism's departure from traditional epistemology begins with its commitment to a causal theory of knowledge. Whether a given belief counts as knowledge depends in part on the chain of causes that led to its formation. The most salient segments of the causal chain—the parts that deserve epistemic appraisal—are belief-forming processes. Each belief forming process operates on a set of inputs and then generates a doxastic belief state as its output. Belief-dependent processes take doxastic inputs, while belief-independent processes only accept non-doxastic inputs. The latter kind of belief forming process (i.e., perceptual processes) serves as a foundation for belief-dependent ones. The concept of a belief forming process is essential to the appeal of reliabilism for it supports the picture of epistemic evaluation as the typing of specific cases.

As types, belief-forming processes are judged in terms of their token outputs. Perceptual and reasoning processes serve an important rhetorical purpose for reliabilism in that they are the

clearest examples of there being well-defined belief process *types*. Perceptual processes, in particular, suggest the core meaning of reliability: percepts most often accurately correspond to physical reality, as evidenced by the comparatively isolated occurrences of illusions or other misperceptions. The guiding idea is that for most people perceptual errors appear against a vastly larger backdrop of accurate perceptions. The accurate output tokens far outweigh the erroneous ones, making perception a reliable process. So unless there are specific reasons to doubt one's senses—such as unfavorable or extreme conditions, a medical diagnosis, the presence of a magician, and so on—then one is epistemically justified in accepting the outputs of perceptual processes. Thus, in addition to its previously noted role as a foundation for belief-dependent processes, perception also represents the core notion of a reliable though imperfect belief-forming process. The notion of a reliable though imperfect belief-forming process, in turn, suggests the appropriate responsibilities of an epistemic agent.

For reliabilism, the primary responsibility of the epistemic agent consists in being aware of what causal processes underlie a given belief. In contrast to the classic picture of directly justifying particular beliefs, performing one's epistemic responsibility is not necessarily sufficient for justification because reliabilism sets a broader task for the agent. In the epistemically praiseworthy cases, the belief-forming processes are truth-promoting. While perception and reasoning tend to promote true propositions (and so are truth-linked), there are other belief forming processes (e.g., wishful thinking and hasty generalization) that tend to promote falsehoods (Goldman 2011). The task of the epistemic agent is only to identify the causal process responsible for his or her belief: if from perception, then the belief is acceptable, whereas if from wishful thinking, then not.

The task of identifying the types of processes underlying a given belief is meant to be less burdensome than the classic picture of justification. In the case of perceptual processes, this is obvious since epistemic subjects only have to consciously scrutinize the grounds for perceptual beliefs in unusual circumstances. Even if an epistemic subject does not understand the causal mechanisms that underlie vision, hearing, and the like, the subject is justified in accepting the beliefs produced by these processes. Goldman's reliabilism generalizes this intuition: subjects also often lack knowledge about the inner workings of other belief-forming processes but are still capable of evaluating their resultant reliability. Yet despite the intent and intuition behind reliabilism, it is unclear just how modest this epistemic task actually is. In the next section, I examine this and other issues as they relate to Goldman's epistemics project.

1.5 Goldman's Epistemics: Some Major Issues

The biggest appeal of Goldman's epistemics is meant to be a dramatically more accurate and thereby practically useful account of epistemic agents and their epistemic pursuits. In this section, I offer a general overview of epistemics and then examine some of the major obstacles that it faces. One common theme among the various issues that I discuss is tempering epistemics' initial appeal of greater accuracy and usefulness.

1.5.1 Epistemics: An Overview

The name *epistemics* signifies Goldman's emphasis on inter-disciplinarity. Epistemics entails using the insights of empirical science, especially psychology, to inform our epistemic pursuits. This dovetails with the causal theory of knowledge discussed in the previous section and is opposed to the picture of epistemology as the self-contained conceptual analysis of epistemic terms (Goldman 1986). Given epistemics' goal to improve how people acquire knowledge, it is

clear that detailed information about the mental processes of cognitive subjects is important: any regulative enterprise should be concerned with the weaknesses of a system in order to address if not obviate them. The range of systems that fall under the domain of epistemics ranges from individual cognitive subjects to institutionally organized epistemic ventures.

There are two main branches of epistemics: the individual and social branches. Within the individual branch, there is, in turn, a distinction between primary and secondary individual epistemics (Goldman 1978). Primary epistemics consists in identifying the native cognitive processes of individual subjects. Most intellectual advances are not the product of native processes alone. Instead, learning methods, heuristics, algorithms as well as the employment of machines often play integral roles; acquired intellectual skills and capacities form the domain of secondary epistemics. Despite the lesser importance of primary epistemics, Goldman still thinks it is necessary for understanding how subjects make use of secondary acquired skills and tools.¹¹

The importance of the social branch of epistemics is closely related to and follows from secondary epistemics. Once the epistemic importance of acquired skills and tools is granted, then other people will necessarily be epistemically important as well. Other people play an obvious and vital role in the skills and tools that any given individual acquires in his or her lifetime. But properly understanding social epistemics requires ultimately rooting all social processes in individual belief-forming processes. Hence, epistemics has three essential levels in which the secondary level is a direct extension of the primary/native one, while the third (social) level is a direct extension of the first two.

In the rest of §1.5, I examine some of the major obstacles epistemics faces. The biggest issue is the so-called generality problem, which appears at multiple levels of epistemics. Another

¹¹ Andy Clark's Extended Mind hypothesis develops this idea to its fullest potential. And John Sutton forcefully criticizes it, both of which I discuss in Ch. 5.

issue is Goldman's conception of intelligence and its attendant picture of problem solving, which separates the intellectual and practical dimensions of intelligence. Each of these issues relates to the transmission view underlying the testimony literature, and both question epistemics' usefulness as a framework for guiding the pursuit of knowledge. Since the generality problem directly follows from discussion of justification in §1.4, I begin there.

1.5.2 Epistemics and The Generality Problem

The generality problem refers to the problem of properly typing the formation of a given belief. Any given belief is meant to result from a token causal process, that is, a concrete process occurring at a particular time and place. But it is unclear how to type a particular causal process, since it can be typed in multiple narrower or broader ways (Goldman 1979). Without a unique way of typing causal processes, there is no determinate reliability value for a token process. At the extreme, if the causal process is typed so narrowly as to only include one instance, then a bad belief-forming process may appear perfectly reliable if the particular belief happens to be true. The reverse could hold with respect to a good belief-formation process.

There have been a number of proposed solutions for the generality problem, but Goldman's favorite response is telling. Among Goldman's own proposals are considering only content-neutral causal processes and restricting the extent of belief forming processes to individual organisms' nervous systems. With respect to the latter proposal, Goldman acknowledges that he has been unable to stick to the restriction in his own explications (Goldman 2011).¹² As for the content-neutral restriction, it blocks the most extremely narrow

¹² Even if Goldman were able to restrict his explications to an individual nervous system, there would still be the possibility of multiple causal processes operant within an organism's nervous system.

cases of typing but still fails to pick out a unique belief forming process.¹³ Goldman (2011) concedes that his own and others' proposals have been inadequate but claims that reliabilism is in no worse shape than any other epistemology.

The generality problem is one manifestation of the basing relation problem. In the case of reliabilism, a belief is grounded on the basis of a belief formation type. Other epistemologies posit different basing relationships: an evidentialist epistemology, for example, will ground a particular belief on available evidence. The subsequent issue for evidentialism is determining the unique set of evidence that a particular belief should be judged in terms of. Parallel to the typing of causal processes, it is unclear how narrowly or broadly to delimit the evidential base. Goldman's claim is that every epistemology's account of justification involves some kind of basing relation; and, furthermore, as with reliabilism and evidentialism, there is no clear way to identify a unique way of basing particular beliefs. Thus, Goldman's ultimate answer to the generality problem is that every epistemology is equally subject to the underlying basing relation problem.

Goldman's diagnosis of the generality problem reflects, I think, the pervasiveness of testimony's transmission view. The basing relation problem is partly an artifact of grounding epistemic interaction in the exchange of a common currency, such as propositional beliefs. If epistemic interaction is grounded in some common currency, then the tokens of that currency are the fundamental unit of evaluation. Evaluation consists in judging discrete entities as tokens of a currency. And it is the discreteness of the tokens that creates the basing relation problem. In Ch. 2, I examine a model of epistemic interaction that does not ground epistemic interaction in the

¹³ Enactivism, as I discuss later, and practically-oriented epistemologies more generally problematize the very notion of a content-neutral epistemic process.

exchange of discrete tokens. This model and others suggest that the basing relation problem is not endemic to every epistemology and is thus a serious cause for concern.

In sum, the basing relation problem is related to a generic picture of epistemic evaluation that is assumed in many epistemologies. Thus far, I have related this picture of epistemic evaluation to testimony's transmission view, but in the next section—while examining Goldman's conception of intelligence and problem solving—I link this picture to an understanding of how knowledge relates to practical action.

1.5.3 Epistemics' Conception of Intelligence and The Basing Relation Problem

The second major problem that I want to discuss in this section concerns Goldman's need for a conception of intelligence. From the outset of his epistemics project, Goldman (1978) recognizes that the goal of acquiring true beliefs is by itself insufficient as a general epistemic standard. The acquisition of truths must be balanced with other considerations, such as error avoidance. The subsequent account of intelligence and problem solving suggests how knowledge relates to practical action. I begin with Ch. 6 of *Epistemology and Cognition*, in which Goldman introduces his notion of intelligence in an attempt to better describe what the acquisition of knowledge entails.

Intelligence incorporates two main elements, power and speed (Goldman 1986). Power largely corresponds to the ability to acquire true beliefs, which Goldman thinks consists primarily in fostering reliable belief-forming processes. Speed, by contrast, is fundamentally in tension with reliability. Heuristic problem solving—essential for solving many if not all problems—necessarily sacrifices at least some reliability. Given this tension, one of the main tasks that Goldman envisions for epistemics is evaluating how best to balance the values of power and speed.

Yet intelligence is much more complicated than the preceding picture of two dueling values suggests. In addition to efficiently acquiring truths while still avoiding errors, it is necessary to take into consideration other factors, such as how difficult it is to gain a truth as well as how significant a truth is. Without these additional considerations, intelligence could be mistaken for the stockpiling of trivial true beliefs or the solving of elaborate though meaningless esoteric problems. The issue of determining the relative significances of different truths highlights how Goldman envisions the relationship between the practical and intellectual dimensions of problem solving.

Goldman narrowly construes the issue, considering it in terms of an intellectual puzzle that is distinct from any practical matters. Practical matters supply epistemic goals: “cognitive mechanisms” try to acquire true beliefs on “*pretargeted* questions,” pretargeted questions referring to any questions relevant to “practical tasks at hand” (Goldman 1986; original italics). Epistemic questions are “pretargeted” in the sense that the resulting cognitive activity is wholly isolable from the practical task at hand. Cognitive activity is distinct from practical activities because the latter only play a subsidiary role in knowledge acquisition.

Practical activities serve epistemic purposes but do not directly contribute to intellectual activity. In addition to supplying target questions, practical activities are only meant to retrieve information from the environment, test possible solutions, and, ideally, realize solutions. This division reflects a distinction between mental and physical operations, which in turn relates back to Goldman’s privileging of native cognitive processes. Mental operations, as native cognitive processes, are thought to control problem solving activities (e.g., recognizing the need for more information and so initiating more retrieval).

Goldman's narrow construal of problem solving implies that knowledge exists independent of its practical implications and uses. After a subject acquires a piece of knowledge, it can then inform his or her subsequent action.¹⁴ Whether the knowledge actually informs the agent's actions depends upon non-epistemic aspects of the agent. This generic picture of how knowledge relates to practical action underlies the transmission view within the testimony literature. Dividing knowledge from practical action implies that knowledge must be grounded in a discrete entity: knowledge is something that can be absent one moment but present the next or something that can be drawn upon or ignored. Tracing the transmission view back to this generic picture of how knowledge relates to its practical uses explains why it is an under-theorized, implicitly presupposed part of the testimony literature. In sum, one of the origins of the basing relation problem is a seemingly obvious but ultimately flawed picture of how knowledge relates to action.

Having used Goldman's account of intelligence to trace the basing relation problem back to another pervasive assumption in epistemology, I now turn to two other aspects of his account. In §1.5.4, I focus on the issue of whether epistemics is founded on an inadequate description of our epistemic pursuits. In particular, its account of intelligence seems to make discoveries an epistemological mystery. To the extent that Goldman can address this concern, it requires appealing to his social epistemology, which I examine in §1.6.

1.5.4 Epistemics' Conception of Intelligence and the Pursuit of Knowledge

As discussed in the previous section, Goldman construes problem solving narrowly. Epistemically significant problems are self-contained intellectual puzzles, isolable from their practical implications and uses. Elaborating on this picture, Goldman emphasizes the

¹⁴ Both the translation and transformation views challenge the strict division between knowledge acquisition and use.

purposiveness of problem solving, using the game of billiards as a metaphor. In the game of billiards, a pocketed ball only counts if the player “called” the shot. Analogously, an intelligent person finds the solution to the problem that he or she had sought (Goldman 1986). Goldman recognizes that when attempting to solve a particular problem, it is possible to serendipitously discover other truths. Lumping such discoveries under the generic label of “incidental beliefs,” Goldman minimizes their importance by indexing them to separate problems that deserve their own “called” shot. One issue raised by this picture—particularly with respect to large-scale epistemic ventures—is whether this is an adequate characterization of how knowledge is acquired.

Goldman minimizes the importance of serendipitous discoveries on principle. This can be explained using the overview of epistemics given above. Serendipitous discoveries cannot be a direct product of an individual’s “native cognitive proclivities.” As for the acquired learning methods and tools of secondary epistemics, Goldman (1978) thinks that these must be understood in terms of how they are employed by native processes. Thus, secondary epistemics is similarly ill-suited to consider the epistemic significance of serendipity. The general approach of individual epistemics is to seek the essential lawful regularities—which are founded on a native cognitive core—that underlie the messy details of a particular instance of cognition.¹⁵ Lawful cognitive regularities—whether manifested in terms of a native capacity, an acquired skill or tool, or some higher-level sociological pattern of behavior—are the only kind of descriptions that Goldman considers epistemically significant. The strictly purposive nature of

¹⁵ John Sutton challenges the premise of Goldman’s focus by critiquing the notion of a substantial immutable cognitive core. I discuss this issue in Ch. 5.

problem solving reflects this descriptive focus on lawful regularities. I think that this picture of purposive intelligence obscures important dimensions of cognitive agency.¹⁶

With respect to the topic of discovery, the immediate concern is whether this picture of intelligence makes discovering knowledge an epistemological mystery. Goldman does not think his emphasis on purposiveness is an impediment to discovery because individuals are members of larger communities. Even if the picture of intelligence suggests that individuals are easily locked into a particular puzzle, it is not a vicious problem so long as another individual's insight can break the deadlock. But for this to be an essential element in the pursuit of knowledge, there needs to be a mechanism that supports constructive epistemic relationships between individuals. Goldman's product epistemology—which bridges the individual and social branches of epistemics—explains how individuals can help each other even when dealing with extremely complex and specialized problems.

Product epistemology marks a shift from “belief epistemology.” The shift consists of identifying and examining specialized forms of communication within particular communities. A community's form of communication is based on intellectual products, by which Goldman means the technical aspects of communication (i.e., the preferred medium(s) of communication) as well as the mode of discourse (e.g., assertive debate). For example, twentieth century scientific communities employed an assertive mode of discourse (as opposed to something like the literary community's use of a depictive mode) whose preferred medium was peer-reviewed journal articles (as opposed to mediums like novels). A community's specialized form of communication is a mechanism for more effectively and efficiently circulating beliefs. The different mediums and modes of discourse represent different problem-solving heuristics, whose

¹⁶ My most substantial critique of the picture of lawful regularities is my discussion of nonrepresentational embodied directedness in Chs. 3 – 5.

suitability depends on the aims of the intellectual venture. For science, the efficient acquisition of knowledge would be hindered by novels' depictive mode while, conversely, novels' goal of personifying characters would be hindered by the assertive mode. The concept of an intellectual product plays a key role, for Goldman, in explaining how individual epistemic agents retain their individuality while contributing to a larger community.

Goldman thinks that social epistemology is simply an extension of individual epistemology. The constraints of creating an intellectual product ensure that an individual's epistemic abilities are channeled into a form that has a greater chance of being useful to another individual working on the same or similar problem. As the name "product" suggests, the specialized form of communication is an instrumental resource that individuals draw upon to augment their own capacities. While product epistemology thus offers some insight into how individuals can *qua individuals* effectively collaborate, the proposal's appeal is tempered by the re-emergence of the typing problem.

Product epistemology is a more fine-grained version of belief epistemology, but it is unclear how fine-grained it should be. Goldman (1986) himself acknowledges, for example, that despite sharing the generic mode and medium, there are important stylistic differences within scientific sub-communities. Analogous to the issue of typing a uniquely important belief-forming process, it is subsequently unclear how broadly or narrowly to delineate the uniquely important level of community. In contrast to the case of belief-formation, product epistemology seems to call for as fine-grained a delineation as possible, since its chief task is to explain how individuals collaborate. But the typing cannot be so fine-grained as to lose its relevance to larger scientific communities.

The typing of intellectual products draws attention to institutional hierarchies. Typing intellectual products is a less pressing issue than typing belief-forming processes more generally for an epistemic agent engaged in a particular problem. While the latter is a central epistemic responsibility, the former only entails broad constraints on how an agent channels her cognitive abilities. Sticking with the example of a twentieth century scientist: the issues of communicative mode and medium (e.g., how to write journal articles, striking the right argumentative tone, and so on) are not necessarily pressing concerns for the practicing scientist. Yet such matters are abiding concerns that are open to change. Such broad concerns and changes are the purview of an epistemic community's authorities rather than a single agent. In the case of science: journal editors, grant committees, and the like are in positions to judge and enact reforms concerning their community's intellectual products and other broad ranging matters. The issue of typing intellectual products thus points to the importance of social epistemology for Goldman's epistemics, since the issue is not as intractable when it is considered from the perspective of authoritative third-person intervention.

In this section, I have further examined Goldman's account of intelligence as well as his concept of a product epistemology. My main concern has been to question whether they adequately characterize the pursuit of knowledge. Purposive intelligence was problematic on this score because it seemed to make discoveries an epistemological mystery; but the issue was partly ameliorated by situating the individual epistemic agent within a larger community. This response raised the question of how individuals could collaborate on highly complex problems, a concern partly addressed by Goldman's product epistemology. Product epistemology, in turn, reiterated the issue of typing belief-forming processes. When considered in terms of product epistemology, the typing issue actually appears less intractable since it highlights the importance of social

hierarchies and third-person regulation within epistemic institutions. In order to more fully evaluate these proposals, I turn to Goldman's social epistemology.

1.6 Truth-Oriented Social Epistemology

Taking an avowedly conservative stance, Goldman argues social epistemology is an extension of individual epistemology. The epistemic significance of social interaction is explicable entirely in terms of social interaction's causal influence on an individual's beliefs. In this respect, there is no qualitative difference between social interaction and other external causal influences—all that epistemically matters is how an external force influences an individual's stock of beliefs. An epistemically praiseworthy force promotes true beliefs, and, in situations where true beliefs are valued over false beliefs or no belief at all, Goldman (1999) refers to the beliefs as possessing “veritistic value” or “V-value.” Hence, Goldman's approach to social epistemology is to examine ways in which individuals promote true beliefs in other individuals, increasing the V-value within a community.

Social epistemology is its own domain of enquiry because there are distinctive social influences on beliefs. Product epistemology is one example of a social process that influences individuals' beliefs by enabling specialized communication. Social processes exert an immense influence on individuals' beliefs simply because people are more finely tuned channels of information than other external causal influences. Most importantly, for Goldman, people share the medium of language, which enables testimonial practices.

1.6.1 Testimony within Goldman's Social Epistemics

Goldman endorses much of the generic picture of what testimony is. Testimony consists in the circulation of beliefs via language.¹⁷ Goldman defines the domain of testimony as factual discourse (i.e., simple observational reports). As such, testimony is the “most elementary and universal social path to knowledge” (Goldman 1999). As the sharing of observational reports via language, testimony is the most basic kind of intellectual product. Later product epistemologies represent specialized forms of testimony, both in terms of developing terminologies (i.e., specialized languages) and mechanisms for sharing the reports. Goldman's approach to testimony, like any other product epistemology, is to identify its most significant causal mechanisms in the hopes of improving its effect on a community's V-value.¹⁸

Goldman recognizes four stages of testimony-related activities. They are as follows: “(1) discovery, (2) production and transmission of messages, (3) message reception, and (4) message acceptance” (Goldman 1999). With respect to (1), Goldman endorses Philip Kitcher's picture of the cognitive division of labor (Goldman and Shaked 1991): the self-interested egoism of individual agents ensures optimal specialization and distribution of risk within a larger epistemic community (Kitcher 1990). Most directly related to discovery: systems of credit reward—best exemplified in the practice of science—entice some individuals to explore radically alternative solutions. Kitcher, like Goldman, thus considers social epistemology an extension of individual epistemology.

¹⁷ Lackey (2006b) argues that some bodily gestures should count as testimonial acts. But this is only a minor revision of the general identification of testimony with language acts, since Lackey considers such bodily gestures as corresponding to propositional beliefs.

¹⁸ Miranda Fricker—whose work on epistemic injustice represents perhaps the most noteworthy ethical application of the concept of testimony—lauds Goldman's idea of the veritistic assessment of communities. I do not extensively discuss M. Fricker (1998) because it limits itself to applying the concept of testimony rather than scrutinizing it.

The second stage of the testimony model corresponds to reporter practices. Goldman gives his most extensive general account of reporting practices in the subsequent chapter on argumentation. In keeping with his narrow construal of intelligence and problem solving, Goldman sharply delineates argumentation from practical decision-making. In strictly epistemic terms, argumentation refers to how a reporter best ‘advertises’ the truth of a conclusion (Goldman 1999). In the simplest case—monological argumentation—the reporter-audience interaction is one-directional, analogous to an isolated observational report.¹⁹ Dialogical argumentation builds off this account—identifying the basic ways in which a recipient can challenge the “advertised” truth. Apart from this model of argumentation, the only other aspect of reporting practices that interests Goldman concerns what types of beliefs are most valuable for reporting. Given the general goal of increasing veritistic value within a community, the most valuable beliefs to report are ones that hold a low “degree of belief.”

While testimony’s first two stages receive only brief attention, the third stage of message reception garners none. In a footnote, Goldman (1999) assumes that testimonial audiences “properly understand” received messages, reducing message reception to the psychology of language comprehension.²⁰ Goldman’s treatment of the third stage draws attention to the decisive cleavage between reporter transmission and message reception, as implied by recognizing monological argumentation as the basic form of reporting.²¹ The receiver’s role is strictly as a passive decoder of the transmitted report. Although Goldman’s cursory treatment of

¹⁹ In his account of argumentation, Goldman considers the issue of enthymemes, the acceptable use of which he claims are regulated by “folk rules.” In discussing coordinated interaction in Ch. 3, I argue that factoring in the pragmatic context of action is a more direct and satisfactory account of this issue.

²⁰ In terms of Goldman’s four stages, it is possible to explain M. Fricker’s (2007) concept of testimonial injustice as occurring at the third stage: that is, due to prejudice a person is denied the capacity of being a knower, which prevents his or her testimony from even being evaluated. This reading, of course, suggests that Goldman’s treatment of the third stage is far too cursory.

²¹ Goldman’s foundation of monological argumentation corresponds to Adler’s description of testimony’s core cases as being only one-directional.

message reception appears to be a harmless assumption from the perspective of the testimony literature, Stephen Turner's weak empathy model emphatically denies its truth. Turner argues that the lack of proper reception and understanding is one of the defining features of epistemic interaction.²² I examine this point in Ch. 2 but, for the moment, only use it to further highlight the premises of testimony's transmission view.

The fourth stage of testimonial activity—message acceptance—is importantly different from the question of epistemic justification. As discussed in §1.4, Goldman uses reliabilism to differentiate between the goals of acquiring true beliefs and acquiring justified beliefs. This sets the stage for Goldman's most distinctive contribution to the testimony literature, namely his proposal for applying Bayesian inference procedures to testimonial practices. When applied to testimonial reports, Bayesian inference is presented as a universal framework for increasing V-value by regulating when a message ought to be accepted.²³ The basic proposal is as follows: given a testimonial report T, if T possesses an objective likelihood magnitude X corresponding to a true answer for a question Q, and if a receiver R possesses X, then employing Bayes' Theorem is objectively likely to improve R's stock of true beliefs regarding Q. Bayesianism does not, however, necessarily justify the belief since R may not be able to justify certain likelihood magnitudes (e.g., perhaps R adopted a likelihood value from elsewhere, relying on another agent's assignment). Apart from the issue of justification, there are number of concerns with the Bayesian proposal.

I will touch on two concerns with the Bayesian proposal before linking it to epistemic paternalism. Both concerns relate to Goldman's "crucial assumptions" concerning the proposal,

²² The importance of "shallow consensus" for collaborative enquiry, in which understanding between individuals may be extremely tenuous—as discussed by Fred D'Agostino (2010)—offers another perspective on the doubtfulness of Goldman's assumption.

²³ The inspiration for Goldman's proposal stems from Richard Friedman's application of Bayes' Theorem to evaluating witness testimony within the legal context.

namely that a given piece of testimony has an objective likelihood value and that the receiver possesses that value. Needless to say, Bayes' Theorem itself offers no special insight on how an epistemic agent ought to acquire the likelihood values. That Goldman does not consider this a pressing issue relates back to the former point regarding objective likelihood values: the Bayesian proposal depends on there being fairly obvious measures of truth. Acquiring likelihoods values would ideally be as simple as having an extensive chart of previous results. But there are good reasons to doubt whether this is an adequate account or even plausible.

Goldman himself acknowledges that Bayes' Theorem may not be the most effective inference practice for particular problems. His argument is only meant to establish that Bayesian inference is likely to have a positive effect for all epistemic practices. While this may be an epistemological curiosity, it does not immediately follow that it should actually be applied to epistemic practices since a given practice may already employ a more effective inference practice. This is an especially telling issue given the second concern, namely the difficulty of actually employing Bayes' Theorem. Largely based on the difficulty of its application, the use of Bayesian inference has been rejected in the British legal system (Kusch 2002a). In discussing the court's rejection, Martin Kusch (2002a, 81) concludes that evaluating the credibility and plausibility of evidence and witnesses—necessary for establishing accurate likelihood values in the legal context—are “not the sorts of things that are best thought of in discrete quantities.” In such cases where relevant factors cannot be expressed in terms of discrete quantities, it is impossible to properly perform a Bayesian calculation. But, setting aside these cases, it should be granted that the difficulty of Bayesian inference is mitigated when considered from the perspective of an epistemic authority, especially an authority who is attempting to evaluate the

efficacy of a large-scale and long-term epistemic enterprise. This brings me to the element of epistemic paternalism within Goldman's social epistemology.

1.6.2 Epistemic Paternalism and Epistemics' Two Loci of Control

The most charitable way of reading Goldman's Bayesian proposal is from the perspective of authoritative regulation. To discuss the importance of epistemic authorities for Goldman, I will examine the notion of epistemic paternalism, which is one form of authority. Discussing the authoritative role within epistemic paternalism is also helpful for understanding the normative role that Goldman envisions for social epistemics. In the present section, I focus on epistemic paternalism and its attendant picture of epistemic agency, while in the next and final section on Goldman's social epistemology, I link epistemic paternalism back to Bayesianism.

Epistemic paternalism entails that not all relevant and available information should always be heard. Epistemic paternalism is defined as "whenever a 'communication controller' interposes their own judgment rather than allowing the audience to exercise theirs" (Goldman 1991). The specific type of authority being exercised in epistemic paternalism is communication control, the authority filtering out any information that would have a veritistically negative effect. A wide variety of domains exhibit this dynamic, such as legal and education systems. If the significance of a piece of information is easily over- or under-estimated by its audience, then the authority may filter it out so that the audience's judgment is not likely misled. In the case of the legal system, for example, hearsay and the accused's criminal history in addition to the previously mentioned case of Bayesian inference are all excluded for fear of being over-estimated by juries.²⁴ In the former two cases—hearsay and the accused's criminal history—the

²⁴ One of the considerations that blocked the use of Bayesian inference in juror deliberation was the claim that jurors are likely to under-estimate or effectively forget the subjective judgments that underlie a set of

information is withheld despite being directly relevant. Yet Goldman considers the withholding as justified on epistemic grounds: blocking relevant information is *epistemically* justified if it improves the chances of acquiring the sought-after truth, such as the guilt or innocence of an accused.

The authoritative, paternalistic filtering of information represents one of two loci of cognitive control that Goldman recognizes. In a footnote from the 1978 *Epistemics* essay, Goldman states that epistemics covers both “the self-regulation of cognitive processes” as well as the “*third-person* control of cognitive traits” (520; original emphasis). Goldman’s goal is to expand traditional epistemology’s purview beyond the individual; in the passage, the specific concern is with how tool-use and education shape cognitive subjects. Goldman implicitly commits himself to the view that first- and third-person regulation are the *only* two loci of cognitive regulation. Either the epistemic subject imposes a change on herself, or a third-person (as a specially tuned external force) imposes the change on the subject. While this is an improvement on traditional epistemology’s exclusive focus on self-regulation, I think there is at the very least one other loci of cognitive intervention, namely second-person participation. But the importance of second-person participation to epistemology does not appear as a possibility to Goldman. In Chs. 2 – 3, I use Turner’s weak empathy model and the enactivist concept of participatory sense-making to argue that second-person interaction is of distinct if not primary epistemic importance. But for present purposes, I mention the possibility of second-person participation in order to highlight another under-theorized element of the testimony view’s transmission picture, namely the implicit passivity of epistemic subjects.

likelihood values. In such a scenario, the conclusions based on Bayesian inference would be over-estimated due to an exaggerated appearance of objective certainty (Kusch 2002a).

One way to examine the passivity of epistemic subjects is by considering the parallels between first- and third-person regulation. Goldman (1978) notes that people do not have direct control over doxastic attitudes and that the most important “command variable” (i.e., cognitive element that an individual does control) is attention. This consideration relates to the importance of secondary epistemics, especially how practicing certain skills can be used to mold cognitive traits. And once the importance of secondary epistemics is granted, it follows that third-person interventions importantly influence individuals’ cognitive traits via secondary epistemics as well. Whatever can be counted as a “command variable” for an individual is susceptible to third-person control, beginning with attention.²⁵ Given the profound influence of social circumstances and external forces more generally on an individual’s attention, the very notion of a “command variable” is at best a difference of degree rather than kind from the non-voluntaristic doxastic attitude.²⁶ The essence of cognitive control thus suggested is that of *filtering*, and it is in this sense that Goldman’s cognitive subject is essentially passive.

The essence of epistemic agency consists in parsing perceptual inputs as opposed to a world-directed behavior. The epistemic agent does not control the causally determinate nature of perceptual input but does exercise some control over what input she is exposed to as well as what aspects of the input are taken as salient. A person, for example, may not be able to control the effect of advertising but can exercise some control as to what advertising they are exposed to (e.g., by avoiding TV). This is the basic sense in which *filtering* is the essence of epistemic agency. And the ultimate result of epistemically skilled filtering is increasing one’s stock of true beliefs. The molding of cognitive traits can be understood as types of filters: such habituated

²⁵ The potency of advertising makes this a ubiquitous and non-trivial point.

²⁶ I argue in Ch. 2 that the lack of a qualitative difference between Goldman’s command variables and non-voluntaristic cognitive states stems from grounding knowledge in causal determination. I further explicate and clarify this aspect of Goldman’s view of the cognitive subject by comparing his simulationist model of the mind to Turner’s weak empathy model.

traits influence what a cognitive subject exposes herself to, what she recognizes as salient, and also forms heuristics for processing the salient input. Likewise, other people—strictly in terms of their epistemic significance—are nothing more than particularly effective filters.

Goldman recognizes social epistemics as a distinct domain because other people often act as specialized filters, filtering information in ways analogous to an individual's own acquired secondary cognitive skills. Other people essentially predigest and condense perceptual input. What another person shares reflects in part what they thought of as salient in addition to being the product of their own acquired epistemic skills: for these two reasons, the testimony of another person has at least the potential and often the actuality of providing true beliefs more efficiently than if an individual had to undergo the corresponding first-hand experiences herself. Epistemic paternalism, in turn, is one special type of social filter; in the case of epistemic paternalism, “pre-digestion” involves intentionally withholding relevant but likely misleading information. As with other kinds of social filters as well as an individual's own acquired secondary cognitive skills, the benefit of epistemic paternalism derives from filtering input in order to lighten the burden on an individual's cognitive processes.

In the present section, I examined epistemic paternalism in order to explore the nature of epistemic agency. Epistemic paternalism provides a helpful perspective on the issue by drawing attention to the second locus of cognitive control, namely third-person regulation. After comparing third-person regulation with self-regulation, I concluded that for Goldman there is no qualitative difference in the nature of control for each locus because individual “command variables” are susceptible to third-person regulation. The comparison also highlighted the passive nature of epistemic agency: the essence of epistemic agency—whether performed by an individual or by another person for an individual—consists in filtering informational inputs in

order to more efficiently stockpile true beliefs. As noted at the beginning of the section, the authoritative role depicted in cases of epistemic paternalism is the most charitable way of considering Goldman's Bayesian proposal. I expand on this idea in the next section, while presenting a summary overview of Goldman's truth-oriented social epistemology.

1.6.3 Concluding Thoughts on Truth-Oriented Social Epistemology and Epistemics

Alvin Goldman's truth-oriented social epistemology includes one of the most comprehensive frameworks for testimony. Beginning with §1.4, I examined how reliabilism enables Goldman to avoid many of the justificatory issues that limit the testimony literature. Epistemics, in both its individual and social branches, is instead directed towards prescribing ways to improve the efficiency of knowledge acquisition.

In §1.5, I discussed the epistemics project as a whole and some of its major issues. The individual branch of epistemics is the most important branch, since all matters concerning knowledge ultimately terminate in an individual's belief-forming processes. Even though social epistemics is merely an extension of individual epistemics, the third major issue that I discussed—namely, accounting for the nature of discoveries—indicates the need for social epistemics. Along these lines, the concept of product epistemology explains how individuals can collaborate on problems such as those found in science.

Language is, in effect, the most basic kind of product epistemology. More than anything, social epistemology is a distinct domain of enquiry because the medium of language makes people particularly influential on each other's stock of beliefs. It is upon this broad framework that Goldman approaches the topic of testimony, rooting testimony in the use of language for transmitting observational reports.

As discussed in §1.6.1, Goldman adopts most of the basic tenets of the testimony literature. His most noteworthy contribution consists in his proposal to universally apply Bayes' Theorem to testimonial practices. Bayesian inference is shown to likely increase the amount of true beliefs within a community despite not necessarily providing epistemic justification for those beliefs. The increase in veritistic value (i.e., "V-value") can be verified by a third-person authority who monitors beliefs within the community.

The specific type of regulatory control that I focused on in §1.6.2 was epistemic paternalism. I used this specific type of third-person intervention to highlight the essentially passive nature of agency in Goldman's epistemics and the testimony literature more generally. Whether first-personal or third-personal, epistemic agency consists in filtering informational inputs in order to stockpile true beliefs. Epistemic agency is essentially passive because it is limited to how an agent receives information. From a strictly epistemic viewpoint, Goldman sees a cleavage between a source of information (whether a person or other external force) and the informational receiver. This cleavage manifested itself in Goldman's testimony account in the claim that monological argumentation is the basic and isolable form of testimony reporting.

The cleavage between testimonial reporter and receiver creates a motivation problem. An individual must make an effort to report information to another person, and it is not always clear why individuals expend such effort to share information. Goldman recognizes and briefly canvasses possible solutions to the motivation problem in the 1999 chapter on testimony. The general solution that Goldman seeks is an evolutionary selective pressure that favors communication, which is most plausible when thought of in terms of examples such as animals warning each other of threats. Over the long-term, individuals would mutually benefit from such communication. But Goldman recognizes that such an explanation is not entirely satisfactory,

since many animals communicate when there is no clear instrumental benefit and even when there is a negative cost—the example that Goldman notes is of chickens who share when they find a food source. While instrumental benefits are obviously a motivating factor in epistemic interaction, I think the attempt to make it the fundamental motivating factor results from the mistaken picture of there being a cleavage between reporter and receiver.²⁷

The very idea of a fundamental motivation to communicate is only plausible given the larger picture of a default cleavage between testimonial reporter and receiver. Seeking a fundamental motivation to be social is an example of what Martin Kusch (2002b)—with regard to the testimony literature as a whole—diagnoses as the dilution of social life. Like other contributors to the testimony literature, Goldman considers the details of actual interaction to be of only secondary/derivative importance to testimony.

The details of social interaction are epistemically significant only insofar as they relate to the transmission of beliefs. Interaction is strictly an instrumental means for gaining information more efficiently than would be required for first-hand experience. So even though Goldman rightfully deprioritizes the question of justification—allowing him to focus on ways to more efficiently transmit true beliefs—he still myopically focuses on message acceptance. The question of whether to accept a belief still crowds out the interactive dynamics between two agents. Goldman’s “aim realism” offers another, broader perspective on Goldman’s distinctive form of social myopia.

Goldman’s truth-oriented social epistemology is an expression of his commitment to aim realism. As presented in Goldman (1986), aim realism rests on a fitting correspondence theory of

²⁷ Another troubling feature of Goldman’s discussion is the implicit equivocation of evolution with natural selection. The adaptationist focus on the survival of the fittest brings with it the presumption that social relationships should be explained strictly in terms of instrumental benefits. In Chs. 3 – 5, using the transformation view, I argue that naturalism does not warrant such a reduction of social relations. Regarding adaptationism, see Gould and Lewontin (1979).

truth.²⁸ Language, as formulated by epistemic agents, provides the conditions of truth but physical reality still determines whether language fits.²⁹ Thus, an agent may focus solely on reformulating terminology (e.g., attempting to solve a theoretical puzzle), but what gives ultimate significance to such exercises is the possibility of providing a better fit to reality. Truth-oriented social epistemology refers to the same dynamic but explicitly at the level of a community. Analogous to esoteric theoretical puzzles with no direct connection to factual truths, certain measures may be instituted in a community even though they may seem to be unrelated or even a hindrance to truth. Epistemic paternalism is a good example in this respect because there is a clear sense in which truth appears to be hindered. But Goldman thinks that the ultimate effects are veritistically positive: a community will ultimately possess a greater amount of true beliefs. Hence, Goldman's social epistemology is *truth-oriented* in the sense that it concerns itself with issues that do not immediately relate to truths but which promote the acquisition of truths in the long-term; it *orients* a community to acquiring truths.

By concerning itself with the means of acquiring truths, truth-oriented social epistemology gives a more comprehensive treatment of testimony. But, as I have argued, Goldman's treatment is still myopic due to neglecting the interactive dynamics between individuals. In §1.5, this issue appears in the form of delineating between the practical and intellectual dimensions of problem solving, while in §1.6.2, I traced the problem back to the essentially passive understanding of epistemic agency. The only means that Goldman concerns himself with—whether from a first- or third-person perspective—is how to filter informational inputs. Both Turner's weak empathy model and enactivism's participatory sense-making offer

²⁸ This conception of truth is the metaphysical underpinning of reliabilism, explaining why it is not merely a heuristic for traditional justification.

²⁹ Martin Kusch (2002a) argues that Goldman's fitting conception of truth correspondence is either vacuous or contradictory.

active conceptions of epistemic agency. But before examining these fundamentally different approaches to epistemic interaction in the subsequent chapters, I first want to discuss Martin Kusch's internal critique of the testimony literature. Kusch's critique serves as a helpful transition because it anticipates the more pragmatically oriented accounts, yet still retains the testimony literature's focus on language as the primary medium of social interaction.

1.7 Kusch's Communitarian Critique of the Testimony Literature

In his call for a communitarian epistemology, Martin Kusch (2002a, 2002b, 2009, 2013) offers perhaps the most thorough and fundamental internal critique of the testimony literature. Kusch largely limits himself to one of the basic premises of the testimony literature—namely that language is the fundamental medium of epistemically significant social interaction—but offers a dramatically different characterization of that medium. Rather than thinking of language as solely or even primarily a medium for transmission, Kusch emphasizes the performative nature of linguistic acts. Kusch's communitarian epistemology thus provides a useful contrast to the transmission view underlying the testimony literature.

Knowledge by Agreement represents Kusch's most extensive attempt to redirect the testimony literature towards communitarian epistemology. Kusch critiques what he refers to as the "individualist testimony view," which corresponds to what I have labeled as testimony's "transmission view." Kusch (2002a) summarizes the individualist testimony view in terms of four aspects: (1) testimony is taken to be a merely derivative as opposed to generative source of knowledge, (2) testimony is meant to consist in the transmission of complete items of knowledge, (3) items of knowledge are delivered solely by perception, memory, and inference, and (4) the "testifier" and "recipient" need possess only minimal social knowledge of each other. The first

aspect concerning the transmission versus generation of knowledge is the most important, as discussed in §1.3.

In §1.3, I evaluated Lackey's argument that testimonial exchanges generate rather than simply transmit knowledge. But Lackey still bases testimonial exchanges on the transmission of preformed beliefs. Given Kusch's four aspects taxonomy, we can explain Lackey's position in terms of adopting all of the four aspects except the first one. By contrast, for Kusch, the testimonial generation of knowledge refers primarily to the performative nature of linguistic acts (e.g., certifying two people as married). The other three aspects help to further illustrate the performative nature of testimony.

The second and third aspects of the individualist testimony view are closely interrelated. Testimony is thought to only consist of transmitting complete items of knowledge (the second aspect) because items of knowledge are supposed to be formed only by an individual's on-board cognitive resources, namely perception, memory, and inference (the third aspect). The individualist testimony view thus endorses a four sources view of knowledge, in which testimony is the fourth and strictly derivative source that encapsulates all social sources of knowledge.³⁰ The fourth aspect of the individualist testimony view offers another perspective on the issues with the four sources picture.

The testimony literature generally disregards the significance of social identities and statuses. The fourth aspect of the individualist testimony view is the claim that testifier and recipient need only possess minimal social knowledge of each other.³¹ This aspect is most clear for theorists who focus on the question of justification in idealized terms. But even John Greco's

³⁰ Turner's weak empathy model challenges this four sources of knowledge picture and thus the first three aspects of the individualist testimony view.

³¹ The transformation view, especially as represented by Sutton's encultured mind, challenges all four aspects of the individualist testimony view.

account of testimony is lacking despite arguing that some social environments are epistemically relevant. As discussed in §1.2.3, Greco's argument is salutary because it alludes to the importance of social roles but remains impoverished. Greco only considers paternalistic relationships—such as doctor-patient and mother-child relationships—because these situations have a dramatic impact on standards of justification. Kusch, by contrast, thinks of epistemic communities in more general and complex terms, claiming that testifier and recipient are usually part of the same community and, furthermore, that the shared membership matters to the testimonial exchange. More specifically, the shared membership influences the normative expectations for both testifier and recipient.³² Paternalistic relationships are just one example of how community membership can influence normative expectations; Greco thus only indirectly alludes to the multifaceted epistemic significance of community membership. Before considering more of Kusch's communitarian alternative to the individualist testimony view, it's worth first noting his diagnosis of the intuition that underlies and guides the individualist view.

In "The Limits of Testimony," Ch. 1 of *Knowledge by Agreement*, Kusch argues that the original legal meaning of testimony still guides the literature. Kusch thinks that the legal sense of testimony serves as an intuition pump that "lurks behind" philosophical theories of testimony, despite attempts to broaden its meaning. In the legal context, an eyewitness ought to pass along knowledge founded on perception and memory, not hearsay. The witness's testimony should not involve any generative element because it would corrupt the testimony.

The legal meaning of testimony nicely encapsulates the four aspects of the individualist view. Testimony in the philosophical sense, as in the courtroom sense, should only transmit knowledge. What is transmitted should be a belief formed by the epistemic agent's/witness's on-

³² Fred D'Agostino (2010) is especially insightful on this point, describing how shared epistemic values within communities facilitate collaboration in the midst of widespread disagreement. This situation corresponds to the previously noted notion of "shallow consensus."

board cognitive resources (the third aspect), which in turn implies the second aspect of having formed a complete belief prior to transmission. Finally, the reason why testifier and recipient need only possess minimal social knowledge of each other is because the physical reality responsible for the first-hand perceptual experience is the defining and isolable element on which the testimonial exchange rests. Similar to how the details of social interaction are treated, social roles and epistemic communities are of only incidental importance. Social statuses may only be useful as a means for determining whether a piece of testimony is credible/uncorrupted or an indication that standards of justification should be relaxed.

Having summarized the individualist testimony view, Kusch points out cases that do not fit the picture. One immediate misfit is the previously mentioned example of certifying two people as married, which is an instance of collective stipulation. Testimony in this case creates the status of marriage as opposed to reporting a fact of nature. Collective stipulation is an obvious example of a linguistic act that is essentially performative. Another example that Kusch gives is of a judge pronouncing an accused person as guilty—the judge’s official capacity entitles him to make the pronouncement and for it to be of consequence. In other words, by recognizing a person as serving the role of judge, the corresponding community endorses his pronouncement and in that sense collectively stipulates that the accused person is guilty. But the performative nature of language goes beyond such obvious examples of collective stipulation.

Three of the primary sources for Kusch’s characterization of language are the late Wittgenstein (1958, 1969), Barry Barnes’s (1983) account of language in relation to the philosophy of science, and Edward Craig’s (1999) conceptual genealogy of knowledge. Most pertinent for social epistemology is Kusch’s use of Craig, since Craig claims that the very concept of knowledge originated as a social status marker. More specifically, Craig claims that

the concept of knowledge originated due to the need to mark certain people as good informants. That is to say, the concept of knowledge (and thus the very foundation of epistemology) is rooted in an act of collective stipulation. Hence, Kusch thinks that knowledge is itself a social kind as opposed to a natural or artificial kind.

Kusch (1998, 2002b) presents the tripartite division of natural, artificial, and social kinds. The tripartite division modifies Barnes's previous dichotomy between natural and social kinds. Elephants and roses are examples of natural kinds—they are organisms with their own particular reproductive histories. Artificial kinds, in turn, correspond to material artifacts: the two examples that Kusch offers are typewriters and tables. The idea is that both typewriters and tables, as material objects, were invented by people but would continue to exist even if humans suddenly disappeared. Social kinds are also invented by humans but their existence depends on people's continued recognition. Money is a social kind because its value derives solely from its being collectively recognized as valuable. If people suddenly disappeared, then money would as well, leaving behind token traces in the form of metal discs and such. More specifically, social kinds like money as well as marriage depend on the existence of language, because language enables collective stipulation (Kusch 2002b). Relating this division back to Craig: knowledge is simply another form of collective stipulation, stipulating people who inhabit the social role of being informative. Without people and the medium of language for stipulating social roles, knowledge would not exist. Although thought provoking and a valuable contrast to the testimony literature, I think Craig's conceptual genealogy and Kusch's use of it are both problematic.

Craig's conceptual genealogy of knowledge is problematic for reasons similar to Goldman's approach to testimony's motivation problem. In both cases, the authors explain a social phenomenon in terms of one type of driving force. For Goldman, the motivation to

communicate is reducible to evolutionary selection pressures; while for Craig, the origin of knowledge is reducible to the function of marking good informants. It does not help Goldman that he acknowledges the need for multiple selection pressures because the underlying presupposition—common to Craig—is that a social phenomenon is reducible to the instrumental benefits derived from it. Craig’s reduction is less objectionable given that it is concerned with a single concept as opposed to all forms of epistemically significant social interaction. I think Craig makes a convincing case that marking good informants played a formative role in the genesis of the concept of knowledge. But what is objectionable is claiming that this is the sole original purpose behind the concept of knowledge.³³ The plausibility this type of one-factor explanation derives in large part from an exclusive focus on instrumental benefits.

Kusch’s use of Craig raises a different issue. More specifically at issue is Kusch’s tripartite distinction between natural, artificial, and social kinds, which Kusch uses to situate Craig’s concept of knowledge. The distinction between artificial and social kinds betrays an overinflated role for language in social epistemology. Kusch thinks that social kinds, unlike artificial ones, depend on the medium of language for their existence. Social kinds are by implication ethereal, separable from material reality and only visible to language users. The tenuousness of this distinction is evident from Kusch’s own examples: a typewriter is thought to be an artificial but not social kind, since a particular hulk of metal that could be used as a typewriter would still exist even if all language users disappeared. But without language users the hulk of metal would cease to be a typewriter for the same reason that the metal discs that were coins would cease to be money: the typewriter would become just a hulk of metal. The significance of language is not distinct from the material reality that it is a part of, which is why

³³ Kusch (2013) alludes to this issue when he suggests that it would have also been important for testifiers to be able to mark good listeners/recipients.

it is misleading to distinguish between artificial and social kinds. John Sutton's (2010) discussion of the encultured mind goes so far as to suggest that there's not even a clear-cut distinction between natural and social kinds. But before challenging this latter distinction, I will use Turner's weak empathy model in Ch. 2 to challenge the view of language operant within the artificial-social distinction.

Despite the two problematic issues, I am largely sympathetic to Kusch's communitarian epistemology. Furthermore, as I will show in the following chapters, there are important affinities with the translation and transformation views. Most importantly, Kusch's emphasis on collective stipulation anticipates their pragmatic-orientation. It is this emphasis on the practical dimensions of knowledge that leads to more careful treatments of the interactive dynamics between individuals. Social interaction is more than a means for transmitting or receiving preformed packets of information.

In this chapter, I have traced some of the major themes and trajectory of the testimony literature. The literature has made it forcefully clear that the social dimensions of knowledge are more than a theoretical afterthought. Yet in describing testimony's "transmission view," I have argued that the testimony view has not fully overcome the myopic individualism of traditional epistemology. By introducing Turner's weak empathy model as a representative of what I call the "translation view," I challenge the presupposition that epistemically significant social interaction must be founded on the exchange of discrete representations or beliefs. In Ch. 5, as part of the transformation view, I challenge the exclusive focus on the instrumental benefits of epistemic interaction—that is, I challenge manipulationism.

Chapter 2: The Translation View

2.0 Introduction

Acknowledging the reality of tacit knowledge raises an immediate concern for the testimony view. If, as Polanyi (1966b) states, “we can know more than we can tell,” then, according to testimony’s transmission view, it appears to be impossible to pass along such knowledge to another person. But this by itself does not pose a problem for the testimony view, since the view need not imply that all of an individual’s knowledge can be passed along. Accordingly, there are conceptions of tacit knowledge that are compatible with testimony’s transmission picture of social epistemology; in the next section, I discuss one such account by Searle.³⁴

Yet other theorists, including Polanyi, think that the implications of tacit knowledge require radically transforming our understanding of explicit knowledge. Tacit knowledge, on this general approach, broadens what is considered epistemically significant. Embodied skills, in particular, take on a greater prominence: Polanyi (1958), for example, examines the epistemic importance of bicycle riding and swimming; in addition to bicycle riding, Dreyfus (1986, 1991, 2005) examines piloting an airplane; Turner (2011) examines driving a car. For all three thinkers, examining the epistemic nature of embodied activities shows how understanding (as explicit knowledge) shades into various skills, skills that are inexplicable in terms of conscious rules or

³⁴ For another account of tacit knowledge that is compatible with the testimony view, see Stanley and Williamson (2001), who acknowledge tacit knowledge but drain it of epistemic significance. Additionally, Gascoigne and Thornton (2013) use this account to argue that McDowellian normativist theories, which I discuss in Ch. 3, are compatible with the tacit.

propositions. And it is this emphasis on embodied skills that ultimately conflicts with the testimony view.

In the present chapter, I draw upon the tacit knowledge literature as a corrective to the testimony view. Analogous to how Schmitt (1994) saw the concept of testimony as a litmus test that establishes the need for socializing epistemology, tacit knowledge shows the need for more detailed analyses of how epistemic agents interact. In terms of the previously discussed cluster of presuppositions and symptomatic tendencies of the testimony view from Ch. 1, tacit knowledge challenges the following:

- (1) the abstract focus on the question of justification,
- (2) basing epistemic interaction on the exchange of discrete tokens,
- (3) viewing epistemic agents primarily as passive informational receivers, and
- (4) dividing the epistemic and practical dimensions of problem solving.

The concept of tacit knowledge shows how these four issues are closely interrelated. Towards this end, in contrast to approaches that are compatible with testimony's transmission view, my understanding of tacit knowledge is heavily indebted to Polanyi and Turner. Turner's (2012a, 2012b) account, in particular, serves as the representative of the translation view, which I offer as an alternative to testimony's model of epistemic interaction.

The present chapter consists of three major parts. In the first (§2.1-2.2), I survey the tacit knowledge literature, beginning with Polanyi's project followed by the two more recent accounts of Dreyfus and Searle. In §2.3-2.4, I examine how Turner uses the issues raised by tacit knowledge to present his weak empathy model of social interaction and criticize what he refers to as "collectivist" social theories. Finally, in the third part (§2.5), I compare Turner's critique of collectivist theories to Kusch's critique of the "individualist testimony view." At first glance, Turner and Kusch's critiques appear to be diametrically opposed to each other, since the latter sees individualism as the primary obstacle to improving social epistemology while the former

sees collectivist notions as the primary issue. I argue that these should instead be understood as complementary critiques that offer two perspectives on one deeply troubled picture. In short, I argue that the individualism of the testimony view creates spurious rifts between epistemic agents that then must be sutured over by vacuous collectivist notions. This discussion sets the stage for the enactivist concept of participatory sense-making in Ch. 3.

2.1 Polanyi on Tacit Knowledge: Imitative Learning within Traditions

Tacit knowledge, as a term of art, originated with the ambitious efforts of Michael Polanyi. First and foremost, Polanyi (1958, 1959, 1964) sought to challenge the positivist picture of science, in which knowledge is detached and impersonal. Polanyi believed this picture of science obscured the psychological processes that underpin knowledge. Most importantly, the basic fact of the tacit—that “we can know more than we can tell” (Polanyi 1966b, 4)—appears paradoxical. Polanyi argues that this basic fact shapes all of scientific practice, beginning with becoming a scientist in the first place. The ramifications of this fact ultimately reshape all forms of knowledge, as indicated by Polanyi’s (1966a, 7) bold claim that “all knowledge is *either tacit or rooted in tacit knowledge*.”

In the present section, I examine the transformative implications of Polanyi’s project. My focus is on the implications that Polanyi’s concept of tacit knowledge has for social epistemology, implications that he only partially develops himself. In the next section, I argue that Dreyfus and Searle both undermine its potential in two closely related though different ways. Because it is the most direct point of contact with social epistemology—the point at which Polanyi most carefully examines how the tacit shapes social interaction—I begin my exposition of Polanyi with his notion of tradition.

The practice of science depends on and is maintained by the tradition of science (Polanyi 1964). Polanyi (1964, 44) famously compared the relationship between great scientists to apostolic succession within the Christian church in which new bishops are ordained by a “laying on of hands.” The comparison emphasizes the need for sustained interaction with distinguished scientists in order to become a distinguished scientist oneself. In modern physics, for example, the Nobel laureate J.J. Thomson mentored Ernest Rutherford who in turn mentored four additional Nobel laureates. The string of tremendous successes—a recurrent pattern in other disciplines as well—suggests that something more than the sharing of explicit observations and theories was responsible for the lineage of great thinkers, since the observations and theories were readily available to researchers elsewhere. In short, the example suggests that learning through imitation is essential to sustaining the tradition of science as well as traditions more generally.

Imitation indicates in what sense tacit knowledge can be taught and deliberately acquired. Turner (2012a, 2012b, 2014) seizes on this aspect of Polanyi’s account, using imitative learning as the framework for modeling the cognitive basis of social interaction. As I discuss in §2.3, this is in tension though not necessarily at odds with Polanyi’s broader emphasis on tradition. For both Turner and Polanyi, imitation enables novices to acquire an indefinite range of skills; in the case of science, some of what a master “imparts” to his or her apprentices include “the way he chooses problems, selects a technique, reacts to new clues and to unforeseen difficulties, discusses other scientists’ work...” (Polanyi 1964, 43). For Polanyi, unlike Turner, acquiring the indefinite range of skills amounts to becoming part of the tradition of science.

In keeping with his rhetorical style, Polanyi’s use of the term *tradition* is somewhat nebulous but does have some clear implications. The term underpins Polanyi’s response to one

possible misreading of his picture of the scientist, which he describes as follows: “the scientist may appear as a mere truth-finding machine steered by intuitive sensibility” (Polanyi 1964, 38).³⁵ Polanyi thinks that intuitive sensibility alone would be insufficient for handling the responsibilities of the practicing scientist. In opposition to this picture, Polanyi describes the scientist as a “detective, policeman, judge, and jury all rolled into one” (38). In terms of this metaphor, the mistaken picture implies that the scientist is only a detective, who collects evidence and tests his or her hunches on nature who is the ultimate judge of the hunches’ truth. But since nature does not play such a determinate role, the freedoms and thereby responsibilities of the scientist are much greater than that of a mere detective. The tradition of science is essential as a resource for guiding individuals, fostering moral ideals including the ideal of objectivity.

In challenging the detached and impersonal picture of knowledge, Polanyi is at pains to offer an alternative account of objectivity. In multiple places, he defines objectivity in terms of a scientist possessing “universal intent” (1958, 341; 1959, 27; 1964, 73; 1966b, 69). Objectivity is thus a product of an individual scientist’s attempt to discover a truth of nature, despite humans’ inherent “infirmity” (1958, 342). Our inherently limited physical constitution ensures that science is an endless inquiry, with further discoveries always awaiting. Tradition is important in this regard, since it consists of the range of values and ideals that new generations of scientists must commit themselves to in order to further the growth of science. These values hold the project of science together—as much as it needs to be held together—not nature as a materially determinate judge of truth.

³⁵ The misreading is notable in part because it is essentially a blend of Alvin Goldman’s view of the scientist with Hubert Dreyfus’s concept of expertise. As noted in §1.6, Goldman views epistemic subjects as essentially nothing more than truth acquirers. While Dreyfus, as I discuss below, focuses on intuitive sensibility to the point of only giving extremely cursory attention to anything resembling a tradition.

The intentional-based account of objectivity serves a dual purpose. On the one hand, it accentuates the abiding importance of making new discoveries in science. Rather than identifying objectivity with substantive facts, the ceaseless striving for more knowledge is the defining feature. The other purpose is more contentious: for Polanyi, humans' physical infirmities illustrate the need for religion: "[we undertake the hopeless task of attaining the universal] because we hope to be visited by powers for which we cannot account in terms of our specifiable capacities. This hope is a clue to God..." (1958, 342). Quite literally, Polanyi (1964, 15 & 45) thinks that being a scientist requires *faith* in its methods, mentors, and in the tradition as a whole; this view is stressed in repeated appeals to the Christian dictum "*fides quaerens intellectum*."³⁶ Given the associations, it should be no surprise that many subsequent theorists have downplayed if not completely jettisoned the importance of tradition in regards to tacit knowledge. This approach, best exemplified by Turner (as discussed in §2.3), offers a fair even if truncated reading of Polanyi, since Polanyi thought empirical considerations motivated the view. The strongest empirical motivation relates to a common regress argument concerning explicit knowledge, which other theorists have drawn different lessons from.

The regress argument is simple. If it existed on its own, explicit knowledge—beginning with simple descriptive statements—would lead to an infinite regress of explicit rules. Polanyi's argument is partly indebted to Gilbert Ryle's (1949) distinction between "knowing that" and "knowing how." Explicit knowledge, which corresponds to "knowing that," implies skillful know how as its precondition. At the most basic level, a cognitive subject must supplement the inherent imprecision of language in order to recognize a statement (whether oral or written) as possessing a meaning. Hence, teaching in general presupposes the "pupil's intelligent co-operation" (Polanyi 1966b, 5). Accordingly, the claim that all knowledge is "either tacit or

³⁶ Polanyi translates the Latin as "to believe in order to know" and "faith in search of understanding."

rooted in the tacit” can be paraphrased as all knowledge either consists of know how or is rooted in know how. Unlike Ryle who limits his argument to a conceptual analysis of terms such as “mind” and “knowledge,” Polanyi relates his version of the regress argument to the biological nature of humans and numerous psychology experiments.³⁷

The biological underpinning of the cognitive subject that Polanyi (1958) identifies resembles that found in Goldman’s epistemics and even Andy Clark’s more recent Extended Mind hypothesis. In all three cases, individual organisms possess a suite of innate drives and cognitive capacities. Goldman and especially Clark think that natural selection favors organisms capable of augmenting their native capacities with acquired skills and tools. While Polanyi (1958, 35) criticizes attempts to explain the origin of personhood and other complex organisms by appeal to natural selection, he still endorses a similar position regarding the immutability of innate characteristics. He writes, “Our thoughts are limited by our innate capabilities. Our senses and emotions can be enhanced by education, but these ramifications remain dependent on their native roots” (339). In this generic picture, the link between innate and acquired capacities is one-directional; innate capacities are the immutable foundation for and executor of acquired capacities.³⁸ It is in this sense that acquired traits are thought to be, strictly speaking, *extensions* of innate capacities.³⁹ Polanyi’s general discussion of biology is geared towards clarifying the relationship between innate capacities and tacit knowledge.

³⁷ N. R. Hanson’s (1958, 1967) conceptual analysis of scientific discovery bridges the linguistic focus of Ryle with Polanyi’s abiding concern with scientific discovery. Polanyi (1964, 12) notes that Hanson’s views concerning the theory-ladenness of scientific facts as overlapping his own views.

³⁸ My own critique of adaptationism, which I detail in Ch. 5, focuses on the issue of whether innate capacities are immutable. I argue that innate capacities are, in fact, plastic and thus are shaped by developmental factors, such as socialization.

³⁹ While Polanyi endorses an immutable cognitive core, he is not committed to Clark’s functionalist framework and is opposed to Goldman’s emphasis on belief-forming mechanisms.

Polanyi (1958) thinks sensory organs are an organism's most important innate cognitive capacity and are inseparable from the organism's innate drives. Perception prefigures all knowledge, while satisfying biological drives prefigures all practical skills (104). Perception and drives thus form a reciprocal relationship: perception guides drives, while the satisfaction of drives, in turn, implicitly confirms the veracity of one's percepts (104). Within the 1958 book, Polanyi thinks of tacit knowledge as interiorizing bits of the universe and offers this biological sketch as an attempt to shed light on the process of interiorization. But by (1966b), though he retains the biological sketch, Polanyi discards the notion of interiorization in favor of viewing the individual as expanding outward towards reality (35).⁴⁰ The defining feature of this expansion is an exponential increase of tacit knowledge, which corresponds to stratifying reality into two cognitive levels. Polanyi describes these two levels in terms of a from-to structure of tacit knowing.

In "The Logic of Tacit Inference" (1966a) and *The Tacit Dimension* (1966b), Polanyi distills tacit knowledge into a from-to structure. This structure is exhibited in various psychology experiments, including one in which human subjects were administered electric shocks after hearing certain nonsense syllables. Subjects were unable to recall the syllables that preceded the shocks yet knew when to brace themselves (Eriksen and Kueth 1956). Polanyi describes the subjects as attending *from* the subsidiary particulars (i.e., nonsense syllables) *to* the focal term (i.e., the electric shock). The focal term corresponds to the meaningful object of conscious attention, while subsidiary particulars are perceived unconsciously only in relation to the focal term. The general trajectory of cognition is for the focal term to become increasingly distal, as acquired tools and skills enable individuals to integrate increasingly large domains of particulars

⁴⁰ Dreyfus's (1972) criticism of Polanyi, as I discuss in §2.2, doesn't acknowledge that Polanyi moves away from the (1958) interiorization picture.

(Polanyi 1966b, 29). Yet in contrast to the case of acquired tools and skills, which even after mastery can be made explicit, the subsidiary particulars in the shock experiment are entirely tacit—known only via attending to the shock. The hands-on education of a new scientist, while perhaps less painful, exhibits the same from-to structure found in the electric shock experiment and the related outward trajectory of cognitive integration.

While attending to the stated goals of a given research project, the new scientist must also “pick up on” the expert scientist’s intuitive skills. These intuitive skills are subsidiary particulars from which the apprentice attempts to contribute to the project’s stated aims, which are the focal terms. If the apprentice focused on learning these intuitive skills, then he or she would not understand their proper significance and also would be unable to focus on the explicit objectives. Hence, *pace* the testimony view, social interaction is all the more important because of what we *cannot* tell, rather than being just a means for transmitting what we can tell. Furthermore, by showing how a subsidiary particular can be inarticulable yet still, in a determinate manner, contribute to a conscious cognitive achievement, the electric shock experiment suggests that emphasizing the role of imitative learning does not necessarily require the broad and perhaps mystifying notion of tradition.

There is, however, an important difference between the electric shock experiment and the education of a scientist. The subsidiary particulars in the latter consist of an indefinite range of various skills and perhaps values rather than a discrete entity, as in the case of a particular nonsense syllable. Hence, in the electric shock experiment, presuming the same nonsense syllables are used for each subject, everyone will be unconsciously aware of the same tacit particular. But in the case of science, there is only a family resemblance between the intuitions of teacher and student, each possessing a somewhat similar though unique skill set.

An individual scientist's unique set of intuitive skills is an example of what Polanyi refers to as a tacit or personal coefficient. Coefficients are personal precisely because they differ amongst individuals, a result of each individual's unique physical constitution and experiential history. Yet these inherent differences are not at odds with the goal of objectivity: Polanyi (1958, 17) describes the personal coefficient as bridging the disjunction between subjectivity and objectivity. The tacit coefficient thus expresses the notion of compensating for personal differences. Individuals "transcend" personal differences by striving to meet universal standards, which relates back to the definition of objectivity as universal intent. Objectivity is defined in terms of intention because it is impossible to compensate for all personal differences.⁴¹

Imitative learning suggests the extent to which compensation for personal difference is possible. In striving to imitate another person, it is common to discover salient differences. Sometimes these differences can be overcome, sometimes not, but many differences correspond to actual differences in how a competency is realized. Competencies are thus multiply realizable functional skills in which personal coefficients refer to differences in how a competency is realized.

In sum, imitative learning is a heavily interactive process that imparts tacit knowledge, but the tacit knowledge is realized uniquely on an individual-by-individual basis. There is thus no determinate body of tacit knowledge that corresponds to science or any other epistemic practice but rather a loose-knit, vaguely bounded range of skills (and perhaps values) that individual scientists have become proficient at. That is to say, there is only a family resemblance between the tacit knowledge of different individual scientists. For Polanyi, this inherent diversity indicates the importance of traditions, which consist in part of moral values that he thinks are

⁴¹ I return to Polanyi's concept of objectivity in §2.3 in order to contrast it with Turner's own.

necessary for maintaining a body of tacit knowledge. In the case of science, relevant values include *faith* (in the history and methods of science), *objectivity* (the intention to meet universal standards despite one's inherent physical infirmities), *honesty*, and so on.

In the following sections, I consider three theorists who are all in some respect dissatisfied with this general picture. Dreyfus and Searle both deny the epistemic significance of diversity, which enables them to each give a more ontologically determinate account of tacit knowledge. More specifically, Dreyfus and Searle, though in importantly different ways, treat tacit knowledge as a form of objective knowledge, which can be transferred between subjects. By contrast, Turner, whom I discuss in §2.3-2.4, retains Polanyi's emphasis on diversity while challenging Polanyi's notion of tradition and the notion of transferrable objective tacit knowledge.

2.2 Dreyfus and Searle: Tacit Knowledge as a Transferrable Objective Precondition

Hubert Dreyfus and John Searle make odd bedfellows. The former's primary interest is in clarifying the nature of performative skills and arguing that such skills cannot be reduced to rule following. Searle, by contrast, models mental consciousness on the rules of language but ultimately supplements this picture with his Background hypothesis, which grounds the mental in pre-intentional *know how*. Despite arriving at the same issue via different routes, both Dreyfus and Searle's treatment of non-conceptual or pre-conceptual *know how* serve as points of contact and comparison to Polanyi's concept of tacit knowledge. In the ensuing section, I argue that despite their differences, Dreyfus and Searle offer similar pictures of the social significance of tacit knowledge. More specifically, Dreyfus and Searle both minimize the differences in tacit knowledge between individuals, which, in turn, dilutes the importance of social interaction. Since Dreyfus's project shares a greater affinity with Polanyi's, I begin there.

Dreyfus's focus on performative skills stems in large part from a reaction against mid-twentieth century artificial intelligence. Good Old-Fashioned Artificial Intelligence (GOFAI), as Haugeland (1985) would call it, sought to create a human-like intelligence based strictly on the computation of explicit instructions. In examining how people perform various bodily skills, Dreyfus argues that *know how* is an essential aspect of human intelligence, an intelligence that cannot be reduced to rule following. With a few rhetorical broad strokes, Dreyfus traces the prejudice against know how through the canon of Western philosophy back to Plato. But, more pointedly, the symptom even appears in Polanyi.

Dreyfus (1972) scrutinizes Polanyi's (1958) bicycle riding example. In an extensive footnote (236-37), Dreyfus seizes on Polanyi's appeal to "hidden rules" and "principles." In the example, Polanyi cites a principle from physics to explain how cyclists maintain their balance. Given the principle's highly technical details, it is safe to presume that most cyclists do not explicitly know and will never explicitly know of it, which is Polanyi's point. Emphasizing the technical details suggests tacit knowledge's complexity and ubiquity. For Dreyfus, however, the appeal to hidden rules signals a conflation between explanation and understanding, in which explanation refers to capturing the essential first-personal features of a performance and understanding refers to a third-personal formal description. While the physical principle is helpful for understanding bicycle riding, Dreyfus thinks that it is incongruous with explaining a cyclist's performance.

Dreyfus's point is somewhat unfair but highlights a limitation of Polanyi's framework. Polanyi, like Dreyfus, thinks that explicit rule-following is inadequate for modeling cognition and has the conceptual resources to support the point. Polanyi's direct response to Dreyfus's criticism is to clarify the ontological status of the hidden rules: as subsidiary particulars, the rules

are not merely unconscious but also functionally bear on explicit knowledge. This corresponds to the previously discussed from-to structure of consciousness. The from-to structure signifies that humans, unlike computing machines, have levels of consciousness. I think Dreyfus's dismissal is unfair, since Polanyi's position is not reducible to GOFAI rule-following, but the criticism does draw attention to Polanyi's need for novel epistemic terms. Polanyi's talk of "rules," "principles," and "premises" stretches their normal meanings beyond recognition. Tacit knowledge deserves and ultimately requires a more original epistemic framework.⁴² Dreyfus develops one such framework, in which rules are only a propaedeutic to a different kind of knowledge.

Over the course of two decades and in collaboration with his brother Stuart, Hubert Dreyfus (1965, 1972, 1980, 1986) distilled his interest in performative skills into a universal model of expertise. Apart from some minor extensions and novel applications, Dreyfus's (1986) five-stage model of expertise represents the mature framework. In the model, rules are initially important, especially in the first stage in which a novice's actions are dictated by rule following. But the significance of rules subsides and eventually is transcended, the crucial moment coming after the proficiency stage. At this moment, the person's skill no longer relies on calculative rationality but strictly on intuition and thereby can no longer be explained in terms of rule-following. Dreyfus thinks that the role of intuition is evident from the felt experience of mastering a skill.

The five-stage model is meant to reveal the underlying phenomenological structure of skills. Dreyfus (1972, 145) characterizes his use of phenomenology as follows: "It [phenomenological description] can give us understanding if it is able to find the general characteristics of such behavior. ... Such an account can even be called an explanation if it goes

⁴² J. J. Gibson (1986, 22) recognizes the far-reaching consequences of Polanyi's concept, labeling environmental affordances as tacit knowledge.

further and tries to find the fundamental features of human activity which serve as the necessary and sufficient conditions for all forms of behavior.” At the very least, phenomenology is understood to reveal the general characteristics of expert performance—namely, acting on intuition rather than rules. But the more ambitious claim, the one Dreyfus commits to, is that the five-stage model reveals the fundamental features of all human skills. Expertise is thus closely related to intelligent bodily coping, which is inherent to any bodily performance.

Intelligent bodily coping is non-conceptual or pre-conceptual, yet on this view is still an objective form of knowledge. Dreyfus’s (2001b) introduction to Samuel Todes’s *Body and World* is helpful in this regard. Todes’s project was to identify the perceptual categories of judgment, which are distinct from and the precondition for Kant’s conceptual categories (Todes 219). Subjects are material beings with needs; the success conditions of perceptual judgments consist in the satisfaction of those bodily needs (Todes 58). The satisfaction of bodily needs provide an objective standard apart from any conceptual judgments, thus relating to what Dreyfus labels “practical objective knowledge” (xxvi). While not invested in the specifics of Todes’s table of categories, Dreyfus heartily endorses the notion of practical objective knowledge, since it articulates a material basis for his phenomenological account of expertise.

Given his unusual use of the term “expertise,” the most charitable way of reading it is as a synonym for intelligent bodily coping. Accordingly, Dreyfus counts many activities as forms of expertise—such as bicycle riding, walking, and talking—that are not normally counted as such, though they are instances of bodily coping. Furthermore, he discounts some activities that are normally thought of as forms of expertise—most importantly, coaching and teaching—that he thinks are inexplicable in terms of bodily coping. Evan Selinger and Robert Crease (2006) link Dreyfus’s usage to his defense of expertise against non-expert demands. Dreyfus (1986, 196)

defends experts against what he sees as an unwarranted and destructive pressure to rationalize their skills. A non-expert's demand for justification, the argument goes, forces the expert to distill her skills into explicable rules, essentially forcing her to regress back to the stage of proficiency. In contrast to this rationalizing pressure, Dreyfus thinks that expertise should be justified in terms of its form.

Dreyfus's defense of experts involves simultaneously stressing their similarities and differences with non-experts. Selinger and Crease (2006, 231-32) express this argument in terms of a Similarity Claim (SC) and Difference Claim (DC). As "experts" in walking, talking, and such, every person can recognize the phenomenal structure of mastering a skill. This is the sense in which all forms of expertise are similar. But if a person has not mastered a particular skill—such as being a doctor, a ballistics expert, and so on—the outsider cannot properly evaluate the skills of *that* expert; only a fellow expert (i.e., an insider) can make such an evaluation. This is the sense in which there are barriers between different kinds of expertise. In order for expertise to be justified in terms of its form, Selinger and Crease note that SC must trump DC; that is, non-experts should accept an expert's judgment based on the shared phenomenal structure of mastery in spite of the barrier. Selinger and Crease's point is that this account is far too socially impoverished to explain how expertise is actually acknowledged with respect to particular individuals, especially in situations where expertise is contested.

There are numerous unacceptable social consequences of Dreyfus's insulation of experts. Most immediate is the naïve presumption that experts have shed all of their personal prejudices with regard to their skills.⁴³ Another consequence is diminishing the importance of coaching and

⁴³ Dreyfus could concede that this is false while maintaining that risking the influence of experts' personal prejudices is a worthwhile tradeoff given the alternative of undermining expertise. But this response would be inconsistent with the notion of *objective* practical knowledge that informs Dreyfus's handling of social questions (that is, his almost complete neglect of them).

teaching. Selinger and Crease (2006, 226) note that while coaching is often understood as a form of expertise, Dreyfus characterizes it as a “dispensing of rules in a standardized way.” This picture of coaching is consistent with the five-stage model—in which all forms of expertise begin with rule following—but is at odds with the complex reality of coaching. Dreyfus (2001a, 171-72) himself hints at this issue in discussing the importance of apprenticeship for “passing along a style.” But his cursory discussion pales in comparison to Polanyi’s original treatment.

Imitative learning, as discussed in terms of the master-apprentice relationship, exemplifies the differences between Polanyi and Dreyfus. “Passing along a style,” for Dreyfus, consists essentially in cloning the master. In the (2001a) passage, Dreyfus notes the danger of becoming merely a virtuoso, a musician who has copied her master’s style. The only way to escape this danger, Dreyfus suggests, is to study with multiple masters, which thereby creates a clashing of each master’s styles that “destabilizes and confuses” the apprentice (172). In other words, the default result of imitative learning is identity between master and student. Dreyfus and Polanyi thus reach opposite conclusions about whether diversity or identity come cheaply through imitative learning: for Dreyfus, it takes extra effort to ensure there are qualitative differences between master and apprentice, while for Polanyi, the moral dimension of traditions are necessary merely to foster a family resemblance between master and apprentice. This result is surprising given that Polanyi was charged with clinging to a rule-based account, but it is Dreyfus’s model that diminishes the diversity across individuals. The notion of practical objective knowledge implies that expert intuitions are so determinate that they are effectively rules, only rules by a different name.

Having diluted the differences between people, social interaction appears, for Dreyfus, as at best an opportunity for copying a piece of objective practical knowledge. But more often

social interaction is expensive if not an outright imposition, as in the case of expecting experts to articulate and thereby justify their judgments. Hence, while identifying know how with a distinctive form of objectivity prevents it from being reduced to rationality, it eliminates what was socially distinctive about tacit knowledge. In terms of the implications that know how has for social interaction, John Searle reaches what is effectively the same position but from a very different direction.

Unlike Dreyfus, Searle (1990, 1995) directly addresses and gives extensive attention to the question of the social. The question follows from Searle's (1969, 1983) larger project of using the methods of the natural sciences to explain ordinary human behavior. From this very different starting point, Searle adduces some additional considerations not seen in Dreyfus, such as a sketch of the evolutionary origins of know how and a proposal for a new type of explanation in the social sciences. As a project in fundamental ontology, Searle articulates some important assumptions underlying Dreyfus's account that also relate to the testimony view.

There are two scientific theories that any educated citizen of the twentieth and twenty-first centuries must accept, at least according to Searle (1995, 6). The two theories in question are the evolutionary theory of biology and the atomic theory of matter. From this vantage point, the question of the social appears primarily as an ontological one. That is, the complexity of social life outstrips bare physical descriptions, so it is unclear how they fit within our fundamental understanding of the universe. Searle (1969) first touched on this issue by distinguishing between brute and institutional facts but felt he could not satisfactorily account for it until *The Construction of Social Reality* (1995).

The difference between brute and institutional facts hinges on constitutive rules. Constitutive rules create the possibility for certain types of action, such as rules that establish the

existence of money or how a sport is played. As such, constitutive rules are the defining feature of all social institutions. With regard to sports, taking baseball as an example, the rules concerning what counts as an out, what a safe hit, what a run, and so on are constitutive in that the game of baseball would not exist without them. Regulative rules, by contrast, such as what type of clothing the players must wear, whether umpire calls can be appealed, and so on can be changed without directly affecting the game itself. Searle (1995) links the notion of constitutive rules to three additional elements that were not present in his early work on speech acts. Two of the elements are the concept of assigning functions and that of collective intentionality, both of which constitutive rules presuppose. The assignment of functions—in which humans and some other animals impose a function on a physical object—illustrates the point nicely. Whether referring to a natural or invented object, Searle argues that all functions are observer relative, presupposing an observer's prior set of values that are reflected in the assigned function.

Searle presents his account as an articulation of people's ordinary notion of functions. The common cold, for example, is not supposed to have a function because it would conflict with the values of observers. In this regard, Searle dismisses the relevance of Ruth Millikan's (1984, 1989) account of proper function on the grounds that hers is a technical concept that conflicts with our ordinary notion and depends on the truth of the causal historical narrative suggested by Darwinian evolution. This is a curious position given that it is the same chapter in which Searle says that every educated person today must accept evolutionary theory and, worse yet, Searle even helps himself to Darwinian evolution on the very same point, using evolution to dismiss teleological conceptions of function. With respect to the issue of whether common colds have functions, Searle's point is unhelpfully circular: colds do not have functions if they are observer relative. But without this stipulation, Millikan's account has an answer readily available: colds

do not, in fact, have a function but rather are an incidental side-effect of the proper functioning of the viruses that cause them in their hosts. Although sneezing—since it helps spread the viruses to new hosts—would arguably count as part of the viruses’ proper functioning.⁴⁴ In short, I do not think that Millikan’s account of proper function and Searle’s characterization of our ordinary functions—if there is even anything substantive to be said about such a thing—are mutually exclusive. But it is important for Searle that they are mutually exclusive, so that he can claim fundamental priority for the latter. That Searle risks such a tenuous line of reasoning indicates its importance to his larger project.

The observer relative model of functions establishes the basic intuition of an ontological hierarchy. If *all* functions trace back to an observer’s imposition on a physical object, then functions presuppose not only an observer but, more fundamentally, the material reality that the observer assigns the function to. This ultimately underpins the claim that social reality is derivative upon material reality. Social reality also presupposes collective intentionality, the other element noted above, but in a less obvious way.

Collective intentionality lacks the patina of concreteness found in the assignment of functions. Searle’s (1990) argument for we-intentions appeals to two pieces of evidence. First, the intuition that group undertakings—such as an orchestra or football team—are more than the aggregate coincidence of multiple individuals’ performances. The second piece of evidence concerns the analysis of the statement form “We intend *x*.” This statement, Searle argues, cannot be substituted for any set of “I intend *x*, and you believe...” statements. From these two pieces of evidence, despite them not referring to anything directly related to biology, Searle infers that we-intentions are a biological primitive. As a result, the biological character of we-intentions

⁴⁴ While I am not endorsing Millikan’s account of proper functions, I think my sketch is helpful for problematizing the intuition behind the observer relative account of functions. I give a more detailed analysis of Millikan’s account in Ch. 5 with regard to functionalist theories of mind.

appears to be somewhat mysterious, we-intentions acting as an explanatory stopping point about which nothing more can be said.⁴⁵ Instead, the explanatory task of we-intentions, in conjunction with assignment functions, is to demystify the ontological status of constitutive rules and the social institutions of which they are a part. But completing this account requires an additional fourth element, which permeates the other three.

The Background hypothesis, the aforementioned fourth element, is the chief innovation of Searle's account. Searle (1983) proposes the Background hypothesis as part of a comprehensive account of intentionality. Intentional states, which are any mental state possessing representational content, presuppose know how. Like Dreyfus, the necessary role of know how is established by an infinite regress argument, which, in turn, relates back to Polanyi and Ryle. Searle (1983, 150) draws the same lesson as Dreyfus, namely that rules become progressively irrelevant as a person masters a skill.⁴⁶ Searle's innovation is formulating this picture of know how in terms of a third type of causal explanation, the first two types being brute physical causation and mental causation. This perspective further articulates the problematic noted at the end of my discussion of Dreyfus, which Searle expresses in terms of the following paradox. Explanations of human behavior, especially in relation to social institutions, face a dilemma: explaining the complexity of behavior, such as the actions of a baseball player or user of money, appears to warrant an appeal to rules, but appealing to rules threatens to undercut the creativity and spontaneity of the behavior. One example that Searle uses is how to explain the actions of a professional baseball player who runs to first base after hitting the ball: explaining the action by appealing to rules such as what counts as a safe hit, how to score runs, and so on

⁴⁵ One of the main selling points of participatory sense-making, as presented in Ch. 3, is that much more can be said about the biological character of the dynamics that we-intentions are meant to capture.

⁴⁶ Searle even reiterates Dreyfus's charge that Polanyi's model fails because it implies rules become interiorized. But, as I stated earlier, I believe that Polanyi's picture of a functional relationship between levels of the mind is more akin to Andy Clark's Extended Mind hypothesis.

fails to correspond to the first-personal performance. Yet if these rules did not exist, the person's behavior would be different or at least mean something very different. Handling this tension requires maintaining the difference between rule-governed and rule-described behavior. With regard to the Background hypothesis, the issue manifests itself in terms of representing something that is itself nonrepresentational (namely, the preconceptual Background).

It is to address this challenge that Searle proposes his new type of causal explanation. The explanatory form is based on an analogy with the principle of natural selection, in particular its diachronic character. By introducing this diachronic element, Searle believes he can reconcile the relationship between the Background's preconceptual capacities and social institutions' constitutive rules. The diachronic element consists of two stages: in the first, a person develops dispositions sensitive to a set of rules, which in the second stage culminate in a set of skills (Searle 1995, 142). Constitutive rules are thus a causal factor in forming dispositions within the agent, but the dispositional structure consists of the agent's Background capacities as opposed to the rules themselves. Having acquired appropriate dispositions, the original rules become increasingly irrelevant as the agent's performance becomes increasingly fluent or, to use Dreyfus's favored term, expert. In sum, the first stage entails a causal explanation, as the rules are a causal factor in the agent's acquisition of dispositions; in the second stage, the explanation is functional, since the agent's behavior functions in accord with the rules but, given its fluency, involves more salient features that may be inexplicable in terms of any rules.

Even though Background capacities are essential to social reality, as suggested by the preceding sketch, Searle is at pains to claim that this reality is mental as opposed to social, biological, or purely physical. As with the paradox earlier, Searle (1983) cuts to the heart of the issue, conceding that the Background is "indeed derived from the entire congeries of relations

which each biological-social being has to the world around itself” but that “all this embeddedness is only relevant to the production of the Background because of the effects that it has on me, specifically the effects that it has on my mind-brain” (154). Culture, the particular people one interacts with, and even one’s own body—apart from the brain—are the contingent means for affecting the mind-brain. Hence, even if I, as my mind-brain, happened to be floating in a vat rather than encased in a skull attached to a motile body, Searle claims that our mental state (including its Background capacities) would be identical. While notable for the starkness of its outline, I think Searle’s sketch lurks behind the testimony view’s reduction of social interaction to belief transmission. It is thus best to view the picture not in terms of Searle’s account of the mental but rather from the starting point of his fundamental ontology.

The problem with Searle’s fundamental ontology is that it is *not* fundamental. While his sketch of evolution and the atomic theory of matter hold wide purchase—indeed, I think this rather than any direct connection explains the parallels between Searle’s mind-brain picture and the testimony view—there are issues with both of the two putatively fundamental theories. With respect to evolution, as was the case with Goldman (§1.6), Searle is only concerned with the principle of natural selection. Evolution thus collapses into adaptationism, which in turn makes social interaction appear relevant only in terms of increasing chances of survival. Regarding the second leg of Searle’s fundamental ontology: there are many educated people who, for various reasons, do not believe in the atomic theory of matter if it entails materialism, as Searle’s ontological hierarchy implies. Dissenting views run the gambit from certain interpretations of Quantum Field Theory within physics to Dynamical Systems Theory within biology. Irrespective of the ultimate status of materialism and evolution, Searle’s breezy attitude towards the

fundamental character of naturalism is contemptible, especially since it is a pervasive attitude that has, I think, obfuscated epistemic conceptions of social interaction.

Searle's conception of the social suffers from most of the limitations noted with respect to the testimony view. In terms of the cluster of presuppositions and tendencies of the testimony view noted at the beginning of the chapter, Searle's account exhibits three of the four issues. His focus on fundamental ontology spares him the myopic fixation on the question of justification, the first issue. But, as evidenced by the picture of the mind-brain above, Searle still considers epistemic agents (i.e., their brains) as primarily passive receivers, the third issue. Likewise, the definitive distinction between the mental and the social-biological reiterates the division between the epistemic and practical dimensions of problem solving. Whether Searle's account suffers from the second issue—namely, basing epistemic interaction on the exchange of discrete tokens—is more contentious. Searle and Dreyfus's attention to know how means that epistemic interaction is not reducible to exchanging propositional beliefs. But I think that Dreyfus's portrait of imitative learning and the concept of practical objective knowledge retains the basic notion of individuals forming complete pieces of knowledge and then imparting it to one another. So even though token statements or statement-like forms are not exchanged, there is still a discrete transaction between two definitively circumscribed individual centers. Turner (1999) also argues that Searle's account of social reality is socially impoverished and, furthermore, he alludes to the shortcomings of Searle's fundamental ontology by challenging the distinction between brute and institutional facts.

The distinction between brute and institutional facts relies on what may appear to be a straightforward and compelling observation. Searle contrasts the question of whether there is snow at the top of Mt. Everest to questions such as the value of money. The latter question

relates to a social institution and is thus, Searle thinks, a social fact, while the former is a brute objective fact that will be true or false irrespective of any individual's belief regarding the matter. As noted above, this distinction follows directly from the ontological hierarchy that the assignment of functions is meant to establish. Turner's (1999, 230-31) critical point specifically regards how Searle uses this distinction to attack social constructionist analyses of science. That is, Searle claims that because the objects of natural science concern brute facts, scientific practice is objective in a way that practices concerned with social facts, such as finance, are not. But as Turner notes and even Goldman concedes (§1.5), contemporary science has an extremely attenuated relationship to anything like brute facts, which undermines the putatively distinctive objectivity of scientific practice. In short, Turner's point is that scientific practice is not concerned with the brute facts of medium-sized dry goods. In Chs. 5 – 6, I argue that not even medium-sized dry goods are *brute* enough to qualitatively distinguish them from the social, but Turner has a different axe to grind with respect to Searle.

For Turner, treating scientific practice as exceptionally objective diminishes the significance of the social interaction between scientists. This is in part a manifestation of the mind-brain picture noted above in which Searle appears to suggest that there is as much epistemically significant about social interaction as there is about my left pinky toe, since such matters are only indirectly important via their effects on my mind-brain. It just happens to be the case that other people—presuming I am not suffering the hallucinations of an invaded brain—offer more stimulation than my left pinky toe. Searle recognizes the behaviorist implications of this picture and appeals to collective intentions as the distinctive and irreducible source and characteristic of the social. And it is precisely this notion of we-intentions that Turner wants to axe as part of his project of restoring the importance of interaction between individuals; a view

that Turner thinks follows from the other aspects of Searle's picture and is suggested by Polanyi's concept of tacit knowledge.

2.3 Substitutes for the Normative

One of Stephen Turner's abiding interests in social theory has been in contrasting collectivist and individualist models of social interaction. The tension is at the forefront of the (1994) *The Social Theory of Practices*, the (2002) *Brains/Practices/Relativism*, and the (2010) *Explaining the Normative*, though in the following I focus on some of his more recent discussions of the issue. I will use Turner's critique of Searle as a jumping off point for his (2012a, 2014) characterization of tacit knowledge. Turner appeals to tacit knowledge, especially Polanyi, to argue against what he sees as a dilution of social interaction by collectivist models. Turner finds empirical support for his individualist picture in cognitive science, which I consider in the next section. But for the present section, I focus on Turner's critique of Searle and his attempt to recapture Polanyi's model of noisy, loose-knit imitative learning.

While the concept of tacit knowledge has been influential, Polanyi's model of imitative learning as apostolic succession has gained comparatively little attention. Dreyfus and Searle's work on know how are two such examples of Polanyi's uneven legacy. The case of Searle is noteworthy because he, unlike Dreyfus, gives extensive attention to the question of the social. Given their complexity, social institutions—such as sports teams and money—and their constitutive rules, represent the most difficult cases and are thus the sole focus of Searle's explanatory efforts. Turner (1999) notes that Searle's strategy rests on a questionable presupposition, namely that all forms of social interaction can be modeled on social institutions even though not all and *probably not even most* social interactions (though who's counting) are based on an explicit set of rules, as is the case with institutions like football and money. As noted

at the end of §2.2, one consequence of Searle's emphasis on social institutions when paired with his Background hypothesis is to move in the direction of behaviorism. Searle's response to this danger also indicates why he thinks social institutions are appropriate exemplars for all social interaction.

Collective intentions are, for Searle, the distinctively social element that enables social institutions and are the source of their normative character. As one example of institutions' normative character, Searle (1995, 146) has us imagine a person who would eat a baseball rather than try to hit it: such a person, he thinks, would fail in a fundamentally different way than someone who swings and misses. The eater would fail to be a member of the baseball community; put another way, the eater would have somehow missed, radically misunderstood, or forgotten his acceptance of "We intend *to play baseball*." It is this normative character, Searle claims, that prevents his account from collapsing into behaviorism. Furthermore, if constitutive rules are simply the formalization or crystallization of we-intentions, then social institutions simply make explicit the underlying dynamics of all social interaction.

The concept of we-intentions, Turner (1999) notes, also addresses a problem in social theory dating back to Hobbes. The problem concerns how to account for the origin of social concepts in individual subjects—in the case of Hobbes, how to ground the concept of state. Building such concepts out of asocial components has proved very difficult but, for many, preferable to the notion of a group mind. The concept of we-intentions is meant to dissolve this dilemma by having the "we" anchored in the individual without presupposing a group mind. Yet Turner thinks that Searle's Background hypothesis, *sans* collective intentionality, provides the necessary tools for addressing the problem. This is the central move in Turner's attempt to excise we-intentions and their attendant normativity from Searle's larger account.

One consequence of the Background hypothesis is to reduce what needs to be explained in terms of conscious I-intentions. Turner (1999, 224) writes, “The mechanism of accustoming and the replacement of conscious rules with the causal structures of the Background that produce skilled performances serve to do much of the work that the ‘internalization’ model formerly did.” Given the primary role of preconceptual know how, it is unnecessary to explain complex behavior in terms of an individual’s conscious intention to follow rules; like Searle, Turner understands this to undercut social theories that claim the existence of internalized systems of rules or concepts. But unlike Searle, Turner thinks that a group endeavor, such as a baseball team, does not depend on intrinsic normativity but rather only requires extending the insights suggested by the notion of preconceptual accustoming.

The differences between Turner and Searle become more pronounced the more poorly bounded and ambiguous a group endeavor is. The case of baseball only suggests a very modest difference: for Turner, all that is consciously needed is for players to individually intend and thereby share the goal of winning the game, while for Searle, each endorses a normatively binding “We intend to play baseball.” But with respect to more diffuse social institutions, such as money, Searle’s account becomes more implausible. In contrast to baseball, it is hard to imagine a we-intentional endorsement of money *per se*, yet Searle needs there to be something at least analogous to conscious endorsement in order for the institution of money to be normatively binding. Towards this end, Searle argues that our acceptance of institutions such as money occurs tacitly, occurring via the Background accustoming of oneself to money use. Turner, by contrast, thinks that Searle overlooks a more straightforward and empirically grounded account.

Diffuse social institutions require three basic elements, Turner (1999) argues, none of which involve any primitive collective element. Mentioned above, the first element consists in

consciously shared goals—whether to win a game or, in the case of money, to exchange goods. The key role of the Background hypothesis, for Turner, is to show that minimal conscious intentions such as shared goals are sufficient, since the Background process of accustoming oneself to an activity accounts for the ensuing behavioral complexity. Given a sufficiently high behavioral frequency of an activity—the second required element—the individual actors constitute a loose-knit community, irrespective of the countless differences in belief that each individual has regarding the activity. Finally, the third element substitutes out the intrinsically *normative* we-intentions in favor of first-personal *normativizing* beliefs. That is to say, if an individual has certain beliefs regarding money use, then he or she may censure others who do not act accordingly. Hence, the normative is not primitive but rather is derivative on individual belief and their enactment by individuals in relation to each other. Turner’s crucial point is that his sketch is implied by Searle’s account, only it is not burdened with the inscrutable notion of biological we-intentions that are tacitly consented to (228).

The substitution of normativizing beliefs for the normative establishes one respect in which social interaction is intrinsically significant. While I return in §2.4 to Turner’s use of Searle’s Background hypothesis in relation to cognitive science, I first want to link this emphasis on social interaction to Polanyi’s model of imitative learning. Turner’s (2012a, 2014) characterization of Polanyi sets the stage for the connectionist account of the next section. Turner’s reading of Polanyi is guided by two features, namely Polanyi’s model of imitative learning and his understanding of tacit knowledge as an actual cognitive phenomenon suited for empirical study. These two features have direct implications for social theory.

The problem of the tacit, for Turner (2012a), is explaining in what sense it can be taught or socially transmitted. Turner thus uses tacit knowledge as a litmus test for social theory,

specifically whether social theories have given a sufficiently careful account of the interactive dynamics between individuals. There is a tendency to base social interaction upon a shared ground: whether the ground is an elaborate system of rules or a binding normativity, the implication is that tacit knowledge must be a distinct entity that exists prior to a given interaction in order to enable it. In contrast to this picture, Turner models social interaction on the act of translation; more specifically, Turner envisions agents as improvising explanations that are aimed at compensating for differences *vis-à-vis* a particular audience rather than tapping into a pre-existing entity. Polanyi's model of imitative learning is central to Turner's account because it exemplifies these features.

Turner's focus on imitative learning is in tension with though does not contradict Polanyi's overall project. As noted in §2.1, Polanyi uses the model in part to emphasize the importance of traditions, which in turn was part of his steadfast critique of logical positivism and his argument for religion as possessing a meaningful role in culture. The science-facing dimensions of the model—the dimensions that Turner seizes upon—concern the education of novice scientists and the fundamental significance of discoveries in science. To briefly recap, acquiring a master scientist's tacit knowledge consists of learning to attend to various focal terms (e.g., concepts, theories, experimental methods), which thereby involves the novice becoming sensitive to countless subsidiary particulars (e.g., discerning salient experimental data) some of which the master cannot even articulate (analogous to the nonsense syllables of the electric shock experiment). This from-to structure of imitative learning thus demonstrates the vital importance of mentorship, which occurs against a backdrop of enduring individual differences that are an endogenous source of scientific innovation. Turner's exclusive focus on these aspects of

imitative learning results in a different conception of objectivity, as seen in his reconstruction of the notion of the tacit coefficient.

Both Polanyi and Turner see the tacit coefficient as denoting the uniquely individual grounding of knowledge. But Turner detaches this idea from the notion of objectivity as bare universal intent discussed in §2.1, favoring instead a less common formulation by Polanyi in which objectivity consists of an individual “striving passionately to fulfill...personal obligations to universal standards” (Polanyi 1958, 17). The key aspect of objectivity thus shifts from away from “universal intent” in favor of “universal standards.” Turner (2012a, 389) writes, “It [objectivity] is a matter of acknowledging individual difference, and correcting it to a public standard.” On this reading, objectivity consists of individuals correcting their personal idiosyncrasies against an agreed upon standard rather than a bare intention that does not necessarily refer to an agreed upon standard. This is a meaningful difference in part because there are not always agreed upon public standards for individuals to correct themselves against, such as in periods of scientific crisis. Turner’s definition reflects in part a fundamentally different orientation to the question of objectivity and imitative learning.

While Polanyi gears his account of objectivity and imitative learning towards characterizing science as a social practice and a historical tradition, Turner is interested in the fundamental nature of the cognitive subject. Issues such as the nature of scientific crises are thus of only derivative import. Rather than being a window onto science as a social practice, Turner sees imitative learning as a window onto the underlying cognitive dynamics of each interacting individual. This can be seen in Turner’s own characterizations of embodied performance and imitative learning.

In one of his most extensive descriptions of imitative learning, Turner (2012a) uses car driving as an example. Car driving shares a feature that is exhibited by many other activities: “Different people may perform the ‘same’ acts, such as driving or riding a bicycle, in different ways that are functionally equivalent” (391). Functional equivalence suggests that despite similar end results it is possible for people to differ dramatically in how they perform an act. Turner continues:

To some extent these differences are smoothed out by training, an explicit process of learning that provides feedback. But to a large extent the feedback is “wild” and consists of actual experiences with other drivers and other situations that go beyond the book or the driving school experience, and are individualized, because the individual has different experiences. (391)

The passage is crucial for representing the overall effect that learning and experience more generally has on individuals. Differences between individuals, which are endemic due to inherent dissimilarities in people’s physical constitutions, are to a very limited extent addressed by training the person in a given skill. In addition to the limits of training, further experience cuts back in the direction of diversity, since different individuals have differing driving experiences. A young adult who moves to a large city, for example, would have very different driving experiences and thereby very different driving habits than her parents. Hence, only the most conspicuous differences—those that would prevent the attainment of functional equivalence—are “smoothed out” by training, which is to say that functional equivalence masks the underlying diversity. Turner builds upon this sketch by detailing the dynamics of the training process.

Smoothing out a novice’s performance involves direct engagement. Turner offers the following sketch of teaching someone how to drive:

What do we do when we articulate driving advice, for example to a novice driver? Do we pull something out of a stock of previously unspoken but already tacitly present sayings, which make the tacit explicit? Or do we repair inferences in a way that provides a functional substitute for what we know tacitly for the particular audience in question, a

novice driver doing something that we recognize that we would not have done—waited too long to get into the turn lane, for example—which we can then formulate as practical “knowledge” or a rule which is a functional substitute for the driving and lane changing habits that we follow unconsciously. (392)

In characterizing what happens when advising a novice, Turner recognizes two basic options—either the advice taps into a pre-existing stock of maxims, or it is spontaneously improvised on the spot when confronted with a salient difference in performance. Given the nature of skills, as described by Dreyfus and Searle, in which rules are irrelevant when it comes to expert performance, the choice is clear for Turner: since there are no rules to tap into, the teacher instead improvises a “functional substitute” for how he or she normally performs the action. The functional substitute reflects both how the teacher normally performs an action *and* how the novice’s behavior differs from it. Since the functional substitute is thus not an introspective report, it cannot be identified as a psychological process in either individual but instead is tied to the actual engaged interaction between the particular individuals.⁴⁷ It is for this reason that Turner warns against conflating the *tacit that is made explicit* and the tacit.⁴⁸

Turner thinks his characterization of teaching someone how to drive generalizes to all social activities. The ability to interact with others in particular contexts, not just learning contexts, is itself a form of skilled know how. This is true whether the context is rush hour traffic, biking in a group, or collaborative scientific research. Goldman addresses the latter context with his concept of a product epistemology (§1.5), in which an individual must learn a community’s medium and mode of communication. But the key piece missing from Goldman’s product epistemology, from the perspective of Turner’s model, is the omnipresent role of the tacit. Goldman’s product epistemology suggests that a community’s means of communication is an

⁴⁷ This account substantiates Selinger and Crease’s (2006) claim that Dreyfus’s picture of coaching is superficial.

⁴⁸ Turner’s point relates to De Jaegher and Di Paolo’s (2007) “Rear Window” diagnosis that I discuss in Ch. 3.

independent mechanism that an individual employs as a tool, whereas for Turner, becoming a member of community consists in the more ambiguous notion of becoming fluent enough to “get by” or “make oneself understood.” This extension of the imitative learning model exhibits the essential features of the general model of social interaction suggested by tacit knowledge, which I have labeled the translation view.

The main tenet of the translation view is that diversity amongst individuals is pervasive, persistent, and thoroughly shapes epistemic interaction. At least two general consequences follow from this fact: first, social interaction is intrinsically important, since it generates knowledge rather than instantiating previously existing knowledge and, second, social interaction is an inherently noisy and protracted process. The translation label stems from the ineliminable need for individuals to translate the behavior and words of others into their own personal competencies. Instructions and verbal reports in general—rather than being standardized maxims—are provocative triggers that elicit and facilitate learning new skills and/or accomplishing a joint task. That this type of translation is endemic to epistemic interaction is to some extent masked by the coarse-grained standard of functional equivalence. The translation view is further substantiated in the next section by filling out more of the underlying cognitive dynamics. But this sketch sets the basic contours of what is to be explained by cognitive science and is to some extent independent of the cognitive account.

2.4 The Cognitive Basis of the Translation View: Connectionism and Mirror Neurons

The relationship between the present and preceding section parallels in some respects the relationship between Turner’s (1994) *The Social Theory of Practices* and his (2002) *Brains/Practices/Relativism*. In §2.3, as in *The Social Theory of Practices*, Turner’s main focus is to show that social interaction does not require a shared foundation. The 1994 book

approaches this issue by criticizing the tendency within social theory of reifying social practices, attributing to them causally efficacious properties on the basis of descriptions of individuals' behavior. The 2012a essay substantiates this critique in connection with tacit knowledge, using Polanyi's emphasis on mentorship to detail an alternative to the reification of social practices. In the 2002 book, as in the present section, I examine Turner's efforts to motivate his alternative picture by appealing to cognitive science. While touching on the 2002 book, my focus will be on Turner's engagement with Goldman's nursery example in (2012a) and its connection to the (2012b) discussion of weak empathy.

Turner's model of imitative learning is detachable from his endorsement of connectionism and mirror neurons but only up to an extent. It is possible to explain the contours set by the imitation model using an entirely different cognitive theory—Turner himself alludes to one alternative—but the connectionist and mirror neuron framework derives some of its appeal from recognizing the imitation model as generalizable to all social interaction. If imitation is taken to be the most basic explanatory target, then any alternative cognitive account would be functionally equivalent to Turner's own connectionist-mirror neuron account. In the following section and chapter, I argue that while insightful the imitation model should be part of a broader framework for modeling interaction, despite Turner's pitching his cognitive account at a very low, mostly unconscious level. To examine this issue, it is important to identify this low-level target, which Alvin Goldman (2005) also addresses in attempting to explain the neural basis for interpersonal understanding.

Goldman's work on a simulationist theory of mind is an extension of his epistemics project. Beginning with (Gallese and Goldman 1998), Goldman has linked his simulationist account to the premotor cortex neurons better known as "mirror neurons." For Goldman,

simulationism and mirror neurons together offer a fine-grained neural account of not only interpersonal understanding but also the underpinning of his fitting correspondence theory of truth (§1.6). Goldman's (2005) crying babies example, which Turner (2012a) directly addresses, is helpful for identifying the essential features of his account. Goldman asks, "Do babies who experience upsetness as a consequence of hearing another's cry represent the latter upset? If not, I don't think it's a case of either mindreading or interpersonal understanding" (n.p.).

Representation is thus understood as a precondition for social cognition, specifically cognizing another person's mental state. But when one baby's crying spreads to another, Goldman thinks this is instead a brute non-mental contagion response.

Two additional elements would be necessary for the second baby's response to constitute interpersonal understanding. First, the contagion response would have had to cause a simulation routine in the second baby, which would represent the crying's significance (e.g., the emotional state of upsetness). Second, the listening infant would have to impute the representation to the first baby and evaluate its accuracy given the other's behavior. With respect to an infant, this seems a bit far-fetched for even a simulationist, like Goldman. Within this picture, mirror neurons are meant to elucidate more of the causal underpinning of the simulationist theory of mind, accounting for how brute causal determination—in this example represented by the spread of crying—can be reliably converted into an internal simulation routine, whose output is a mental representation. Epistemic interaction thus reduces to representation matching, which corresponds to the testimony view's generic notion of discrete belief exchange. The need for each individual to test and refine an internally generated representation provides the most charitable way of reading Goldman's "fitting" correspondence theory of truth.

The mirror neuron-simulationist picture allows for some latitude in correspondence relationships. The picture replaces the notion of there being one essential causally determinate correspondence in favor of individually unique yet reliable isomorphisms. That is to say, each individual's mental representation of an "external" state of affairs is unique but, presuming that it is a truthful representation, the representation is causally determined and is thereby reliably produced. The need for discussion and enquiry, which may appear to be of little-to-no importance given the seemingly all-or-nothing implications of causally determinate correspondence, is explained by the fact that individuals must work to converge on and match each other's internally generated representations. While this latitude opens up a theoretical space for more familiar notions of epistemic agency, I think it is still too cramped to adequately account for creativity, as suggested, for example, by Polanyi's characterization of scientific discovery.⁴⁹ While I press this point further in Ch. 3, Turner offers a more direct response to Goldman's representation matching picture.

Turner's alternative to Goldman's representation matching picture is notable in part because it also begins with mirror neurons. Analogous to his argumentative strategy in relation to Searle, Turner seeks to show that what Goldman considers to be the foundation of social interaction is, in fact, unnecessary and most likely impossible. In terms of the crying babies example, Turner's starting point is the second baby's mirror-like contagion response. Rather than being the basis for producing a discrete internal representation, the second baby's mirroring response instead admits of greater refinement through continuous feedback. The initial contagion response is thus on a continuum with, say, an adult's ability to empathize, in the strong sense of

⁴⁹ I do not think the latitude offered by individually unique isomorphisms is enough to recapture the basic notion of voluntary agency. This relates back to my claim in §1.6 that there is no qualitative difference between Goldman's non-voluntaristic and voluntaristic cognitive states (e.g., belief versus attention, respectively).

consciously controlling one's feelings in response to another person's emotions. An adult's response possesses greater specificity in large part, Turner thinks, due to language, which enables the mature agent to mimic and express more. This implies that empathy, in its strong sense, doesn't consist of a cognitive agent identifying a correspondence between one's own representation and another's mental state, but rather the much more modest and open-ended picture of the empathizer acting in a manner that is sensitive to the other person's grief. One reason to endorse Turner's more modest picture of empathy is to address the basing-relation problem that Goldman claims that all epistemologies suffer from.

Pace Goldman's claim, I proposed in §1.5 that the issue stems from treating knowledge in isolation from the practical dimensions of cognition. Turner's characterization of the nursery example implies a pragmatic standard of epistemic success: empathy consists in acting in an suitable manner, not in identifying the other person's mental state; teaching someone how to drive consists in smoothing out any salient differences (i.e., any noticeable lack of functional equivalence). The general shift suggested by Turner's account is from justifying discrete beliefs to exhibiting a competency. As a result, there is no need to identify a unique justificatory base, since the "proof is in the pudding," so to speak. One of Turner's recurring lessons is that the opaqueness of social interaction doesn't just generate misunderstandings but is also what enables their resolution. If epistemic success required each subject to end up with matching representations, then all epistemologies would, in fact, be subject to the basing relation problem. But the practical success of "getting by" or "making oneself understood" is all that epistemic interaction requires and actually involves. The basic idea of continual refinement via feedback *sans* discrete representation—which underpins the claim that representation matching is

unnecessary—predates Turner’s interest in mirror neurons, relating back to an earlier interest in connectionist models of learning.

In the (2002) *Brains/Practices/Relativism*, connectionism represents the general need to scrutinize and displace much of social theory’s terminology. Turner’s point echoes the motivation behind Goldman’s epistemics project: epistemology and, by extension, social epistemology should be intimately concerned with the nature of the cognitive subject. For Turner, the most important lesson of contemporary cognitive science is that “every mind is the product of a distinctive and individual learning history” (1). Although other cognitive theories can capture this essential idea—Turner specifically mentions the Turing-Chomsky model as an alternative—Turner endorses connectionism. Connectionism pictures learning in terms of neural networks that are gradually modified by ongoing experience. The shifting strengths of synaptic connections correspond to the gradual refinement of a person’s behavior, the key idea being that the differential response is produced directly by the empirical input’s affect on the synaptic connections. Hence, while discrete representations may play a role at higher neural levels, they are unnecessary for learning skills or, more generally, learning any manner of acting (e.g., comforting a grief-stricken friend). Turner’s point is that since representations are unnecessary, it’s a bad idea to read them into places where they don’t seem to be. Turner appeals to mirror neurons in an attempt to build on this basic connectionist picture.

One of the general problems for connectionism concerns the speed of learning. Some learning appears to happen too rapidly for it to consist of the gradual modification of a neural network. Imitative learning, for instance, can appear to be more direct than the brute stimulation and consequent modification of a neural network, as connectionism by itself would imply. In this context, Turner sees mirror neurons as a mechanism that detects salient aspects of empirical

input (i.e., another person's embodied performance) and simultaneously creates the potential for the observer to mimic the behavior or at least aspects of the behavior. Mirror neurons thus offer an explanation of why social contagion responses are of distinctive importance as well as how some forms of social learning happen more quickly than the brute modification of neural networks suggest. Turner supports this connectionist-mirror neuron picture by linking it to empathy.

In (2012b), Turner presents his connectionist-mirror neuron account in terms of weak empathy. In contrast to strong empathy, which corresponds to the familiar notion of consciously controlling one's feelings, weak empathy is mostly unconscious. Weak empathy refers to the basic idea of the social contagion response described in Goldman's nursery example. The key move in (2012b) is to argue that weak empathy is not only explanatorily useful for unambiguously social cases of learning but also for clarifying the nature of objectivity. Turner's argument is filtered through Franz Brentano and Max Weber's respective conceptions of *Evidenz*—which roughly means immediately intelligible or self-evident—which anticipates Turner's own conception of objectivity as an inferential regress-stopper (397).

The weak empathy model's main interlocutors are normativist accounts, such as Searle's concept of we-intentions. Normativism offers its own account of objectivity within social interaction: we-intentions, for instance, are meant to explain why it is *wrong* to eat a baseball rather than try to hit it. The wrongness of eating the baseball doesn't, on this type of account, reduce to any combination of physical facts (e.g., the behavioral frequency of swinging or the inedibility of leather-coated cork and rubber) but rather the norm created by the we-intention of *playing baseball*. Appealing to physical facts appears to lead to an infinite-regress, leading to further questions such as, why is it common to swing at baseballs? why do many players chew

inedible tobacco product? and so on. Turner argues that his *Evidenz*-inspired concept of weak empathy can provide such a regress-stopper, but without needing to appeal to intrinsic normativity.

There are two symptomatic issues with the normativist approach. First, normativity is an inscrutable primitive about which nothing more can be said. As noted above, in order to square this with Searle's naturalistic project, this leads Searle to the very curious and questionable claim that we-intentions are a *biological* primitive. The second issue involves trying to characterize normativity. Attempts to do so face a dilemma: either attempt to identify a universal normative structure (i.e., the essence of rational discourse), or identify specific forms of normativity (e.g., natural language use, the discourse of particular scientific disciplines, legal proceedings, and so on). If choosing the former approach, then one has to explain the persistent lack of consensus on what the universal normative structure actually is, while the latter leads to a welter of hidden normative structures, since there must be one to cover every potential regress. Turner thinks that this dilemma is an artifact of the first issue, namely treating normativity as a primitive; thus by offering an alternative account of norms, Turner argues that weak empathy invalidates the very premise of normativism (2012b, 392).

Rather than tapping into a hidden structure, objectivity appears as an extension of learning contexts. Turner uses arithmetic ($2 + 2 = 4$) as an example of what may appear to be a universally objective truth and reconstructs it in terms of weak empathy (2012b, 394-95). There are two types of input in learning contexts: empathic input (e.g., figuring out what the teacher wants and observing the other students' responses) and empirical input (e.g., counting objects). Turner thinks that such experiences "empathically universalize" in the sense that any person who could learn arithmetic would come to take $2 + 2 = 4$ as self-evident after undergoing the

learning experience. Objectivity is thus an interpersonal understanding that reduces to something “seeming to be evident” rather than to being true (397).⁵⁰ One upshot of this picture is that it provides a straightforward explanation for possible future alternatives. Analogous to the case of Riemannian space in geometry, it is possible that a radically different form of arithmetic could be invented; given such an event, “ $2 + 2 = 4$ ” may lose the appearance of necessity but not its intelligibility, at least not to those people who have learned arithmetic as we know it (397). But the most important validation of the weak empathy model, which relates back to undermining the premise of normativism, is how it opens up for empirical enquiry what appears to be inscrutable from the perspective of normativism.

The weak empathy model converges with various other empirical considerations. This includes, of course, providing a suitable framework for explaining the underlying dynamics of tacit learning and also providing support for generalizing the imitative model of learning. With regard to the latter: by reconstructing the concept of objectivity in terms of interpersonal understanding, imitative learning is not merely an efficient means for acquiring factual truths. Hence, even though weak empathy’s main interlocutor is normativism, it substantiates the shift away from justification to pragmatic competency as indicated by Turner’s reconstruction of the nursery example. In terms of the arithmetic example, the equation “ $2 + 2 = 4$ ” reflects the ability to follow other people’s patterns of inference and recognizing those patterns as self-evident since everyone else follows the same inferential pattern; it does not reflect the discovery of a material truth, which would need to be justified by a strictly epistemic base. Accordingly, the basic lesson of the critique of normativism and the modification of Goldman’s nursery example are the same: interpersonal understanding does not require appealing to hidden elements—whether an intrinsic

⁵⁰ This account of objectivity underpins Turner’s reconstruction of Polanyi’s own account.

form of rationality or representational analogues to factual truths. It is worth more closely examining Goldman and Turner's respective accounts because despite their important differences, there is also a shared presupposition that I want to challenge in Chs. 3 – 5. To identify this shared presupposition, it is helpful to first more carefully examine the differences between Goldman and Turner's accounts.

The weak empathy model, in effect, dramatically loosens Goldman's fitting correspondence theory in two general respects. First, weak empathy implies there is a vastly greater variety of empirical input: this is because the physical constitutions of individuals inherently differ and also because of their unique experiential histories. While true of Goldman's account as well, the positing of individually realized isomorphic correspondence minimizes diversity's epistemic significance. Second, weak empathy implies that the articulation of our experiences is profoundly shaped by our social interactions. That is, my understanding of my experiences partly reflects what tacit aspects of experience I have made explicit, and what tacit aspects I have made explicit depend on who I have interacted with. In sum, the first point questions to what extent people correspond to the "same things," while the second implies that even if individuals could perceive the same things, an individual's experience is significantly influenced by contingent social interactive contexts, both past and present. Goldman's fitting correspondence theory is thus loosened to the point that the most epistemically salient goal is no longer finding a correspondence between a representation and an independent reality but instead the more modest and life-sized goal of attaining functional equivalence between my behavior⁵¹ and another's (whether trying to teach or learn a skill). Yet, as implied by describing the weak

⁵¹ Behavior is meant broadly to include verbal performances and any sort of cognitive activity.

empathy model as *loosening* rather than breaking the correspondence relationship, both Turner and Goldman are committed to materialism as a fundamental presupposition.

The basic picture of the cognitive subject, for Turner and Goldman, is of material input impinging on the subject, which the subject then processes,⁵² and which then affects the individual's behavior. Goldman and Turner differ over the details of each stage but agree on the basic input-processing-output picture. As discussed in regards to each specific account, the mirror neuron is an attractive concept because it addresses one of the basic difficulties of this picture, namely accounting for how individuals determine the salient aspects of brute empirical input: for Goldman, this issue appears in terms of the origin of internal simulation routines; for Turner, how to account for rapid learning processes. The challenge of accounting for perceptual salience relates to the Meno Paradox: it is difficult to characterize how agents discriminate what is perceptually salient without thereby implying they *already* know what they're actually learning from the perceptual experience.⁵³ I think the notion of brute empirical input as the first isolable stage of cognition ultimately limits Turner's account. To clarify the issue, it helps to evaluate the weak empathy model in terms of the four presuppositions/symptoms of the testimony view that I noted at the beginning of the chapter.

While it problematizes all four of testimony's noted features, I think that the weak empathy model only undermines three of them. The first and fourth features—testimony's abstract focus on justification and its division of epistemic and practical matters—were undermined jointly by weak empathy's conception of objectivity as interpersonal understanding. The second feature regarding the circulation of discrete tokens was directly undermined via

⁵² I use "processing" in the broad sense of any cognitive activity internal to an organism. In Ch. 3, I contest the notion of delimited and isolable *internal* cognitive activity.

⁵³ In this regard, Polanyi's concerns overlap with Hanson (1958) on discovery and Gibson's (1966, 1) analysis of "to sense" as *detecting* something as opposed to having a sensation.

offering an alternative to Goldman's representation matching picture. But it is the third issue—viewing epistemic agents as passive informational receivers⁵⁴—that the weak empathy model *still* endorses.

All of the epistemologies that I have examined thus far consider empirical input to be the isolable first stage of cognition. Cognitive subjects must receive empirical input before any processing can occur and are thus *passive* in the sense that cognitive processing is derivative on the received input. Materialism provides strong intuitive support for this picture: if agents and objects are individuated as extensionally delimited entities, then perceptual input is required for bridging the material gap. Weak empathy's particular conception of this relationship—its connectionist element—depicts an individual's experiential history (i.e., the sum of empirical input) as being inscribed on or effectively imparted to an individual's brain (or the brain and central nervous system). Hence, while weak empathy rightly emphasizes the pragmatic dimensions of cognition, its endorsement of the input-receiver picture places it in something of a hybrid position: the cognitive agent's passivity has been thinned out to a bare minimum, suggesting that the agent is a very active receiver but a receiver nonetheless. The enactivist concept of sense-making, which I discuss in Ch. 3, challenges the input-receiver picture of cognition. But in the present context, I want to examine the input-receiver picture by linking it to a common understanding of naturalism, which Searle (1995) articulates when claiming that every educated person should accept two fundamental scientific theories.

As discussed in §2.2, there are reasons to doubt whether Searle's two fundamental theories, as he understands them, are in fact fundamental. I think Searle's sketch is, nevertheless, helpful in articulating a pervasive though often assumed understanding of naturalism. The

⁵⁴ I am using "information" in the broad sense of any empirical input.

materialism noted in the above paragraph—which appears in Searle as the atomic theory of matter—is often paired with an adaptationist interpretation of evolutionary theory, in which the principle of natural selection is the sole general principle for explaining organisms’ evolutionary history. Adaptationism’s exclusive focus on the survival of the fittest dovetails with materialism in that scarcity of material resources is considered to be the formative influence on all evolved characteristics. Since Turner has ably dispatched Searle’s (and Goldman’s) accounts, I will focus on how the materialist-adaptationist framework manifests itself in the weak empathy model.

Naturalism as materialism-adaptationism manifests itself within the weak empathy model in terms of three general issues. They are as follows:

- (1) an impoverished account of the origin of normativizing beliefs,
- (2) the epistemic significance of social interaction is reduced to instrumental benefits,
- (3) the brain is viewed as the executive controller of all cognitive activity.

With regard to all three issues, Turner would grant each point but deny that they are, in fact, liabilities. Substantiating my criticisms will occupy Chs. 3 – 5; so for the moment, I only want to clarify each issue in order to set the explanatory targets of each chapter. With regard to (1), I accept Turner’s critique of intrinsic normativity insofar as it implies human exceptionalism, but, on the other hand, I think there needs to be a more substantive account of the origin of *normativizing* beliefs.⁵⁵ More specifically, Turner takes it as a given that an individual has normative beliefs about the world and from this shows that he can account for normativity within explicitly social contexts. That Turner doesn’t see the need to account for the origins of an individual’s normative beliefs reflects, I think, the belief that something like the generic notion of a survival instinct unproblematically explains the origin of normativizing beliefs.

⁵⁵ In Ch. 3, I use enactivism and the related Dynamic Systems Theory (DST) framework to discuss the origins of normativity. Furthermore, DST provides an alternative understanding of naturalism broadly referred to as methodological naturalism.

Granting a fundamental role to a survival instinct suggests that the fundamental importance of knowledge is the material benefits it provides, an inference that leads to issue (2). If the fundamental purpose of knowledge is the material benefits it provides, then it is a very small step to inferring that the epistemic significance of social interaction should be explained in terms of its instrumental benefits.⁵⁶ To some extent, weak empathy can obviate this concern by emphasizing the primitive role of contagion mirroring, but the attendant epistemic standard of functional equivalence still reduces other people to a means for acquiring a skill. This latent manipulationism also manifests itself in terms of the one-factor explanation issue discussed in §1.7. There I noted a similarity between Goldman's approach to the motivation problem (§1.6) and Craig's definition of knowledge as a social good (§1.7); the presumption that the primary function of knowledge is to serve an instrumental benefit leads both to posit there being one essential factor underlying diverse social phenomena. Issue (3) of the weak empathy model exhibits this one-factor tendency in the form of trying to reduce all cognitive activity to the weighting of synaptic networks.

Whether the brain is the executive controller of all cognitive activity is an empirical question. In this respect, given the shifting sands of cognitive science, it is a difficult question to approach. Yet, even conceding this proviso, I think that the weak empathy model places too much explanatory weight on neural networks (i.e., connectionism) and particular kinds of neurons (i.e., mirror neurons). The explanatory tasks that connectionism and mirror neurons confront are, I think, better addressed by avoiding them, which is possible by avoiding the input-receiver picture and its underlying egoism. I see this shift in the spirit Turner's own insistence on

⁵⁶ Clark's EM hypothesis pursues this line of reasoning to its logical conclusion, treating other people strictly in terms of their functional role in a cognitive process, on par with any other tool. In Ch. 5, I consider two appropriations of EM that suggest that intersubjectivity involves a primitive or at least non-instrumental cognitive dimension.

not diminishing the importance of social interaction by appealing to hidden rational structures or representations. While neural networks and neurons are not hidden in the same sense as something like intrinsic normativity, the widely divergent and conflicting explanatory uses that mirror neurons have been put to—such as feeding input to internal simulation routines or the instantaneous weighting of connectionist neural networks—should be serious cause for concern. The rhetorical uses of mirror neurons suggest they are analogous to a Rorschach test in which the sundry strands of materialist accounts read into the mirror neurons whatever suits their particular needs. Regardless of the ultimate status of mirror neurons, I think that recognizing the role of community-level epistemic processes indicates that not so much should hang on the outcome.

One basic difference between weak empathy and the transformation view, which I will be arguing for, is that the former reduces all social interaction to the one-on-one framework of the imitation model. The latter, by contrast, recognizes epistemically significant processes that emerge and operate at the level of communities, communities greater than two individuals. These processes presuppose their individual participants but are irreducible to any individual cognitive process.⁵⁷ If there are emergent cognitive processes operant at the level of communities, then *not every* cognitive process needs to be anchored in an individual center. Hence, I think that such community-level processes suggest that appeals to neuroscience should be more circumscribed and more judicious.

From this discussion of the weak empathy model, it is possible to round out the sketch of the translation view offered in §2.3. The weak empathy model offers a cognitive basis for generalizing the importance of imitative learning, as was suggested by tacit knowledge. Weak empathy sketches a form of primitive, pre-conceptual imitation that is an innate ability of the

⁵⁷ I discuss this idea in Ch. 4 within the context of the neo-Kuhnian social epistemologies of Fred D'Agostino and William Rehg.

human brain (and some other animals). Although beginning at this primitive level, Turner believes it is on a continuum with and bleeds into conscious processes, especially strong empathy. Weak empathy attempts to explain how, despite the profound and ineliminable differences between individuals, social interaction is epistemically valuable. In sum, other people amount to opaque exemplars that, by virtue of mirror neurons, we are predisposed to emulate.

While the translation view represents a significant advance beyond the testimony view's model of social interaction, its account is ultimately too corrosive. That is to say, the epistemic significance of social interaction is *not* reducible to the sparking of mirror neurons. The shortcomings of the translation view has been previously hinted at, such as with Polanyi's concept of tradition and Kusch's insistence on the epistemic significance of community membership. The latter is particularly noteworthy since despite the very real differences between Kusch's communitarian epistemology and Turner's translation view, I think their respective critiques of social theory offer two complementarity perspectives on one larger issue. Thus, before pursuing an alternative to the translation view, it is first worth synthesizing these two insightful critiques.

2.5 Social Theory: Individualist or Collectivist?

The dichotomy between individualist and collectivist/communitarian social theories, while of some value, ultimately confuses the issues it is meant to address. This is seen, for instance, in comparing Turner and Kusch's respective uses of the dichotomy. While the former's critique of collectivist concepts and the latter's critique of individualist testimony appear to pit each in diametric opposition to the other, I think a closer look paints a much more complex picture, one that transcends the individualist-collectivist dichotomy. Assessing the translation view in terms of Kusch's four point critique of individualist testimony is a helpful starting point.

Turner's individualist social theory raises many of the same concerns with the testimony view as Kusch's communitarian epistemology. Of Kusch's four main criticisms, Turner agrees on three. For both, testimony is wrong to posit social interaction as a merely derivative source of knowledge since, leading to the second point, interaction does not consist of the transmission of complete items of knowledge. The translation view also corroborates Kusch's third criticism—that items of knowledge are delivered solely by perception, memory, and inference—but for different reasons. The weak empathy model grounds all knowledge in perception (i.e., the input-receiver picture) but undermines the notion of itemized knowledge in favor of the continuous refinement of skills. Since I scrutinized the input-receiver picture in the previous section, I will focus on the translation view in relation to Kusch's fourth criticism of testimony, namely that the “testifier” and “recipient” need possess only minimal social knowledge of each other.

By emphasizing the individual differences and the noisiness of social interaction, the translation view claims that people, in fact, only possess minimal knowledge of each other, social or otherwise. For Turner, Kusch's first three criticisms reflect the misguided belief that epistemic interaction depends on a shared foundation. This shared foundation takes the form of either a shared currency (e.g., Searle's rational normativity) or the basis for a shared currency (e.g., Goldman's view that objective facts are the basis for sharing propositional beliefs). Turner's attack on collectivist concepts aims to counteract the glossing over of individual difference and the whitewashing of social interaction wrought by the appeal to a shared foundation. For Kusch, of course, the three criticisms point in the opposite direction—to the neglect of epistemic communities.

The basic problem with the testimony view, for Kusch, is the conviction that all cognition occurs within a discrete isolable individual. The fourth point regarding the importance of social

knowledge serves as a jumping off point for Kusch's communitarian epistemology. Two interacting cognitive subjects are, Kusch thinks, often members of a shared epistemic community and this shared membership directly affects the epistemic interaction. While I noted some issues with Kusch's proposal in §1.7, I think the basic idea that community membership and social identities play a direct role in epistemic interaction is a valuable insight worth retaining. Towards this end, Turner's critique of collectivist concepts is a useful standard for refining Kusch's understanding of community. By way of introducing this revision, it is first worth noting an underlying connection between the collectivist and individualist critiques.

Vacuous collectivism goes hand-in-glove with the disconnected individualism that Kusch diagnoses. As touched on in relation to testimony's motivation problem, picturing epistemic agents as isolated units creates spurious gaps. In order to present something that resembles a *social* theory, it is necessary to suture over these gaps, creating the need for vacuous collectivist concepts. In short, if you carve epistemic agents up into self-contained units, then you have to somehow explain the stickiness of other people. This is true even if it's a very weak adhesive, such as an innate disposition to mirror others. The fundamental issue—what I think is antithetical to a properly *social* epistemology—is the conviction that there must either be an individual mechanism or perceived instrumental benefit that steers an individual towards interacting with other individuals. D'Agostino's portrait of epistemic communities is an example of an account that transcends this understanding of the individualist-collectivist dichotomy.

D'Agostino's concept of shallow consensus—mentioned previously in Ch. 1—depicts a form of community membership that shuns the collectivist notion of a shared foundation. Shallow consensus refers to D'Agostino's (2010) claim that agreements within epistemic communities occur against a backdrop of extensive disagreement. This applies to all aspects of a

community—from values of deliberation and evaluative standards to, in the case of scientific communities, the particular judgments relating to theory choice. Shallow consensus is essential for reconciling an epistemic community's dual need for solidarity and diversity (89). With regard to solidarity, research teams and individuals within a research team are able to reach consensus despite important disagreements, many of which they are not even aware of, which is necessary when dealing with extremely complex and multifaceted problems. This fragile solidarity simultaneously ensures that consensual judgments do not preclude diverse approaches to and explorations of a problem space (91). Unlike Kusch and Turner, D'Agostino sees the individualist and collectivist poles in terms of a productive tension.

Social epistemology is itself a noisy and multifaceted problem space. In terms of this picture, I see Turner and Kusch as exploring different patches of the problem space—imitative learning and performative language-use, respectively—and D'Agostino mapping the problem space itself. By themselves, Turner and Kusch's models are only one-factor accounts of social interaction but when assembled together as part of D'Agostino's heterogeneous map, each offers a valuable perspective on what are in fact thoroughly heterogeneous phenomena. But in order to motivate this motley assemblage—which I detail in Ch. 4—it is first necessary to displace the input-receiver picture noted above, since it is this picture that implies that it is possible to isolate and dissect a fundamental nature at the root of the cognitive subject. In the following chapter, I draw upon enactivism in order to reconceptualize the cognitive subject in terms of an essentially active agent who cannot be examined apart from interactions with his or her contingent environment.

Chapter 3: Coordinated Interaction

3.0 Introduction

Over the course of Chs. 3 – 5, I assemble together a joint approach to cognition and social epistemology, which I call the transformation view. The central tenet of the transformation view is that cognition, especially cognition within a social context, is a transformative process. Or to put the same point polemically, the transformation view objects to the notion of there being an *immutable* innate cognitive core. While humans and other animals are not born as a *tabula rasa*, I present a framework in which innate cognitive structures are profoundly plastic, shaped by the contingencies of organisms' individual and idiosyncratic development.

In the present chapter, I employ the enactivist approach to cognition in order to supplant the input-receiver picture identified in Ch. 2. Accomplishing this goal involves the following three tasks:

- (1) provide a non-exceptionalist, non-saltationist naturalization of cognition,
- (2) provide an account of intrinsic teleology, and
- (3) offer an empirically grounded framework for evaluating intersubjectivity that transcends the individualist-collectivist dichotomy.

To achieve these tasks, my use of enactivism centers on the concept of participatory sense-making, as articulated by Hanne De Jaegher and Ezequiel Di Paolo (2007, 2008). Participatory sense-making sets the stage for examining community-level dynamics within social epistemology in Ch. 4 and displacing manipulationism as a foundational assumption of cognitive science and social epistemology in Ch. 5.

While I endorse the concept of participatory sense-making, extending its insights beyond dyadic interaction comes with difficulties. I examine what I think are some flawed attempts to expand the scope of participatory sense-making. Most striking is Steiner and Stewart's (2009) argument that participatory sense-making presupposes a background social domain that consists of structural norms. While intrinsic normativity plays an essential part within enactivism—as indicated by (2) above—Steiner and Stewart's proposal reflects, I argue, a re-entrenchment of a normativism similar to Searle's (1990, 1995) and which is antithetical to enactivism. In criticizing Steiner and Stewart and others, I chart a different path for enlarging the scope of participatory sense-making, one that links it to neo-Kuhnian social epistemology rather than to Émile Durkheim.

The present chapter consists of three major parts. In the first (§3.1-3.2), I trace enactivism from its beginnings in autopoiesis up to the concept of sense-making. As argued by Evan Thompson (2004) and Di Paolo (2005), the concept of sense-making involves a break from the strict structural determinism of autopoietic machines, a break that Varela alludes to in Weber and Varela's (2002) argument for intrinsic teleology. This opens up the conceptual space for De Jaegher and Di Paolo's concept of *participatory* sense-making, which I explicate in §3.3. The last part of the chapter (§3.4-3.5) considers attempts to expand the scope of participatory sense-making beyond discrete two-agent interactions. Despite the difficulties and dangers of expanding the scope of participatory sense-making, I argue that it is a fruitful tool for analyzing larger-scale and longer-term interactive contexts, an application that is mutually beneficial for social epistemology and cognitive science.

3.1 From Autopoietic Machines to Sense-Making: The Enactivist Approach to Cognition

The concept of sense-making—one of the core concepts of enactivism—represents the culmination of more than two decades of work that began with Humberto Maturana’s interest in the biology of cognition. Maturana’s (1970, 13) claim that “living as a process is a process of cognition” shifts towards the more refined notion of “living is sense-making” (Thompson 2004, 386). Explicating the concept of sense-making sets the stage for De Jaegher and Di Paolo’s concept of *participatory* sense-making, which, as the name suggests, is a direct extension of the older concept. The concept of sense-making, however, has something of a tangled interpretive history, which I unravel in the present section, beginning with its roots in the theory of autopoiesis.

In a retrospective article, Francisco Varela (1996) attributes his joint work with Maturana on autopoiesis to two general dissatisfactions. First, like Dreyfus (1972), Varela and Maturana were dissatisfied with the notion of information as the key to understanding cognition (410). While Dreyfus filters this concern through Heidegger and the topic of expertise, Varela and Maturana were concerned with providing an adequate biological framework for cognition. Hence, the second dissatisfaction concerns accounts of the biological cell: “one talked about the molecular constitution of the cell, and used terms like self-maintenance, but no one, not even two reunited Noble prize winners [George Wald and James Watson], knew what was meant by that” (412). It wasn’t simply that accounts of cellular self-maintenance were inadequate but that most biologists dismissed it as an idle question. In this context, autopoiesis was an intervention in cellular biology, which Maturana and Varela hoped to extend to more complex life forms. Autopoiesis signaled a fundamental epistemological and ontological change in how to

understand living organisms, a change that has immediate implications for the nature of cognition.⁵⁸

In the first major English language publication of the theory of autopoiesis, Maturana and Varela (1980) offer a concise definition of it and present the minimal case of cellular self-maintenance. An autopoietic machine is defined as follows:

A network of processes of production (transformation and destruction) of components that produces the 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 the space in which they (the components) exist. (79)

Cells are the paradigmatic exemplar of an autopoietic machine, absorbing external nutrients in order to produce monomers, which, in turn, polymerize to form the cell's membrane. Because maintaining the cell's membrane is a continuous undertaking, as expressed in (i), Maturana and Varela identify the cell in terms of the homeostatic process of renewal rather than in terms of its material composition at any given moment. While autopoiesis, as cellular self-production, is a spatially bounded phenomenon, other biological processes of self-renewal need not be.

Autopoiesis is the most basic type of living organization, from which different forms of biological autonomy can emerge.

Varela (1979, 1997) defines biological autonomy in terms of operational closure. Operational closure is “a circular reflexive interlinking process, whose primary effect is its own production” (Varela 1997, 73). As circular self-production, operational closure corresponds to (i) above. But by abstracting away from the cellular context, the principle of biological autonomy possesses a much wider scope of application, including multicellular organisms such as humans

⁵⁸ Maturana (2002, 16) describes the basic insight of autopoiesis as shifting from the question of “What is the essence of that which I observe?” to the question of “how do I do what I do as an observer in observing?”. This departure from cognitive realism will be further detailed in the present section.

and various biological processes within such organisms.⁵⁹ Enactivism is founded on this more abstract principle of autonomy rather than autopoiesis, since it is the former that has direct implications for characterizing all cognitive subjects, including humans (Thompson and Stapleton 2009).

In collaboration with Evan Thompson and Eleanor Rosch, Varela introduces the “enactive” approach to cognition in the seminal 1993 book *The Embodied Mind*. Enactivism is presented in contrast to cognitive realism, which the authors characterize in terms of a theory of representation that is similar to what I labeled the input-receiver picture. The realist theory of representation has three essential features: “(1) the world is pregiven; (2) our cognition is of this world—even if only to a partial extent, and (3) the way in which we cognize this pregiven world is to represent its features and then act on the basis of these representations” (135). The defining characteristic of cognitive realism is thus positing an independent pre-given world as the foundation for and target of cognition [(1) and (2)], while (3) denotes the representationalist account of action. In choosing the “input-receiver” label, I emphasize the essentially passive first stage of cognition shared by *all* realist theories and thus avoid overemphasizing the question of representation that is—as Turner’s weak empathy model shows—to some extent a red herring.⁶⁰ In this originary text, in spite of marshaling a great deal of empirical evidence in support of enactivism, the authors still lean heavily on the theory of autopoiesis and enactivism’s contrast to cognitive realism.

After an extensive discussion of color vision, the authors present enactivism as asserting two key points. They are as follows: “(1) perception consists in perceptually guided action and

⁵⁹ Varela was especially concerned with the autonomy of the immune and nervous systems, which are distributed systems. Thompson and Stapleton (2009) offer insect colonies as another example.

⁶⁰ For another characterization of cognitive realism, see Susan Hurley’s (1998) discussion of the “input-output model.”

(2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided” (173). The first point corresponds to the idea of a perceiver-dependent world—perception is an endogenous activity of an organism, not a recording of external reality. This point reformulates one of the original consequences of autopoietic organization. Maturana and Varela (1980) write:

Autopoietic machines do not have inputs or outputs. They can be perturbed by independent events and undergo internal structural changes which compensate these perturbations. ... Whichever series of internal changes takes place, however, they are always subordinated to the maintenance of the machine organization. (81)

Autopoietic machines are affected by external events but only in relation to their available range of internal compensations, which are subordinate to self-maintenance. Anything falling outside this range is either indifferent to the autopoietic machine or results in its disintegration. This is the basic sense in which enactivism’s perceiver-dependent world does not reduce to solipsism or an attenuated realism. Autopoietic self-organization is the minimal form of activity from which more complex forms of activity, including cognition, emerge; documenting this emergence corresponds to the second point above, with vision studies playing a central role.

Visual perception has played a special role in autopoiesis and enactivism since Maturana’s earliest ruminations on biological cognition. Maturana (2002) attributes his original doubts regarding cognitive realism to his attempts to understand pigeon vision in the 1960s. Furthermore, visual perception lends strong intuitive support to the realist standpoint and so being able to explain it from an enactivist perspective helps address some deep-seated resistance to the framework. Studies of visual perception are so important to enactivism that they form the principal topic of one of enactivism’s two major strands (Torrance 2006).

For Varela, Thompson, and Rosch, the color vision study is a bridge-building exercise that links their broad interest in cognition and naturalism to technical accounts of perception. In

contrast to this comprehensive project, Steve Torrance (2006) recognizes a second major strand of enactivism, exemplified by O'Regan and Noë's (2001) account of visual perception. In the spirit of enactivism, O'Regan and Noë explain visual perception strictly in terms of an agent's sensorimotor coupling with the world, though they avoid the broader concerns of Varela, Thompson, and Rosch. Unlike the color vision study, O'Regan and Noë don't offer their account as a bridge between minimal and human forms of cognition but instead as a standalone account of human visual perception. Furthermore, O'Regan and Noë's sensorimotor contingency theory diminishes the importance of first-personal experience and thus the relevance of phenomenology, as noted by Thompson (2007) among others. The two strands of enactivism thus mark a productive tension—the narrow focus of strictly sensorimotor accounts substantiating the general enactivist orientation—but also an indication of some unresolved difficulties. In §3.4, for example, I argue that if enactivism is identified with the narrow strand, then it is vulnerable to Clark's (2008) criticism that it reduces to a species of extended functionalism. One of the most important differences between the broad and narrow strands—an issue indirectly alluded to in regards to the role of first-personal experience and phenomenology—concerns the status of teleology within enactivism.

One of the chief values of autopoiesis is its mechanistic characterization of self-organization. As suggested in the definition above and reiterated in Maturana (2002), autopoietic machines are structurally determined, characterizable strictly in terms of their present coupling state. While Varela never dismisses autopoiesis, he ultimately reinterprets it such that it includes a teleological component. This marks an important dividing line between broad and narrow enactivism, but before examining the details of Varela's re-interpretation, it is helpful to first consider Thompson's (2004) intimate perspective on Varela's previous reluctance. Thompson

relates Varela's response to a June 1999 email arguing that phenomenology is inconsistent with anti-teleology as follows: "he [Varela] preferred to 'shift the accent' from teleology to original intentionality, understood as the sense-making capacity proper to autopoietic units" (395).

Thompson was dissatisfied with this response since he thought that both the notions of original intentionality and sense-making possess a teleological element. In a posthumously published paper co-authored with Andreas Weber, Varela indirectly concedes the point by endorsing a specific type of teleology.

Weber and Varela (2002) reappraise the role of teleology in relation to modern biology as a whole. Teleological descriptions are common in modern biology, despite the consensus view that Darwin's evolutionary theory proves that teleology is absent from biological nature. The authors note that this tension is usually explained away as a methodological fiction: biological facts have a semblance of goal-directedness due to the influence of natural selection, a position known as teleonomy. In the next section, I examine the teleonomy position in greater detail, but for present purposes, it is sufficient to note that Weber and Varela argue for a form of teleology that is compatible with Darwinian evolution. Weber and Varela distinguish between two different but often conflated senses of teleology: "external seemingly purposeful design" and the "internal purposes immanent to the living" (100). While Darwinian evolution undermines the former, the authors argue that the latter is essential for understanding the organism as a unitary locus of experience. Immanuel Kant's ([1790] 1987) discussion of purposiveness is instructive both as an anticipation of Weber and Varela's formulation of intrinsic teleology and also as an illustration of the strategy by which many theorists explain away intrinsic teleology. Kant's concept of purposiveness is often interpreted as *only* referring to an artifact of human cognition: on this reading, if humans could intuit the physical universe, then organisms would appear as

mechanisms like the rest of nature. While conceding that Kant's own position is unstable, Weber and Varela argue that there has been enough progress in science and philosophy to show that purposiveness is, in fact, an empirical phenomenon common to all known organisms.

The progress that Weber and Varela have in mind consists in two parallel trends in twentieth century philosophy and biology, namely the aforementioned autopoiesis theory and Hans Jonas's (1966) biophilosophy. Jonas notes that metabolism is common to all life but absent from matter (3). Inherent to metabolism, Jonas argues, is a value-laden perspective, a "vital identity" or "living form," from which organisms exist as a continuous and total material flux (76). Jonas's ruminations on the character of life were inspired in part by a frustration with existentialism's myopic focus on humanity, metabolism providing a broader means for understanding the nature of agency and emotion, both human and otherwise. In this vein, but going in the opposite direction, Jonas generalizes Heidegger's description of humanity's primordial concern with its mortal existence as an indication that nature shouldn't be viewed as an aggregate of indifferent forces (233). This is one of multiple ways in which Jonas argues for intrinsic purposiveness within nature, another being his critique of cybernetics (116). While the affinities between Jonas's notion of metabolic value and autopoiesis are clear, Weber and Varela argue for a more substantive relationship.

Weber and Varela (2002) present autopoiesis as empirically grounding Jonas's characterization of metabolism. That is, in conceptualizing the self-organizing properties of matter, autopoiesis suggests how life could have emerged from brute matter. Mechanical causality can thus account for the nature of life when understood in terms of circular causation. Looking at the connection from the other direction, Varela uses Jonas to guide his reinterpretation of autopoiesis as involving two modes of intrinsic teleology. The two modes are

as follows, “First, a *basic* purpose in the maintenance of its own identity, an affirmation of life. Second, directly emerging from the aspect of concern to affirm life, a *sense-creation* purpose whence meaning comes to its surrounding” (117; original emphasis). The first mode corresponds to the original definition of autopoiesis as a self-maintenant network of processes, while the second is presented as a consequence of the first. This elaboration relies on the notion, noted above, of an autopoietic machine’s range of internal compensations: the basic idea is that every organism has a range of compensations and, accordingly, a range of interactions available to it, which encompasses the sense it can make of its environment. Thus, *sense* is a product of an organism’s interaction with its surroundings. In sum, Weber and Varela’s endorsement of intrinsic teleology is presented as a clarification of autopoiesis that redresses modern biology’s inadequate conception of the organism.

Varela’s modified interpretation is a fork in the road for the theory of autopoiesis. On the one hand, it is possible to claim that autopoiesis—by revealing more of the mechanistic underpinnings of living organisms—further suggests that teleology is a superfluous metaphysical extravagance to be avoided. Introducing intrinsic teleology would be seen in this regard as muddling the picture. Such a response is suggested by O’Regan and Noë’s displacement of first-personal experience. On the other hand, Varela’s endorsement can be seen as an advance that unlocks more of cognition’s dimensions. Following this second path, I align myself with Di Paolo’s (2005) emphasis on operational terms, though in §3.3 I note a danger in exclusively focusing on them.⁶¹

⁶¹ Thompson (2007) offers a distinct second way of pursuing Varela’s endorsement of intrinsic teleology. Thompson emphasizes phenomenology and is primarily interested in supporting Godfrey-Smith’s (1994) strong continuity of life and mind thesis. While I am more closely aligned with Di Paolo’s focus on operational definitions in enactivism, I note the irreducible significance of first-personal experience in §3.3 and Ch. 5.

For Di Paolo (2005), Weber and Varela's argument for intrinsic teleology points in the right direction but falls short. The crux of the issue relates to the concept of sense-making: while autopoiesis provides the language needed to talk about sense-making, Varela's re-interpretation conflates its conservative and adaptive aspects.⁶² Autopoiesis entails the norm of conservation but not adaptivity. To delineate these two aspects, Di Paolo uses the example of falling off of a cliff. Balancing on the edge of a cliff and falling over the edge are—as far as autopoietic conservation is concerned—viable interactions, since neither cause system disintegration (436). This short-sighted perspective results from autopoiesis's structural determinism, which characterizes a system strictly in terms of its present state; hence, it is only the thud against the ground that violates conservation. This is, of course, useless to the organism that fell, since the value judgment comes only after disintegration.⁶³ This limitation doesn't indicate a flaw with autopoiesis, since its structural determinism underpins its systemic grounding of cognitive terminology, which, in turn, is the foundation for its critique of cognitive realism (434). Rather than jettisoning or re-defining autopoiesis, Di Paolo introduces adaptivity as a separate operational term.

The concept of sense-making entails both autopoietic conservation *and* adaptivity. In order to show that it is consistent with enactivism's systems framework, Di Paolo offers the following operational definition of adaptivity:

A system's capacity, in some circumstances, to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the boundary of viability. ... 1) Tendencies are distinguished and acted upon depending on whether states approach or recede from the boundary and, as a consequence, 2) Tendencies of the first kind are moved closer to or transformed into tendencies of the second. (438)

⁶² In the next section, I clarify how intrinsic teleology relates to sense-making. For now, it is enough to note that sense-making articulates a form of intrinsic teleology.

⁶³ Di Paolo also recasts the same point in terms of bacteria, one of Varela's favorite examples. Given only the norm of conservation, autopoiesis accounts for why bacteria metabolize sugar, but not why bacteria swim up a sugar gradient to higher concentrations of sugar (437).

Adaptivity presupposes but is distinct from autopoietic robustness. While robustness provides the basic all-or-nothing norm of viable self-identity, adaptivity specifies an organism's current status in relation to its viability. Adaptivity implies that an organism possesses *some* kind of awareness of its viability boundaries and that the organism attempts to mitigate if not reverse a destructive tendency via modifying its internal state or environmental coupling. Thus described, adaptivity is grounded in systemic operational terms, though not the structural determinism of autopoiesis. It is for this reason that Di Paolo views adaptivity as emergent upon autopoiesis.

The chief obstacle to recognizing robustness and adaptivity as distinct is that all known organisms exhibit both characteristics. Yet there are synthetic cases of autopoiesis without adaptivity, such as Tibor Ganti's Chemoton reaction (Di Paolo 2005, 450). Furthermore, the robustness-adaptivity distinction coheres with the two distinct connotations of sense-making, namely uncovering and generating meaning (450). On this reading, uncovering meaning or "making sense" refers to the basic process of self-maintenance (i.e., robustness); generating meaning, on the other hand, consists in retroactively attributing significance to previously neutral encounters based on whether they caused the organism to approach towards or recede from the boundaries of viability. Enactivism has emphasized sense-making's generative connotation only because it contrasts more sharply with cognitive realism. While Di Paolo remains committed to these two basic connotations, he later modifies the concept of sense-making by distinguishing it from adaptivity.

Di Paolo, Rohde, and De Jaegher (2010) argue that adaptivity is a necessary though insufficient condition for sense-making. In the definition of adaptivity above, there are two ways an organism can address a destructive tendency: changing its internal state and/or changing its relation to its environment. In the 2010 article, Di Paolo, Rohde, and De Jaegher restrict

adaptivity to the former capacity, namely the ability to change one's internal state. Sense-making, by contrast, implies the additional capacity of motility. Hence, the most basic way of "changing" or "regulating" one's environmental relation is moving to a different place. The tumbling and swimming of many types of bacteria exemplify a very primitive form of motility that enables the bacteria to find higher concentrations of sugar. While all known organisms exhibit adaptivity, the authors offer green sulfur bacterium as an example of a just-adaptive, nonmotile organism (50). The upshot of distinguishing between just-adaptive and sense-making organisms is not a neat taxonomy—the corresponding taxonomy, in fact, has a number of borderline cases such as sessile or mostly sessile aquatic animals—but rather the distinction's implications for understanding agency. In this respect, the borderline cases redound to the credit of the distinction, since they illustrate subtle differences within minimal forms of agency.⁶⁴

The refinement of sense-making supplements Di Paolo's (2005) sketch of agency. Agential behavior is not simply an organism's structured coupling with an environment but additionally an organism's regulation of the coupling (442). The latter, unlike pure autopoietic coupling, has an intentional structure for the organism: it is possible for the organism to detect whether a regulation has succeeded or failed in improving its viability condition. Furthermore, as alluded to regarding the generative meaning of sense-making, this primitive agency entails historicity, since an organism's regulative responses mark and constrain its future interactions (444). Reading this sketch in terms of the refined notion of sense-making, the three terms of autopoiesis, adaptivity, and sense-making denote two constraining but underdetermined relationships. In the first, autopoietic self-identity sets the open-ended goal towards which

⁶⁴ Furthermore, I do not think a lot hangs on whether adaptivity alone should be counted as a form of agency. More important is recognizing how adaptivity informs motility and how motility enhances an organism's range of possible adaptive responses. Hence the value of having operational terms for each distinct aspect. See Fitch (2008) for a related analysis of different minimal forms of cognition, which he formulates in terms of nano-, micro-, and proto-intentions.

adaptive responses are geared. Autopoiesis and adaptivity together capture the essential nature of metabolism. In the second constraining relationship, metabolism acts as the open-ended goal for sense-making activities. Sense-making activities generate values that are constrained but not determined by metabolism and, given sense-making's historical dimension, the constraints set by metabolism can become highly mediated and thereby attenuated.

It was Di Paolo and De Jaegher's (2007, 2008) work on the generation of values within social contexts that prompted them to emphasize the specifically motile dimension of sense-making. The basic premise of *participatory* sense-making is that individuals influence each other's bodily movements ubiquitously and often through simple, non-mentalistic means. Having identified the regulation of bodily movement as the most primitive form of agency, the concept of sense-making suggests that anything that consistently shapes an individual's movement is *much more* than an incidental input for an "actual" cognitive process. By grounding cognition in viable interactive regulation, the concept of sense-making provides a less saltationist and exceptionalist account of human cognition, a framework that undermines the input-receiver model of cognition. My chief concern in the present chapter is to show how this naturalization of agency and cognition provides a more comprehensive and straightforward framework for social epistemology, as initiated by the concept of participatory sense-making. But before examining participatory sense-making, it is first necessary to situate the basic enactivist model of cognition in relation to two criticisms that I leveled in Chs. 1 – 2.

As noted in the present section, the enactivist approach that I am endorsing is committed to a specific kind of teleology. Furthermore, this teleology is expressed in terms of an organism's adaptivity. Adaptivity may appear to be at odds with my previous critiques of adaptationism, as discussed in regards to Goldman's testimony model and Searle's Background hypothesis. And

my endorsement of the more general notion of intrinsic teleology may appear to be at odds with my critique of Searle's concept of intrinsic normativity. In the next section, I differentiate my enactivist commitments from these positions. Doing so offers a broader perspective on the issues that enactivism has negotiated in relation to its shifting interpretations of sense-making.

3.2 Teleology, Adaptivity, and Other Forbidden Things

The shifting interpretations of autopoiesis and sense-making reflect tensions within modern biology as a whole. André Ariew and Mark Perlman (2002) offer a general perspective on the topic, boiling the issue down to the following question: is there “a naturalistic, non-pragmatic way to sustain the distinction between how an item functions and *the* function of the item?” (2). Teleological explanations play a legitimate role in biology if there is a (potentially) meaningful difference between an item's present and its actual functions. For those who answer in the affirmative, many use Darwinian natural selection as a grounding mechanism for functions. This is an attractive strategy in the philosophy of mind because it is thus possible to account for the nature of misrepresentations by branding it as possessing negative survival value (3). As touched upon in §2.2, Ruth Millikan's concept of proper function has been the most influential attempt in this regard. Ariew and Perlman note the potential for Developmental Systems Theory to offer an alternative grounding, a potential that I articulate by using the enactivist framework discussed in the previous section in conjunction with Mark Bickhard's interactive model of representation. To set the explanatory target for this alternative grounding of functions, I first examine Ariew's (2002) discussion of Aristotelian teleology.

Ariew (2002) recognizes two basic types of teleology, both of which are present within Aristotelian philosophy. There is (i) agency-centered teleology and (ii) teleology pertaining to natural organisms (9). While both of these types are present within Aristotle, the former unlike

the latter is also found in Plato and so Ariew labels (i) Platonic and (ii) Aristotelian. Agency-centered teleology roots functions in an agent's conscious intentions; when scaled to the level of the universe, nature appears as an artifact created by a very powerful agent (11). The second type of teleology, by contrast, corresponds to an inner principle of change within organisms. One form of this teleology corresponds to developmental processes that occur for the sake of an organism's self-preservation. Ariew considers the example of a plant's roots growing downwards: a plant is not consciously aware that it is good for its roots to grow downwards but instead plants whose roots don't grow downwards simply do not flourish (9). Rather than the product of conscious intention, the second type of teleology is immanent to organisms, activating potentialities specific to a species rather than a universal form. As a result, immanent teleology doesn't imply striving for what is best; this *telos* only implies relative usefulness. So, for example, just because a plant's roots grow downwards does not mean the development is ideal; it is possible that a different root system would enable the plant to flourish *even more* (12). Hence, immanent teleology implies a gradation of norms (i.e., worse-to-better) rather than an all-or-nothing norm of best versus the rest. As discussed in the last section, with regard to sense-making, having graded norms is essential to accounting for learning. Ariew sticks to the more abstract notion of development in order to argue that Aristotle's basic notion of immanent teleology is integral to contemporary biology.

Ariew claims that Darwin's principle of natural selection presupposes immanent teleology. More specifically, natural selection presupposes organisms' striving for self-preservation, which corresponds to the graded norms of development (23). The basic issue is that "blind necessity" alone is insufficient to account for organismal development, adaptation, and the like (29). To bolster this reading, Ariew cites correspondence in which Darwin thanks Asa Gray

for his remark that evolution restores a role for teleology (23). Ariew's larger argumentative strategy is to show that grounding biological functions by appealing to natural selection presupposes Aristotle's immanent teleology. To evaluate this strategy, it is worth returning to Millikan's account of proper functions.

Millikan (1984) introduced her concept of proper functions to account for what underpins judgments of defectiveness. Hence, the concept of proper function applies to anything about which it makes sense to talk of its dysfunction, whether mental representations, toasters, or hearts (Millikan 2002, 116). With regard to cognitive realism, it is important to be able to claim that instances in which a mind misrepresents external reality are dysfunctional cases. Millikan (1984) believes that this requires appealing to the principle of natural selection for two reasons. First, cognitive capacities are the products of evolution and they have survival value (93). Second, and more decisively, Millikan thinks that defectiveness presupposes a historical dimension as a basis for evaluating a given instance. The way something "should" function in this context refers to what has been selected for in an evolutionary lineage (94). There are various issues with this account, of which I will draw upon two objections from Robert Cummins's since his objections will set up an alternative view of teleonomy.

Expressing a common sentiment, Cummins (2002) views neo-teleology derisively. He levels two major objections to Millikan's general project. First and most fundamental, natural selection does not create functions but instead only shapes what is already present (163). But, given this criticism, it is possible to claim that identifying the function of a trait identifies the dimension of performance relevant to assessing its adaptiveness (167). Cummins discounts this line of reasoning on the grounds that it oversimplifies actual selection histories. An adaptation need not correspond to improving a function—perhaps the adaptation nets an increase in

efficiency or robustness or contributes to another function (168). The key point is that there is no direct correspondence between functions and adaptations. It is possible to read Cummins as indicating the need for supplementing Millikan's position with immanent teleology (Ariew's position) or as undermining any substantive notion of teleology. One influential articulation of this latter position is Colin Pittendrigh's concept of teleonomy.

Pittendrigh (1958) coins the term teleonomy to discuss goal-directedness without implying that the goal is causally operant in its own realization.⁶⁵ There have been many explications of teleonomy. Ernst Mayr (1974) defines a teleonomic process as one controlled by a program, wherein a program is any coded information (101). Ernest Nagel (1977), by contrast, offers a systems-property account in order to isolate specific functional aspects of complex systems. Common to both approaches and various others is reducing teleology to a heuristic. For Mayr, functional ascriptions aid in identifying behavioral programs, while for Nagel they simply reduce to generic causal properties. For Mayr and Nagel, the aim is to identify the deterministic mechanisms that underpin an organism.⁶⁶ Enactivism, however, shows that teleology is in fact a special form of causation, the recognition of which undermines the teleonomy view, offers an alternative account of defectiveness, and improves on Ariew's formulation of immanent teleology.

Weber and Varela's (2002) central insight was in recognizing the teleological implications of operational closure. *Pace* Pittendrigh, operational closure identifies how "ends" play a role in circular causation without implying "backwards" causation. Di Paolo's (2005) concept of adaptivity addresses Millikan's question of defectiveness: dysfunction is whatever

⁶⁵ Similarly, Ariew (2002, 7) notes that most theorists have dropped "teleology" in favor of "functions" simply in order to avoid the metaphysical associations of the former term.

⁶⁶ Teleonomy thus corresponds to the interpretation of Kantian purposiveness as an artifact of the limits of human cognition, in which heuristics partially reveal the actual causal mechanisms.

moves an organism further away from its viability conditions. By thus relieving natural selection of the burden of grounding all functional norms, Ariew's argumentative strategy appears somewhat myopic. While Ariew is right that natural selection presupposes an immanent teleology, the central issue concerns adequately accounting for the organism as a unitary locus of activity as opposed to a coincidence of forces. Providing such an account, as is enactivism's goal, shows that intrinsic teleology is not merely derivative upon nor even solely geared towards natural selection. This is the point at which enactivism parts ways with Ariew's account.

Ariew's understanding of immanent teleology is at variance with enactivism regarding purposiveness and intentionality. In contrast to enactivism, Ariew claims that immanent teleology is non-purposive and non-intentional, but this is because he considers purposiveness to imply conscious design and intentionality as implying representational aboutness. Last section's discussion of sense-making showed the rudiments of how enactivism departs from these claims. This departure is important for it enables enactivism to avoid appealing to Aristotle's notion of "species form," which may be more tractable than universal forms yet is still deeply problematic.⁶⁷ To buttress enactivism's conception of intentionality and purposiveness, I now turn to Bickhard's interactive model of representation.

Bickhard's (1996, 2000, 2002) interactive model of representation buttresses Di Paolo's (2005) sketch of adaptivity and brings into focus the issues of perceptual salience and learning. Bickhard's interactivist model complements enactivism, approaching many of the same issues from the context of psychology. Offering an alternative account of representation is central to reconceptualizing intentionality, as evidenced by Bickhard's (1996) shifting the question of representation from the usual one of content to the act of representing itself. Representational

⁶⁷ For a notable though ultimately flawed attempt to ground the concept of species, see Hull (1990).

content is a product of a system's functional usage of a perceptual signal and thereby derivative upon the act of representing.

Representations, first and foremost, serve to differentiate between different interactive contacts. Bickhard (1996) writes, “differing environments may leave that (sub)system [whatever initiates the interaction] in differing final internal states or conditions when the interaction is ‘completed.’ Such possible internal final states, then, will serve to *differentiate* possible environments” (60; original emphasis). The term differentiation expresses the semantic opaqueness between environment and interacting system. The final state reached does not bear informational content from environment to organism. Instead, the only information that perceptual differentiators contain about the environment is that it caused the given final state to be reached; the internal state thereby only *implicitly* defines the environment (61).

Differentiation and implicit definition are the two foundations of interactive representation.

The concept of implicit definition dovetails with enactivist adaptivity. An organism does not need to represent its environment as such but rather only what has a differential effect on its viability state. Thus, built directly into implicit definition is perceptual salience—the organism only perceives what is of differential significance. Because this account of salience treats the act of representing as a primitive, it presupposes goal-directedness. Bickhard's interactive model clarifies the nature of primitive goal-directedness, showing that it is itself non-representational. Rather than representations, at the most primitive level, goals are functional switches with a success result (e.g., further processing/anticipation of another interaction) and a failure result (e.g., a trial and error interactive process) (62). To illustrate the difference Bickhard reinterprets one of Fodor's (1990) examples concerning frogs. For it to eat, a frog does not represent, say, a fly or a worm. Instead, the fly and worm appear as two potential types of interaction: “tongue-

flick-*at-a-point*” and “tongue-flick-*between-two-points*,” respectively (64; original emphasis). The frog thus anticipates specific kinds of eating opportunities, the salient issue being how to flick the tongue rather than what the tongue flicks at.

Representation (as implicit definition) is inherently opaque or coarse-grained, though capable of becoming more fine-grained with increasing interactive sophistication. Frogs flick their tongues rather indiscriminately; fortunately, they live in a world where many of those flicking opportunities bear edible fruit. If massive numbers of teenagers or philosophers took it upon themselves to shoot BBs across their fronts, this would change the frogs’ world dramatically. Perhaps frogs would become able to discriminate between metal projectiles and buzzing things, maybe their diet would shift more towards crawling things, or maybe they would die off. Regardless of this (thankfully) hypothetical situation, the key point is that what the frog represents is inextricable from its interactive world because its representations are the product of its actual interactions. Ingestion of metal spheres *would* constitute error and *could* trigger learning new behavioral patterns, but for the vast majority of frogs who have been spared the machinations of philosophers of mind, BBs *don’t* count as errors. Bickhard’s rendition of the frog example thus illustrates how implicit definition entails error-guided learning.

For the interactionist model, misrepresentation is no longer an aberrant phenomenon in need of explanation. Misrepresentation is inherent to implicit definition and is the basis for detecting and then learning from error. There’s a profound sense in which you can’t be right without having first been wrong. Error prompts refining one’s interactions and implicit definitions; but *refining* implicit definitions does not make them bearers of external content. They are still subordinate to the self-maintaining network of which they are a part and whose self-identity is the motivating constraint from which the anticipatory interaction issues. Fitch (2008)

highlights the serial nature of such learning and an attendant need for tagging, a proposal that I buttress in Ch. 4 by further articulating Bickhard's (2002) levels of knowing model. But for now, in preparation for discussing enactivism's specifically social implications, it is enough to clarify why the individual as a self-maintenant locus of activity does not entail egoism. Towards this end, Searle's account of intrinsic normativity serves as a useful contrast.

Searle's (1990, 1995) account of intrinsic normativity ties together the issues of adaptationism, egoism, and intentionality discussed in this section. His identification of intentionality with representational aboutness implies that individuals have a more determinate relationship to the world than is actually the case. This, in turn, dovetails with viewing individuals as motivated by egoistic self-interest: an individual's desire appears in terms of what *thing(s)* she wants to acquire from the world. The world is comparable to a shopping list.⁶⁸ The essential difference between egoism and enactivist self-maintenance is that, in the former case, the default position is a static individual as opposed to the continuous interactive coping of the latter. From the position of stasis, according to egoism, the individual chooses to interact with her environment. This creates the false problematic that Searle's we-intentions aim to solve: we-intentions are meant to flip the switch from the isolated self to self "as member of a community."

Searle's normativity thus lacks the gradation essential to any learning process: someone has committed to a we-intention or they haven't. In place of a first-personal learning process, the broad stroke of natural selection is meant to ensure that those for whom we-intentions do not flip on are selected against. Furthermore, making the we-intention commitment tacit and then trying to cash it out in terms of stronger and weaker only has the regrettable consequence of turning

⁶⁸ As in much else, Jerry Fodor is instructive: "What philosophers call 'linguistic deference' is actually *the use of experts as instruments*....Epistemology...is mostly about money; e.g., about what I'm prepared to pay to insure that one or other of my beliefs is true" (1995, 36-37). I take Fodor's position as a *reductio* of much of epistemology.

normativity into an ethereal mystery rather than an integral part of a learning process. It is this latter depiction of the social that De Jaegher and Di Paolo's concept of participatory sense-making offers.

In the present section, I have addressed two apparent tensions between my commitment to enactivism and my previous critiques of adaptationism and Searle's intrinsic normativity. In doing so, I situated enactivism's intrinsic teleology within the larger context of the philosophy of science and buttressed Di Paolo's (2005) sketch of adaptivity with Bickhard's interactive model of representation. The opaqueness of goal-directedness, as conceived by implicit definition, together with sense-making's emphasis on regulated movement opens up new ways for examining social interaction.

3.3 An Enactivist Approach to the Social: Participatory Sense-Making

Enactivism makes possible an empirically grounded framework for analyzing and understanding intersubjectivity that transcends the individualist-collectivist dichotomy. Hanne De Jaegher and Ezequiel Di Paolo tap into this potential with the concept of participatory sense-making and its central theme of coordination. Coordinated movement appears as an intrinsic dynamic endogenous to interaction rather than an output of an individual cognitive mechanism, which in the following is an important point of comparison between the enactivist approach and the translation view. There are two rough groupings of essays that I consider in explicating participatory sense-making: the first is devoted to the basic formulation of the concept and the

second, centering on De Jaegher's efforts, attempts to develop the concept beyond dyadic face-to-face interaction.⁶⁹

De Jaegher and Di Paolo's introductory (2007) article on participatory sense-making has two general concerns. First, the authors needed to show that enactivism, despite previously neglecting the topic, has valuable insights for social interaction. Toward this end, they situate their concept in relation to the five core enactivist themes of autonomy, sense-making, embodiment, emergence, and experience (486-88). The second task is to show that participatory sense-making is more explanatorily parsimonious and powerful and better captures our everyday experience of social interaction than what is commonly found in the philosophy of mind. De Jaegher and Di Paolo diagnose philosophy of mind's treatment of social interaction as the Rear Window approach.

The Rear Window approach refers especially to the Theory-Theory versus Simulation-Theory debate that has dominated the philosophy of mind's concern for the social. On the Rear Window approach, social understanding stems from "figuring out someone else's intentions out of our individual observations of them" (489). Theory Theory and Simulation Theory differ over whether the observations of another person are fed into an inferential or simulation process, but both share the basic picture of observations as the basis for imputing an intention to others. In presenting an enactivist approach, De Jaegher and Di Paolo want their reader to feel just how myopic this picture is: my social life does not occur through a pair of binoculars, nor did Jimmy Stewart's.

In the initial article, De Jaegher and Di Paolo (2007) state they are focusing on "the face-to-face encounters of everyday life" (486). In order to explicate the concept of participatory

⁶⁹ The first group consists of De Jaegher and Di Paolo (2007, 2008), with Rohde (2010), and with Gallagher (2010). The second group includes Fuchs and De Jaegher (2009), McGann and De Jaegher (2009), De Jaegher and Froese (2009), and De Jaegher (2013).

sense-making at this level, I primarily draw upon De Jaegher and Di Paolo (2008) since this article focuses on the two concepts most relevant to participatory sense-making (i.e., autonomy and sense-making) and its central theme of coordination.

As noted in §3.1, biological autonomy expresses cellular autopoiesis in more general terms. An autonomous system consists of “several processes that actively generate and sustain an identity under precarious circumstances” (35). Precarious conditions are simply situations in which a process would tend to “run down or extinguish” without being part of an organized network, which in the case of biological autonomy is self-organized. Participatory sense-making identifies a novel kind of autonomy emergent in social contexts, namely the relational dynamics of interaction (De Jaegher, Di Paolo, and Gallagher 2010, 443). Autonomy is also the foundation of sense-making, the latter specifically entailing the additional capacity for spontaneous movement and thus the presence of agency (§3.1). On the enactivist picture, motility remains the center of mental activity for complex animals, such as humans (2008, 36). Given motility’s primitive and abiding significance, anything that shapes an organism’s movement is integral to its cognition.

Participatory sense-making examines social interaction in terms of how interactants ubiquitously influence each other’s movements via coordinating factors. Two systems are correlated if their behavior coheres beyond what is expected of them given each system’s capabilities (2007, 490). Coordination simply refers to cases in which a correlation is non-accidental (490). One of the chief tasks of the 2007 paper is to show that coordination often only involves simple mechanisms of coupling. Towards this end, the paper lists a wide range of coupling mechanisms, including the synchronized light flashing of fire flies and the synchronized pendulum swinging of two people in view of each other (490). In short, because it

is realized by countless mechanisms, coordination is so pervasive in nature that it is hard to avoid. The larger implication for De Jaegher and Di Paolo is that coordination does not depend on mental mechanisms.

Although it is only one form of coordination, synchronization receives the most attention because it conveys the significance of timing to interaction. Synchronization need only be relative (i.e., phase attraction between two systems) and, importantly, can involve very different behaviors on the part of the two systems (2007, 491). For example, as noted by De Jaegher and Di Paolo, Kelso (1995) considers an adult and child walking alongside each other. With little-to-no mental effort, the two will adjust their gait to roughly match the other's pace. This can be realized in multiple ways (e.g., adjusting the length or frequency of strides) and, importantly, it entails opposing compensations on the parts of the child and adult, which is at variance with the translation view's emphasis on contagious mirroring. Hence, it is helpful here to examine some of the fundamental differences between the enactivist approach to the social and the translation view.

For the enactivist approach, there is no substantial foundation for social interaction. Coordination is an abstract notion that refers to any non-accidental correlation between two systems. Mirroring is one particular form of coordination—salient within certain contexts, such as apprentice training—rather than being the fundamental basis of *all* epistemically significant social interaction. Avoiding a substantial foundation enables participatory sense-making to capture a wider array of epistemic significance. In Ch. 4, for example, I argue that one of the most significant aspects of social interaction is the *sui generis* creation of epistemic goals, which the translation view obscures by positing functional equivalence as the most basic epistemic goal. The positing of any general epistemic standard results in reducing social interaction to a means

for accomplishing that end (e.g., acquiring a tacit skill). And it is this pervasive manipulationism that I directly scrutinize in Ch. 5.

It should be noted, however, that De Jaegher and Di Paolo's early discussions of coordination do not sound all that different from the translation view's model of tacit learning. There are four types of coordination mentioned in the two introductory articles—synchronization, mirroring, imitation, and anticipation.⁷⁰ De Jaegher and Di Paolo's interest in these four types is connected with how they facilitate fluid, seamless social interaction. If coordination were synonymous with smooth social interaction, then it would be tempting to read it as merely broadening the enabling factors that underpin tacit learning, with tacit learning still being defined by opaque mirroring. But the authors are clear that their list of coordinating factors is not exhaustive; furthermore, I will be arguing that their emphasis on fluidity is a partly an artifact of focusing on two-person face-to-face interaction, which becomes problematic if directly extrapolated to larger-scale institutional contexts. Broadening coordination beyond factors conducive to fluidity will further differentiate the enactivist approach from the translation view.

While the broad range of coordination factors provides multiple avenues for understanding social interaction, for some social theorists this becomes a liability: the abstract concept opens up *too* many avenues, so many that social interaction becomes indistinguishable from other kinds of coupling. After all, among the original examples of coordination is the synchronized oscillation of clocks, which is certainly not a *social* phenomenon. To address this concern, De Jaegher and Di Paolo (2008) define social interaction as:

The regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the

⁷⁰ De Jaegher and Di Paolo are not explicitly clear on the difference between mirroring and imitation. Presumably the former specifically designates mirror neurons, whose exaggerated use in social theory De Jaegher and Di Paolo (2008) directly criticize.

autonomy of the agents involved (though the latter's scope may be augmented or reduced).
(39)

So while interaction itself emerges as an autonomous process, social interaction requires the agents to maintain at least some individual autonomy. The qualification that social interaction may augment or reduce the autonomy of the involved agents—which can be interpreted as implying that autonomy falls on a one-dimensional spectrum—signals what becomes the central theme of De Jaegher's extension of participatory sense-making.

As a contrast to *social* interaction, De Jaegher and Di Paolo offer the example of heat transfer within a crowd of people waiting for a bus. This sort of coupling is explicable strictly in terms of brute physical determination and does not help hold the crowd together because it is incidental to the situation (the prospect of a bus holds everyone together). It should be noted, however, that on closer inspection the difference between social interaction and the intra-crowd heat transfer is not necessarily cut and dried. Perhaps it is a cold windy day and people want to stay warm, so they huddle close together rather than spreading out along the sidewalk. In this scenario, the desire to stay warm may initiate conversations and a sense of camaraderie (or disgruntlement) that would not have otherwise occurred. This is not meant as a rebuke to De Jaegher and Di Paolo's distinction but rather a cautionary note: what counts as a *social* coordinating factor depends on the intimate details of particular situations including the actual individuals involved.

In this respect, I see the enactivist approach as advancing the translation view's emphasis on the importance of improvisation within social interaction: the role of spontaneity only increases when social interaction is not anchored to a particular subpersonal neuronal process, such as mirroring. Participatory sense-making enables analyzing the interactive process as such, in all of its haphazard and idiosyncratic glory (and terror). Indeed, capturing the spontaneity of

social interaction is one of the concept's chief selling points, since much of social theory suffers from a staleness that results from treating the idiosyncrasies of interaction as merely incidental to the actual cognitive processes. The Rear Window approach exemplifies this blandness by taking another's behavior as merely data for judging his or her individual intentions.

The Rear Window approach is in part an artifact of the previously discussed input-receiver picture of cognition. This picture places pressure on reducing any cognitively salient factor to an individual mechanism, since cognition is meant to be precisely what the individual does to the input, which then informs the individual's subsequent behavioral output. Discarding the input-receiver picture allows participatory sense-making to avoid the proliferation of homunculi dressed up as subpersonal neuronal processes. This indicates what is participatory sense-making's most rhetorically persuasive selling point—namely, its greater explanatory parsimony. When compared to the enactivist approach to the social, the Theory-Theory/Simulation-Theory debate, looks like a Ptolemaic exercise in epicycle adjustment rather than a conceptual clarification of our best science. Participatory sense-making is accordingly first and foremost an intervention in the philosophy of mind.

For De Jaegher and Di Paolo, however, the most substantial appeal of participatory sense-making is its usefulness for informing empirical research. The (2010) article that they coauthored with Shaun Gallagher is a primer for how to employ the enactivist approach to social interaction within a controlled experimental setting. The article canvasses studies that share an affinity with the enactivist approach, cites a case in which the framework has already informed research, and offers a set of questions for future research. The authors thus address what they consider to be the most pressing issue facing the concept of participatory sense-making; to wit,

showing that it has a well-defined and valuable research use. The second most important issue is whether the approach can be extended beyond two-person face-to-face interaction.

From the outset, De Jaegher and Di Paolo believed that participatory sense-making could directly inform larger-scale forms of interaction. In (2007), they cite works in sociology that take face-to-face encounters as the basic frame of analysis (e.g., Goffman 1983, Sacks 1992), while in (2008) they pitch participatory sense-making as a corrective to sociological theories that focus exclusively on collective processes. The (2008) article thus identifies the converse error to philosophy of mind's Rear Window approach, positioning participatory sense-making as a corrective to the excesses of both overly individualist cognitive science and overly collectivist sociology, the two errors having created a spurious gulf between the fields. De Jaegher and Di Paolo see participatory sense-making as balancing and thus bringing into focus the respective roles of individual and interactive factors. Yet the manner in which participatory sense-making has been extended beyond the two-person context—spearheaded by De Jaegher—is, I think, founded on some troublesome premises. De Jaegher's efforts center on the spectrum of social coordination, an abstraction that appears in the introductory articles but which comes to take on a much greater significance in De Jaegher's subsequent works.

As presented in (2007, 2008), the spectrum of participation offers some suggestive insights on two-person interaction. The spectrum is based on degrees of coordination: on one side of the spectrum lies mere “orientation” and, on the other, “joint sense-making.” In the latter case, it may be very difficult or even impossible to tease out each individual's distinct contributions. Compare this to an example of orientation, in which I happen upon a person looking for something and ostend in order to direct his or her search (2008, 42). In this case, it is

quite easy to delineate each individual's contributions: one is the desirous searcher and the other a helpful pointer; importantly, the searcher's desire would exist irrespective of my presence.

De Jaegher, Di Paolo, and Gallagher (2010) tentatively propose that the spectrum is a general indicator of the affective character of interaction. The proposal has some empirical support within the psychotherapeutic context: there is a correlation between, on the one hand, high bodily coordination between therapist and patient and, on the other, a patient's positive assessment of the therapy session (2010, 442). This link is buttressed by Fuchs and De Jaegher's (2009) pairing of coordination with phenomenology's concept of mutual incorporation. Fuchs and De Jaegher claim that the reciprocal shaping of interactants' movements (i.e., coordination) causes and is caused by a decentering of subjective experience for each person (i.e., mutual incorporation). While De Jaegher and Fuchs are right to link coordination and mutual incorporation, it is a mistake to identify a general correspondence between high bodily coordination and positive affective states. Phenomenology constrains cognitive science in part because first-personal experience is stubbornly idiosyncratic.⁷¹ While high bodily coordination might correspond to positive emotional states in the therapeutic setting, this does not generalize to all social settings.

McGann and De Jaegher (2009) recast the correspondence between coordination and first-personal experience as a dichotomy between fluid and rigid interactions. For the remainder of the section, I focus on problematizing these one-dimensional spectra of sociality, especially as it is expressed by the fluid-rigid dichotomy. The issues raised here reappear in starker terms in Steiner and Stewart's (2009) account of the social domain and participatory sense-making's role therein, which I scrutinize at length in §3.5.

⁷¹ For more background on the relationship between cognitive science and phenomenology, see Merleau-Ponty ([1945] 2012), Gallagher (1997), and Thompson (2007).

The main goal of McGann and De Jaegher (2009) is to give an account of social perception in terms of social contingencies. One hope is to reconcile the Noë-O'Regan sensorimotor contingency theory with the broader Varela strand of enactivism; participatory sense-making's emphasis on operational terms makes it an attractive point of contact. In this context, fluidity appears as sensitivity to social contingencies. That is to say, being socially skilled entails the ability to improvise based on the changing emotional states of ourselves and others (427-28). Lack of social skill, by contrast, implies rigidity and predictability, an inability to engage one's particular audience. In making this connection with Noë and O'Regan's sensorimotor contingency theory, I think that McGann and De Jaegher erode some of participatory sense-making's value. Indeed, the sensorimotor contingency theory reflects the guiding intuition behind the various one-dimensional spectra of greater or lesser sociality: interpreting greater sociality in terms of unpredictability dovetails with Noë and O'Regan's privileging of quantifiable motility over first-personal experience.

De Jaegher (2013) bolsters the fluid-rigid dichotomy by articulating it in terms of political theory. Drawing upon Gilligan and Richards (2009), De Jaegher frames the dichotomy in terms of horizontal democratic forms of association in contrast to top-down patriarchal social organizations. The connection is ambitious, representing an attempt to show participatory sense-making's relevance to analyzing entire societies. This extension is predicated on yet another spectrum, namely a spectrum of power symmetry. De Jaegher writes, "When interacting with another person, or with an institution, one or other partner may be more or less 'dominant,' more or less influential. In such asymmetric relations or interactions, the degree of influence that each partner has is different" (23).

De Jaegher's understanding of dominance as a matter of "more or less" corresponds to the earlier interpretation of autonomy as one-dimensional. Accordingly, this symmetry of power roles directly corresponds to the spectrum of coordination. On the less social end, highly asymmetric institutionalized interaction corresponds to orientation, since it is easy to delineate the differing contributions by the authority and the subordinate. On the other, De Jaegher offers the example of a lawyer-client relationship as a case of greater symmetry/joint sense-making: the lawyer and client work together to interpret the law to their advantage. But even in this case of a messier, horizontal type of interaction, institutions still appear as rigid things—as an inert resource for individuals to draw upon—since the laws and courts are simply means to advance one's interests. Institutions appear as inert objects because they are identified with rules, similar to how Searle interpreted institutions as founded upon we-intention statements. It is this reified understanding of institutions that I challenge in §3.5 and Ch. 4. But there is a more immediate problem with the fluid-rigid dichotomy.

The basic issue with the dichotomy between fluid and rigid interaction is best illustrated by one of the studies repeatedly cited in its support. Granic (2002) uses Dynamic Systems Theory to model parent-child relationships. Granic is specifically interested in how patterns of hostility between parents and children can emerge despite each party's amicable intentions, a phenomenon she attributes to behavioral attractors (273). In both McGann and De Jaegher (2009) and De Jaegher (2013), Granic's cases of hostile parent-child behavioral patterns are glossed as instances of rigid patterns of social interaction. That is, since the hostility occurs *against* each individual's intentions—thereby reducing the autonomy of each—the pattern is rigid in the sense that it predictably occurs irrespective of the circumstances of the encounter. But Granic herself does not think of hostile and other negatively valenced interactions as being

more “rigid” than positive ones. Behavioral attractors *stabilize* both positively and negatively valenced patterns of interactions. From a therapeutic perspective, Granic is most interested in identifying and figuring out how to destabilize negative patterns, but this does not imply that negative patterns of interaction possess any greater stability or “rigidity” than positive ones.⁷²

While the motile profiles described by De Jaegher’s various spectrums are relevant and may in some circumstances strongly correspond to particular affective states (e.g., in the psychotherapeutic context), they cannot serve as context-independent indicators. This is because sociality simply does not consist of a one-dimensional spectrum of greater or less. The quantifiability of this kind of spectrum is attractive from the perspective of a controlled experimental setting but ultimately obscures important dynamics within the larger-scale and longer-term institutional contexts that De Jaegher aspires to examine.

In Ch. 4, I detail novel kinds of sociality that can emerge within the constraints of hierarchical organizations. Moreover, these emergent social organizations can enable epistemic achievements that would be otherwise impossible. But before turning to this matter, I first want to consider a critique and also a refinement of participatory sense-making. In §3.5 I examine Steiner and Stewart’s (2009) attempt to situate participatory sense-making within sociology. I think Steiner and Stewart’s revision brings into starker relief the issues with the one-dimensional spectra discussed here. In §3.4 I consider Andy Clark’s (2008) critique of enactivism as a whole.

⁷² Consider the following, not terribly far-fetched scenarios: a friendship of mine may center on a shared hobby or interest, let’s say football. Whenever I see the other person, whether at work, the grocery store, or the post office, we chat about the latest football news and commiserate over the latest setbacks. The person, him- or herself, becomes an attractor for the topic of football. Importantly, this may be a positive or negative relationship. If he or she is just a casual acquaintance, then there’s nothing necessarily wrong with our predictable pattern. And this is so even if we are not particularly mindful of each other’s affective states, since such mindfulness may be largely superfluous insofar as our encounters stick to the topic. Alternatively, there may arise special occasions when the pattern becomes objectionable: maybe we encounter each other at a mutual acquaintance’s funeral and can’t help ourselves. Yet, to complicate this latter scenario, if the mutual acquaintance also shared a passion for football, raising the topic may actually be an appropriate form of coping. All of this is by way of saying—indeed, insisting—that the significance of a social interaction’s rigidity or flexibility depends on the particular circumstances and the idiosyncrasies of the people involved.

Clark argues that enactivism is simply a species of “wide functionalism.” I argue that Clark’s critique holds water only if enactivism is identified with Noë and O’Regan’s sensorimotor contingency theory. But, contrary to the tenor of McGann and De Jaegher (2009), the lesson I draw from Clark’s critique is to sharply distinguish the sensorimotor theory from enactivism as a whole and participatory sense-making in particular.

3.4 Extended Functionalism?

Given that mind is embodied and environmentally embedded, Andy Clark (2008) recognizes two defining stories that divide how these facts inform cognitive science. First, the “Special Contribution Story” claims that “specific features of the body (and perhaps the world) make a persistent, non-trivial, and in some sense special contribution to our mental states and processes” (39). This is contrasted with the “Larger Mechanism Story,” which claims that “aspects of body and world can, at times, be proper parts of larger mechanisms whose states and overall operating profile determine (or minimally, help determine) our mental states and properties” (39). As is suggested by characterizing the stories in terms of operating profiles, Clark believes that the body’s epistemic significance should be understood in terms of the functional role it plays in cognition. Accordingly, Clark endorses the latter story, in which the basic units of analysis are functional mechanisms, of which the body is just one among many. I will detail and scrutinize Clark’s position in Ch. 5 since it occupies a peculiar place in regards to social cognition and epistemology.⁷³

Presently, I am interested in how Clark (2008) mutes the enactivist approach by reducing it to Noë and O’Regan’s sensorimotor contingency theory. Clark summarizes enactivism as

⁷³ Even though Clark himself doesn’t consider social interaction to be of any special cognitive significance *whatsoever*, both Gallagher (2009, 2011, 2013) and Sutton (2010) employ his Extended Mind hypothesis to argue for the distinctive epistemic import of the social.

follows: “The central claim is that perception is nothing other than implicit knowledge of so-called ‘sensorimotor contingencies’; that is to say, knowledge of the effects of movement on sensory stimulation” (41). Enactivism is thus reduced to the claim that perception depends on our implicit knowledge of the likely effects of our body’s movements. In other words, our sensory organs continuously factor in the motion of the body of which they happen to be a part and thus the body contingently plays a role in cognition. Having pinned the enactivist position to a particular claim about the body’s functional role, Clark opens the door to his larger functionalist framework: “For now, I simply note that from the fact that (as seems highly likely) our human experience really does depend in part on many idiosyncratic aspects of our embodiment, it does not follow that only a creature thus embodied could have those very experiences” (42). If the body plays a specific functional role, as Clark interprets enactivism, then it is theoretically even if not yet practically possible to design an artifact that will serve the same function.

Clark’s gloss of enactivism misses the importance of intrinsic teleology. It is not simply that sensory organs happen to anticipate their body’s movement but, more fundamentally, perception is tied to a nonrepresentational normative directedness (§3.2). This normative directedness, as explicated in Bickhard’s interactive model of representation, accounts for perceptual salience. In short, salience corresponds to what is anticipated, unless the organism encounters error, which is itself salient and affords a learning opportunity. Clark skirts the issue of agency and simply presupposes an answer to the problem of salience.

Within the functionalist framework, desire is compartmentalized from cognition. As a result, Clark’s functionalist analyses take the problem-to-be-solved as the given starting point and attempt to show how the goal is achieved. This method of analysis is tied to a version of the well-worn input-receiver picture: “For what embodied experience actually delivers as the

baseline for learning and metaphorical thought surely depends on some complex mixture of bodily form, environmental structure and (possibly innate) downstream internal processing” (54). Cognition is meant to be founded on sensations impinging upon the body. The impressions on the body correspond to what is perceptually salient, which is then passed “downstream” to neural processors. The body is thus merely a “gross physical bedrock,” serving a particular instrumental role for the more important neural processing (56). Clark leaves open the possibility that some of the neural functions may be learned and thus perhaps some of the filtering occurs in the nervous system. Overall, this schematic harkens back to the testimony view’s reduction of other people to being informational filters.

The issue of perceptual salience is raised by the input-receiver picture critique but not Hurley’s (1998) more familiar input-output sandwich critique. Directly addressing the latter, Clark writes, “Extended functionalists thus reject the image of mind as a kind of input-output sandwich with cognition as the filling.... Instead, we confront an image of cognition quite literally bleeding into the world” (49). The input-cognizing-output sandwich refers to the more specific claim that cognition only occurs within the body or skull of an individual. The input-receiver critique, by contrast, applies to Clark’s extended functionalism, and the testimony and translation views, assigning a version of the Meno Paradox to each. Simply put, if passive reception is the primitive foundation of cognition, how could an organism filter out what is perceptually salient without already knowing what it is they are perceiving. To ape the title of Clark’s (2008) article, no matter how hard you “press the flesh,” no amount of pressure will tell the organism what matters, what is salient, what should be passed along to processors “downstream.”

The issue with Clark's gloss of enactivism highlights the problem I raised in the previous section regarding De Jaegher's efforts to develop participatory sense-making beyond dyadic interaction. De Jaegher too closely aligns participatory sense-making with enactivism's narrow strand and its privileging of motility over idiosyncratic first-personal experience. Such a link, Clark's critique suggests, reduces participatory sense-making to a species of extended functionalism, an avowedly anti-social cognitive theory. In the following section, I turn to Steiner and Stewart's (2009) direct criticism of participatory sense-making. Partly because they are attempting to improve the concept, Steiner and Stewart's perspective is even more telling. Their argument for situating participatory sense-making within structural normativity obscures participatory sense-making's most insightful and radical ramifications for large-scale social contexts.

3.5 Structural Normativity?

Like Clark (2008), Pierre Steiner and John Stewart (2009) believe that enactivism is ill-equipped when it comes to higher-level cognition. But unlike Clark, the authors argue that enactivism need only be supplemented in order to address this lacuna, supplemented by recognizing the role of structural normativity. Toward this end, Steiner and Stewart propose a "strong" definition of social interaction: "a truly social domain ... is always defined by a set of structural norms. ... [in which the structural norms are] impersonal, culturally inherited and to a large extent independent of the individuals" (527). A social domain is here understood to be more than simply interacting individuals; it additionally entails heteronomous submission to and inheritance of a normative order. Steiner and Stewart thus present what they consider the second pole of enactivism, from which it can address higher-level cognition.

The proposal is meant to be in part a corrective to De Jaegher and Di Paolo's concept of participatory sense-making. The specific issue with participatory sense-making, as they see it, is the concept "fails to distinguish between the mechanisms and processes which contribute to the maintenance of social interactions, and the constitution of the domain as properly social in the first place" (544). The claim is that the coordinating mechanisms identified by participatory sense-making only help to sustain social interactions as opposed to constituting them. As evidenced by the strong definition above, constitution of a *social* interaction presupposes structural norms.

The appeal to structural norms finds ample traction in much of contemporary philosophy. The most prominent formulation is by what Chauncey Maher (2012) refers to as the "The Pittsburgh School," and its central theme of the "space of reasons" (Sellars 1956). There are two defining characteristics that link Steiner and Stewart's account of structural norms to the Pittsburgh School: (1) both endorse human exceptionalism in which social interaction between persons entails a qualitative leap from other organisms' intra-species interactions and (2) human language is the focal point and source of humans' distinctive social normativity. The weight placed on human language in (2) lends the strongest support for (1). The exceptional human world is analogous to and enabled by acquiring the conventions of a language. Additionally, (2) is often motivated by a distinction between regulative and constitutive rules, in which human language is understood to be a precondition for the creation of rules necessary to uniquely human behavior. Language, according to this line of reasoning, is thus constitutive of uniquely human behavior. Consider the distinction between regulative and constitutive rules—which underpins

Steiner and Stewart's criticism of participatory sense-making above—an optional but prominent third characteristic of the structural normativity position.⁷⁴

In concert with (1) and (2), Steiner and Stewart argue that there is a “gulf” between (human) language and animal communication because the former involves a reflexive element (541). Stewart (2010, 15) further clarifies the putative gap between language and animal communication, describing animal communication as “stereotyped reactions” to signals exhibited by “all normal members of the species.”⁷⁵ Language, by contrast, is not stereotyped because of its combinatorial power. Due to its exponentially greater expressive power and thereby greater potential for error, Stewart argues that the combinatorial mechanisms of language must be supplemented with a mutual intention to communicate that fosters the repair of meanings (16).⁷⁶ The mutual intention to communicate noted by Stewart corresponds to the strong definition of the social in Steiner and Stewart (2009). Hence, Steiner and Stewart argue for human exceptionalism based on the unique properties of human language, specifically its combinatorial power and the expressly social domain generated by the mutual intention to communicate.

Adopting a label from Turner (2010), I will refer to theories committed to (1) and (2) as instances of “normativism” in order to contrast it from the concept of normativity that I have

⁷⁴ As discussed in §2.2, the distinction between regulative and constitutive rules also figures prominently in Searle (1969, 1995). It was first articulated by Rawls (1955). Haugeland (1998) provides a more recent perspective on the distinction's significance for the philosophy of mind.

⁷⁵ Tellingly, Stewart (2010, 17) grants a modicum of linguistic normativity to primates who are capable of sign-language and leaves it as an open question whether domesticated cats and dogs possess a rudimentary form as well. In the following, I argue that this places too much weight on the nature of language taken by itself and has the effect of turning social normativity into a ghostly alter universe that haunts the physical one.

⁷⁶ It's worth noting that Stewart (2010, 24) also sketches more of the putative evolutionary progression from stereotyped communication to language: on this sketch, the earliest written forms of language were lists of persons and things; as the lists grew, the pictorial symbols were ordered phonetically, which was the precursor to the combinatorial power of a fully-fledged phonetic language.

endorsed in §3.2.⁷⁷ Steiner and Stewart's account of language mitigates one of the chief complaints leveled against normativism, namely that its human exceptionalism is anti-naturalistic.⁷⁸ Stewart's (2010) evolutionary sketch, in particular, suggests how the exceptional nature of human language could have emerged out of more primitive forms of communication. Furthermore, the account maintains the linguistically generated and enforced social normativity that is championed by normativism. In §2.2-2.3, I examined this type of social normativity in connection with Searle's (1995) we-intentions.

Despite their differences, Steiner and Stewart's (2009) heteronomous submission to a social domain serves the same purpose as Searle's we-intentions. Both found social interaction on an all-or-nothing commitment: for Steiner and Stewart, either someone is or isn't a part of a given social domain, depending on whether the individual has heteronomously inherited its normative order; for Searle, either someone has or hasn't formed a we-intention. Both formulations describe a special form of objectivity generated by social normativity. In Searle's baseball example from §2.3, for instance, social norms make eating a baseball qualitatively different from the mistake of swinging and missing at a baseball. Steiner and Stewart put the same point as follows, "Unlike natural laws or biological norms, it is quite possible for an agent to behave in a way that does not respect a social norm; the sanction is no more, and no less, than that the behaviour in question will not be recognised as a socially meaningful and appropriate action" (530). Language enables the creation of a normative order, such as baseball, in which many actions may be possible but would be sanctioned as wrong, such as eating a baseball. And the basic sanction is to exclude the individual from the social domain. Conceptualizing this

⁷⁷ See Turner (2010) for a comprehensive review of the various types of normativism and their general pattern of argumentation. My present task is comparatively modest: I am only sketching enough of normativism to gain a better perspective on Steiner and Stewart's proposal.

⁷⁸ See Ch. 3 of Turner (2010) for a review of other attempts to naturalize normativism. I am concerned specifically with Steiner and Stewart's attempt since they draw upon enactivism.

special form of objectivity is meant to be one of the principal appeals of normativism, since it provides a clear basis for moral and political philosophy.

But the cost of endorsing this linguistically generated objective social normativity is dear. Later in the section, I challenge Steiner and Stewart's claim that enactivism needs to be supplemented by normativism. But first, I want to examine the price of endorsing the objective social normativity. There are three issues common to the different varieties of normativism, including Steiner and Stewart's enactivist flavor, all of which have been raised by Turner (1994, 2010).

First and most general, is the puzzle of inheritance. How exactly are normative orders acquired by individuals? It is important to notice that Steiner and Stewart's evolutionary sketch does not address this issue, since it only identifies population level trends. A social theory worth its salt should be able to account for how actual interactions produce the definitive characteristics of social normativity in individuals. Or, at the very least, a social theory should not undermine the possibility of providing such an account. The standard normativist line glosses over this issue by simply appealing to the evident fact that the vast majority of humans acquire languages. With particular languages come particular cultures, customs, and so on; the acquisition of normative structures is analogous to acquiring the syntax of a specific language. The troubling tendency of this line of reasoning is to avoid teasing out the actual dynamics of inheritance, which as a result conflates descriptions of behaviors with causal explanations. Turner (1994) diagnoses this tendency in connection with the more general appeal to "social practices."

Probing the actual dynamics of language acquisition and social interaction highlights the second issue. Explanations based on objective normative structures are faced with a two-horned dilemma: either attribute characteristics to the normative structure that are outrageously complex

and thus implausible or, taking the other horn, are empirically well-supported but do not require appealing to a normative structure at all. Normativist theories side-step this issue by claiming that appeals to structural norms are necessary for giving a coherent account of social interaction.⁷⁹ Brandom's (1998) deontic scorekeeping exemplifies the former horn, implying that there is effectively one currency within linguistic communities for enforcing its norms. This scorekeeping system is meant to be obliquely intimated in the hodge-podge of beliefs one has regarding other people's credibility and the like.⁸⁰ By focusing on naturalizing normative structures, Stewart and Steiner represent an instance of the latter horn, in which normative structures are an unnecessary fifth wheel. But I will wait to substantiate this claim until I evaluate their interpretation of enactivism later in the section.

The third and most damning issue with structural norms, however, is that—contrary to their very purpose—they are *anti-social*. The impersonal nature of structural norms—from whence their special sort of objectivity is meant to derive—annihilates the details of interaction and the idiosyncrasies of individuals. As a result, structural norms dilute the very experience of social interaction. Steiner and Stewart attempt to characterize the experience of structural norms by linking it to Merleau-Ponty's ([1945] 2012) phenomenological description of anonymity. Structural norms, on this reading, exert their normative force by appearing as an anonymous other. But anonymity is a relationship that holds between people—a random stranger or the mailman is anonymous to me—whereas structural norms are objective and “independent” of individuals. More generally, cultural artifacts—whether rules of etiquette or an artwork—are not

⁷⁹ More specifically, accounts that do not appeal to normative structures are thought to be subject to an infinite regress. Turner (2010) attributes this claim to a vicious circularity, in which the need for normative regress-stoppers is smuggled into pseudo-empirical descriptions of social interaction.

⁸⁰ In this regard, Brandom's scorekeeping account shares an affinity with reductionism within the testimony view, as discussed in §1.2. In the context of distributive justice, Walzer (1983) challenges such reductionism—which he refers to as one currency views—in arguing for “complex equality.”

anonymous to me even if the person who made the artifact is. The ascription of subjective agency to an artifact is analogous to the slippage from behavioral descriptions to causal attributions noted above in connection with the first issue. By drawing attention to the fleeting yet non-trivial nature of coordinating factors, participatory sense-making strengthens Turner's suspicion that appeals to objective normative structures—structures that exist prior to and independent of particular circumstances—undercut the uniquely *social* aspects and experiences of interaction.

Having examined three general issues with normativism and noted that Steiner and Stewart's variety of normativism—despite its naturalistic credentials—fails to address them, it is worth evaluating their claim that enactivism in general and participatory sense-making in particular presupposes normativism. Thus far, I have focused on the normativist elements of their proposal, so I now turn to examining how their variety of normativism relates to enactivism.

The enactivist flavoring of Steiner and Stewart's proposal has two aspects. First and most obvious, they endorse participatory sense-making, though limiting it to single isolated interactions. Second, emergence and circular causation figure prominently in Steiner and Stewart's naturalization of higher-level cognitive abilities. Both of these aspects are most evident in Stewart (2010): the evolutionary sketch suggests how human exceptionalism emerged from more primitive precursors; regarding circular causation, Stewart also hypothesizes that written language both caused and was caused by the spatialization of consciousness, with these two factors ultimately producing fully-fledged reflexive consciousness. I consider this second point to be a salutary contribution to enactivism, even if it exaggerates the differences between human and animal communication and does not support Steiner and Stewart's larger conclusions, such

as limiting the scope of participatory sense-making. Instead, participatory sense-making's limited scope is a result of Steiner and Stewart's flawed interpretation of enactivism.

The apparent need for structural norms is predicated on an overly narrow understanding of enactivism. This narrow reading is hinted at in the passage above concerning the purported differences between social and biological norms. In the passage, Steiner and Stewart claim that social norms can be violated, whereas biological norms cannot. The basic sanction for violating a social norm is exclusion from the corresponding domain, but violating a biological norm, on this reading, is lethal. Steiner and Stewart thus interpret biological norms as reducible if not identical to autopoiesis. Accordingly, the authors elsewhere conflate enactivism with autopoiesis:

The paradigm of Enaction is rooted in a strong theoretical definition of 'life': to wit, the theory of autopoiesis. This is a key element in rendering plausible the idea that sensorimotor interactions between a living organism and its environment can sensibly be considered as 'cognitive,' because it places a constraint on these interactions that renders them meaningful. (527-28)

Conspicuously absent in this sketch of enactivism is any recognition of intrinsic teleology and Di Paolo's (2005) concomitant distinction between autopoiesis and adaptivity. Likewise, Thompson and Stapleton (2009, 24) argue that biological autonomy—in contrast to autopoiesis—is the foundation of enactivism. Biological autonomy entails adaptivity, which in turn does not merely constrain interaction but involves a primitive normative directedness.⁸¹

As a result of disregarding intrinsic teleology, a spurious division appears between brute material determination and socially normative behavior. Steiner and Stewart write, "Because of heteronomy [i.e., submission to a normative order], behaviours are not mere physical movements, mere reactions to some external stimuli or constraints; they become actions involving

⁸¹ Stewart (2010, 3-4) reiterates the conflation of enactivism with autopoiesis: "Now the key point [regarding the connection between enactivism and cognitive science] is this: what the world 'is' for the organism amounts to neither more nor less than the consequences of its actions for its sensory inputs." Because Stewart only notes how perception is linked to consequences of an organism's actions, enactivism appears reducible to Noë and O'Regan's sensorimotor contingency theory.

responsibilities” (529). The authors thus suggest that humans—*in contrast to other animals*—do not merely react to external stimuli or constraints. Here, Steiner and Stewart are at odds with the entire enactivist project: autopoiesis and biological autonomy entail that organisms do *not* merely react to external stimuli but rather ENACT their world. Normative directedness is built into the very notion of adaptivity rather than depending on the ethereal notion of linguistically constituted normativity.

Since it is only a primitive directedness, intrinsic teleology cannot by itself account for the specialized norms exemplified in large-scale knowledge communities. My proposal, however, is to use primitive normative directedness as a bootstrap for developing participatory sense-making beyond isolated dyadic interactions. Turner’s (1999) account of social institutions, as presented in §2.3, serves as a useful guide. To briefly recapitulate, Turner recognizes three general factors at play in social institutions: (1) consciously shared goals, (2) a high behavioral frequency of an activity, and (3) first-personal normativizing beliefs. The first factor alludes to the importance of explicitly written rules and other artifacts, such as a nation’s laws or a team’s playbook; the third factor refers to one individual’s beliefs about another and her willingness to impose them. The second factor involves individually habituated patterns of behavior, which are the definitive dynamic of tacit learning.

Habituation carries the brunt of the explanatory burden for Turner. Social habituation involves explicit conventions and instruction—thus explaining why certain behaviors are more frequent in one community as compared to another—but habituation is also inherently idiosyncratic. This is ensured by the ubiquitous differences in individuals’ bodies, a fact that is partially concealed by the tacit dimension of learning. That is, individuals learn how to perform roughly the same action but in very different ways, often without even realizing the differences

in their performances. Enactivism can incorporate this model of social institutions and thereby their specialized norms within a more expansive framework.

Given its rejection of the input-receiver picture, enactivism does not lean as heavily on habituation and, furthermore, reconceptualizes the process of habituation itself. With regard to the latter, Bickhard's model of error-guided learning (§3.2) offers one such reconceptualization: habituation for Bickhard, is in part the product of repeatedly successful anticipations. By thus indexing habituation to future-oriented cognition, Bickhard's model helps demystify what habits and other aspects of an individual's experiential history are relevant in a given circumstance. Indeed, enactivism suggests that an individual's experiential history is significant in more ways than just the brute weighting of neural networks. The forward-looking orientation of enactivism thus alleviates the explanatory burden of habituation while maintaining its significance.

Enactivism similarly incorporates the role of normativizing beliefs. Coordinating factors entail that individuals influence each other in more ways than just imposing beliefs on one another. Drawing attention to the inchoate, messier dynamics of social interaction pulls further away from normativism's human exceptionalism, since the formation of a belief (e.g., a normativizing belief) can be understood to presuppose linguistic normativity (and all that this entails on the normativism view) whereas coordinating factors most decidedly do not.

Participatory sense-making thus corroborates the tenor of Turner's approach even if not its details. As suggested by the incorporation of habituation and normativizing beliefs above, enactivism does not so much undermine Turner's account as reframe it within a larger suite of coordinating factors. The enactivist framework does, however, depart from the translation view in regards to the individualist-collectivist dichotomy. Whereas for Turner and the translation view, the defining characteristic of social interaction is the opaque boundaries between

individuals, the ubiquity and cognitive significance of coordination undermines the notion of there being well-defined boundaries between individuals. In Ch. 5, I focus on the issue of egoism and its manifestation in cognitive science in terms of the manipulationist presupposition. But I want to approach this issue after having discussed enactivism's implications for social theory. Having a more careful explication of community-level epistemic dynamics will help overcome the egoist presuppositions found within cognitive science and social epistemology.

My approach to social theory unsurprisingly cuts against Steiner and Stewart's use of sociology. They appeal to the "hard-won achievement of social science" in support of claiming that autopoiesis and structural norms are the two poles of enactivism (534). Specifically, they endorse Giddens' (1976, 121-23) "duality of structure," which posits human agency as both constituting and constituted by social structures such as language. In this regard, participatory sense-making merely serves to elucidate the micro-dynamics that occur within the constraints of objective normative structures. But, as suggested by my appropriation of Turner (1999), I think that participatory sense-making has more radical implications for social theory.

Participatory sense-making undermines the plausibility of there being a qualitative gap between linguistic normativity and material determination, a gap that gives an ethereal character to normativism's depiction of social interaction. Coordination occurs at multiple timescales and via a variety of mechanisms, such as linguistic signification. While there may be some heuristic value to speak of social structures as objective entities set over against individuals, and there may even be some heuristic value in talking as if the structures possessed agential powers, actually positing their existence would presuppose a stability that simply does not exist and, even more importantly, would obscure the individual idiosyncrasies at play. In contrast to normativism, I believe all forms of normativity emerge from and are only differences of degree from the

primitive sense of normativity as nonrepresentational directedness. As a result, my position shares a much closer affinity with Turner, since both prefer a firmly naturalized, non-saltationist, non-exceptionalist account of social interaction.⁸²

While participatory sense-making's implications for social theory are dramatic, they are not unprecedented. Many of enactivism's insights have been prefigured by an offshoot of sociology—neo-Kuhnian social epistemology. The present chapter has thus shown some of the far-reaching ramifications of rejecting the input-receiver picture in cognitive science. The one I have given the most attention to is overcoming the individualist-collectivist dichotomy in social theory, a task that I have initiated by sketching how the concept of participatory sense-making can be used in larger contexts than isolated dyadic interaction. Ch. 4 builds on this sketch of participatory sense-making by showing its explanatory value to neo-Kuhnian social epistemology.

⁸² This is analogous in spirit to Hans Jonas's (1966) criticism of existentialism's exclusive focus on humanity, previously noted in §3.1. Importantly, Jonas sought to gain perspective on humans by noting the role of broader biological factors.

Chapter 4: Coordination in Epistemic Communities

4.0 Introduction

Some of the most important and dramatic implications of enactivism cannot be realized without linking its account of the cognitive subject to social epistemology. Cognitive science and social epistemology mutually constrain each other, with social epistemology offering longer-term and larger-scale perspectives on the cognitive subject. The task of this chapter is to use social epistemology's population-level perspective—its identification of emergent community-level cognitive processes—to further undermine the belief that the brain is the executive of all cognitive activity.

One obstacle to realizing this task, however, is that social epistemology to date has not adequately accounted for the embodied nature of cognition. In §4.1, I revisit Bickhard's interactive model of representation in order to clarify the relationship between embodied cognitive science and social epistemology. Enactivism's embodied cognitive subject, I argue, dovetails with the social epistemology inspired by Thomas Kuhn's philosophy of science.

K. Brad Wray (2011) carefully explicates in what sense the epistemology of Kuhn's philosophy of science is social. Importantly, Wray argues that Kuhn's epistemology is also “evolutionary,” possessing deeper connections to biology than Kuhn himself acknowledges. Wray's explication, which I detail in §4.2, is helpful for establishing the potential of and also lacunae within Kuhn's work that neo-Kuhnian social epistemology has subsequently built upon. This sets up my analysis of two major works in neo-Kuhnian social epistemology: Fred

D'Agostino's (2010) *Naturalizing Epistemology* (§4.3) and William Rehg's (2009) *Cogent Science in Context* (§4.5).

The central theme of D'Agostino (2010) is how communities of enquiry balance both conservative and innovative impulses, which Kuhn (1977) refers to as the “essential tension.” Like Kuhn, D'Agostino is primarily interested in how institutional mechanisms facilitate balancing these competing needs. But D'Agostino broadens his account beyond scientific communities and the question of theory-choice, focusing on the more general question of epistemic divisions of labor. My analysis of D'Agostino in §4.3 has two ends: first, identify dynamics endogenous to epistemic divisions of labor that enable individuals to engage in novel cognitive activities and, second, explain how D'Agostino's account complements and strengthens the enactivist cognitive subject.

In §4.4, I consider two varieties of disembodied social epistemology. Importantly, Fuller's (2002, 2011, 2012) social epistemology is a direct inspiration for D'Agostino's account. My critique of Fuller thus serves two ends: it shows that an embodied epistemology provides a more nuanced perspective on D'Agostino's mapping of epistemic communities and, second, it provides the resources for further revising Kusch's communitarian epistemology (§1.7, §2.5). From this vantage point, the issue with Kusch's communitarian epistemology is its linguistically-centered normativity. This shift away from disembodied normativity is extended via Rehg's concept of cogency.

In the final section of the chapter, I consider Rehg's (2009) argumentation theory, which details “microsocial” dynamics within communities of enquiry. This helps to clarify the relationship between the institutional mechanisms described by D'Agostino and participatory sense-making's treatment of two-person interaction. The key concept of Rehg's theory is

cogency, which entails that argumentation is a context-specific process that involves an indefinite range of possible factors. Examining the immanent nature of argumentation gives a more fine-grained perspective on how epistemic agents are simultaneously empowered and constrained by their communities.

4.1 Relating Embodiment and Social Epistemology: Two Ways of Regulating Error

In §3.2, I used Bickhard's interactive model of representation to resolve an apparent tension between enactivism's endorsement of both intrinsic teleology and adaptivity and my previous criticisms of Searle's normativity and Goldman's adaptationism. Bickhard's model, in its barest form, consists of nonrepresentational goal-directedness that is refined via the dual functions of environmental differentiation and implicit definition. I return to Bickhard's model in this section in order to identify two types of learning, whose reciprocal development bears on the link between cognitive science and social epistemology.

As noted in §3.2, Bickhard (1996) regards learning as an ability to cope with error feedback. Bickhard (2002, 1) develops this view by identifying two basic ways of regulating error: (1) the regulation of the interactive processes between a rational system and its environment and (2) the construction of new interactive "(sub)systems" that act as a "kind of metaregulation." That is to say, upon encountering error, the two non-exclusive options for an organism is to (1) modify how it interacts with its environment or (2)—to use mentalistic vocabulary weighted toward persons—refine its understanding of the environment. The first and most basic type of error regulation—which most organisms are limited to—corresponds to the first level of Bickhard's "levels of knowing" model (Bickhard 2002, 8). Animals⁸³ such as

⁸³ In the following, I prefer the term "organism" or "animal" to "rational system." While I endorse Bickhard's (2002) conception of rationality—which complements the previously discussed accounts of learning and

humans can abstract from (1) and thereby represent aspects of the environment in terms of interactive possibilities. Importantly, this new level—which is the first instance of (2)—is itself open to abstraction. Abstracting from (2) would in turn create a new higher level of knowing that is also open to abstraction and so on, with no determinate upper bound on the levels of abstraction.

For the present discussion, I am specifically concerned with the process of abstracting from a lower to a higher level of knowing and its implications for the general relationship between (1) and (2). As opposed to a passive mirroring or encoding relation, abstraction entails an “unfolding” of values implicit in lower levels (Bickhard 2002, 13). Importantly, such unfolding can generate errors between levels, such as a conflict between a newly articulated value and a previously held one. Given such a circumstance, there are various ways of resolving the conflict: for example, one of the values may be modified or discarded or the domain of one or both may be restricted (Bickhard 2002, 15). The key is that lower and higher levels of knowledge mutually constrain each other: it is not predetermined whether a value from a lower or higher level will take priority or, taking up the other general alternative, whether it is possible to reconcile the values. Furthermore, abstraction—insofar as it corresponds to the internal development designated by (2)—is not an internalization of but rather runs in parallel to external processes (Bickhard 2002, 13).⁸⁴ Understanding the parallel development of (1) and (2) holds the key, I argue, to understanding the related though distinct insights of cognitive science and social epistemology.

perceptual salience—I do not have the space to rehearse his argument and so refrain from using “rational system” in order to avoid any misleading associations.

⁸⁴ It’s worth noting that the role of abstraction thus addresses the criticism leveled by Dreyfus (1972, 236) and Searle (1983, 150) against Polanyi, which claimed Polanyi’s framework required an internalization of rules. The criticism occludes some of the empirical dimensions of tacit knowledge that were argued for by Polanyi and also Turner.

The parallel development of (1) and (2) is reciprocally enabling. The abstractions of (2) facilitate the creation of more specialized environmental niches that in turn dramatically reshape (1). A notable example of an epistemic niche is an institutionalized community of enquiry—which, as extensively detailed by social epistemology, facilitates novel forms of cognitive activity and levels of abstraction. This latter implication thus designates a way in which (1) reshapes (2). That is, the construction of epistemic niches, such as research institutions, play an integral role in new levels of abstraction that fall under the direct purview of cognitive science. The development of (1) and (2) thus run in parallel in the sense that there is no isomorphic correspondence between them—that is, (2) doesn't internalize (1)—even though the two mutually influence and enable developments within each other.⁸⁵ De Jaegher and Di Paolo's concept of participatory sense-making occupies a special place with regard to this parallel development, since participatory sense-making relates directly to both (1) and (2).

Participatory sense-making is Janus-faced. As an extension of enactivism, participatory sense-making is concerned with the cognitive subject as such; while as a model of face-to-face interaction, participatory sense-making is concerned with a vital type of epistemic niche. It should be noted, however, that one of enactivism's most important insights is that the cognitive subject cannot be modeled in a vacuum. Hence, participatory sense-making's concern with two-person interaction is not simply an add-on to a core enactive subject but rather the articulation of a particular interactive mode. Separated from all modes of interaction, the cognitive subject is a meaningless abstraction. The present chapter is concerned with how participatory sense-making's insights on two-person interaction shade into the larger-scale domain of social

⁸⁵ The reciprocal relationship I have in mind is a reflective equilibrium, as articulated by John Rawls (1999) who was inspired by Goodman (1983). I develop this connection in Ch. 5.

epistemology. To help situate this proposal, I offer the following outline of Chs. 3 – 6 in connection with the reciprocal relationship between (1) and (2) above.

The following outline of Chs. 3 – 6 complements the one presented at the end of Ch. 2. There I noted the main tasks of Chs. 3, 4, and 5; to wit, to undermine the input-receiver picture of cognition, to identify community-level cognitive processes, and, finally, to undermine the egoist presupposition within cognitive science and social epistemology. From the present viewpoint, Ch. 3—via the concept of participatory sense-making—counteracts the tendency to inflate the importance of abstract representations by offering a balanced characterization of how refinements in an organism’s interactions with its environment [i.e., (1)] relate to refinements in the representations of its environment [i.e., (2)]. The present chapter articulates participatory sense-making’s implications in terms of larger-scale epistemic niches, thus building upon enactivism’s basic characterization of (1). Ch. 5 focuses on the nature of (2), using the present chapter’s more comprehensive account of social niches to challenge egoist reductions of the cognitive subject. Finally, Ch. 6 concludes that given the reciprocal nature of (1) and (2), there is no fixed cognitive core that underpins individual subjects.

4.2 Kuhn as a Bridge to Contemporary Social Epistemology

Thomas Kuhn is the most influential intellectual figure behind social epistemology.⁸⁶ Wray (2011)—as the title *Kuhn’s Evolutionary Social Epistemology* suggests—explicates Kuhn’s philosophy of science with an eye towards the contemporary field of social epistemology. In the

⁸⁶ The label “social epistemology” gained currency in the late 1980s with Steve Fuller’s founding of the eponymously named journal. As evidenced by my interest in Kuhn’s influence on the field, my understanding of social epistemology has undertones of Fuller (2002). Yet I find Fuller’s (2012) mapping of the field unhelpful—he pits his own politically oriented approach against “analytic social epistemology”—because Fuller disregards the embodied nature of cognition. In fact, Fuller (2011) proposes the chimerical goal of disembodied minds. I return to this issue in §4.4.

present section, I draw upon Wray's explication in order to give more context to two neo-Kuhnian works that I examine in §4.3 and §4.5. On the one hand, Kuhn sets the explanatory targets for much of social epistemology while, on the other, suffering from some major lacunae that have been forcefully highlighted by the contemporary field.

To understand Kuhn's uneven contribution, it is helpful to start with his formative interest in the psychology of discovery. In particular, Bruner and Postman (1949) influenced Kuhn greatly. In Bruner and Postman's study, subjects were asked to successively identify five playing cards, some of which were painted in the opposite color (e.g., a black three of hearts) (209). It took on average four times longer for test subjects to identify such "trick" cards (210). The key lesson that Kuhn drew from the study is that "perceptual organization is powerfully determined by expectations built upon past commerce with the environment" (Wray 2011, 51; Bruner and Postman 1949, 222). This psychological finding bears on the process of discovery in science, offering one explanation for why discoveries can be protracted affairs.

Speaking to the pressing issues of science, Kuhn's epistemology is intimately concerned with the continual growth of knowledge and the obstacles to said growth. As discussed in §2.1, Polanyi (1958) also approaches the issue of scientific discovery using psychology, a connection that is much more substantial in Polanyi than in Kuhn. Despite his interest in the Bruner and Postman study and Gestalt psychology more generally, Kuhnian epistemology is ultimately far more indebted to the philosophy of language than to psychological accounts of perception or any other details of embodiment. This priority is evident in how Kuhn modified his understanding of scientific revolutions.

In the wake of criticism of *The Structure of Scientific Revolutions*, Kuhn eventually stopped characterizing scientific revolutions as paradigm changes, which—akin to the Bruner

and Postman study—he had described as gestalt shifts in perception (Kuhn 1962, 85; Wray 2011, 15). In place of gestalt shifts, scientific revolutions came to be understood as lexical or taxonomic changes within a particular discipline. This shift dovetailed with Kuhn’s emphasis on theory choice and the overarching role of the philosophy of language. Like Rehg (2009, 44) and in concert with my previous criticisms of linguistically based conceptions of normativity, I find Kuhn’s dependence on the philosophy of language of dubious value.

But the development within Kuhn’s thought did have the beneficial effect of bringing more attention to the complexities of “normal” science. Wray (2011, 137) goes so far as to claim that Kuhn offers revolutionary *and* evolutionary perspectives on science, a decidedly more nuanced picture than *The Structure of Scientific Revolutions*’ stark dichotomy between normal science as mere articulation of a paradigm and revolutions as irreversible conversions to new paradigms (Kuhn 1962, 111). Wray’s characterization refers to Kuhn’s use of the term “evolution.” Kuhn (1992, 115), for instance, uses “evolution” as a metaphor for the proliferation of scientific disciplines.

Wray argues that despite Kuhn’s limited metaphorical intentions, the specialization of science reflects an important feature of scientific theories’ ontological status. The proliferation of specialties suggests that science is not converging on one universal ontology.⁸⁷ Science is “pushed from behind”—constrained by its previous theories and their attendant vocabularies and taxonomies—rather than converging on a single picture of the natural universe (Wray 2011, 203). In short, Wray judges Kuhn’s epistemology to be evolutionary since Kuhn depicts science as a historical process that lacks a fixed *telos*.

⁸⁷ For a complementary perspective, see Nancy Cartwright’s (1999) *The Dappled World*. Rather than considering the sociological dimension of scientific practice, Cartwright argues against universalism in science based on the nature of scientific laws, which she claims are only domain-specific regularities.

Kuhn (1992) views himself as avoiding the excesses of both realism and relativism. Even though science is not converging on a unified account of a theory-independent reality, Kuhn is an internalist with regard to theory change (Wray 2011, 160). In this context, internalism entails the claim that scientific disputes are settled on the basis of evidence rather than factors putatively “external” to science, such as political concerns or other such interests. Political concerns and the like are of *no* epistemic significance, according to Kuhn. The internalist position is directed especially against the so-called strong program in the sociology of knowledge, which Kuhn invested a great deal of effort in criticizing, such as in his 1992 essay.⁸⁸

As evidenced by his critique of the strong program, Kuhn construes the social dimension of epistemology narrowly. The essential feature of science’s social dimension is, on Kuhn’s view, how individuals weigh objective criteria differently (Wray 2011, 161). Given the domain-general values of accuracy, simplicity, consistency, scope, and fruitfulness, for example, individual scientists will prioritize each differently, differences that impact the evaluation of competing theories. This subjective dynamic is understood to cause risk spreading within scientific communities and also offers another explanation for why scientific disputes can be protracted. Yet, insofar as it’s epistemically significant, the subjective weighting of criteria acts *only* to spread risk within a community by ensuring that scientists will work on competing theories (Kuhn 1977, 332). The subjective factors ultimately *do not* play a role in theory choice: after some period of time, the community’s efforts either produce a clear winner (one that everyone endorses) or the community branches into different specialties.

⁸⁸ Kuhn took particular issue with the strong program’s “symmetry principle”—in which the same types of explanations should be employed for successful *and* unsuccessful knowledge claims. Kuhn believes that the principle obscures the distinctive role that nature plays in scientists’ formation of beliefs. See Bloor ([1976] 1991) for the first and most famous defense of the symmetry principle; for a general overview of the strong program, see Barnes, Bloor, and Henry (1996).

Philip Kitcher challenges Kuhn's exclusion of political concerns and interests for very different reasons than those of the strong program. Taking up the theme of risk spreading, Kitcher (1990) argues that a broad range of interests are in play. In a scenario reminiscent of James Watson and Francis Crick's discovery of DNA's structure, Kitcher imagines a community of scientists pursuing the structure of a very important molecule ("VIM"). In a situation where the discovery would earn a coveted prize (e.g., a Nobel Prize), Kitcher argues that a community composed of ruthless egoists would achieve a better spread of effort and risk than if all of the scientists were pursuing high-minded ideals (e.g., acquiring truths) (16).⁸⁹ Additional and less sordid interests—such as national or personal loyalties, personal investment, and so on—also ensure a diverse distribution of effort. It is important to note that Kitcher's incorporation of these diverse interests rests on the same justification as Kuhn's own understanding of the social dimension of science: for both, these social factors are instrumentally valuable for furthering research.⁹⁰ Kitcher thus shows that one of Kuhn's most important themes—the distribution of risk within a community—involves a diverse range of interests that promote rather than detract from scientific enquiry and thereby should not be counted as external to science. In other words, Kitcher makes the compelling (and, in retrospect, somewhat obvious) point that the credit mechanisms of science are not incidental to its progress and growth, indeed they are more than incidental to its very subject matter.

The role of what Kuhn considered external factors does not undermine Kuhn's evolutionary picture of science. With regard to the latter, Kitcher (2001) introduces the concept of "significance graphs" to model the trajectory of scientific research. As with Kuhn's picture,

⁸⁹ D'Agostino's (2010, 53-54) discussion of communication inhibitors, such as first-mover bias and social comparison pressures, bolster's Kitcher's point by suggesting that there are common factors that can overwhelm any differential weighting of objective criteria.

⁹⁰ In Ch. 5, I challenge this type of manipulationist reduction of epistemic interaction.

Kitcher recognizes a profound and ineliminable historical dimension to scientific research: “Because some instruments, techniques, sites, or model organisms become embedded in the significance graphs of different fields, so that researchers know how to use them, the evolution often shows a kind of inertia” (81). The breakthroughs of a field act simultaneously as constraints, with their associated terms and techniques possessing an inertia that can inhibit further discoveries. Although Kuhn’s picture is not undermined, it is significantly broadened with the incorporation of political concerns.

The contingent trajectory of research, as described by significance graphs, highlights the need for what Kitcher calls “well-ordered science” (117). That is, without actively scrutinizing and revising its direction, science will not serve the collective good of human society, falling far short of what it could contribute. Kitcher (2011) further develops the concept of well-ordered science and includes a more extensive critique of the notion of value-neutral science. Once it is granted that science should help human society with particular types of problems—such as battling disease and improving access to life necessities—any use of resources within science reflects a value judgment. This is not to say, as Kitcher notes, that science should not pursue purely theoretical problems, since such problems satisfy human curiosity—which in itself carries value—and also has the potential of indirectly addressing practical problems (109).

Although I find the Kitcher’s concept of well-ordered science problematic—especially its use of an “ideal conversation” as a normative standard (106)—well-ordered science serves well as an immanent critique of Kuhn’s internalism. Kitcher shows that Kuhn is wrong to judge political and other social interests as of no epistemic value, a broadening of Kuhnian epistemology that has been even further developed by other theorists. In the next section, I

examine one such theorist, Fred D'Agostino, who considers the division of labor within research communities in contrast to Kuhn and Kitcher's focus on risk distribution.

4.3 D'Agostino's Federal Model of Enquiry

The best starting point for D'Agostino's treatment of epistemic divisions of labor is his broader understanding of social epistemology's domain. D'Agostino organizes his account around the concept of an "assembly bonus," a concept adopted from social psychology and management science. The concept of an assembly bonus refers simply to a benefit wrought from pooling the cognitive efforts of different individuals without specifying any particular type of task or activity (Collins and Guetzkow 1964). Accordingly, the subject of *Naturalizing Epistemology* is not just scientific disciplines as such but communities of enquiry in general.

One reason for drawing upon fields such as management science is a lack of substantive analyses of epistemic divisions of labor within social epistemology, despite it being a familiar theme. Kitcher (1990), for example, takes up the theme in "The Division of Cognitive Labor," but only addresses its "subjective" dimension (D'Agostino 2010, 113). That is, Kitcher addresses how individual attitudes to a problem differ but does not consider the actual process of breaking down a problem into smaller tasks within a collaborative group. The subjective dimension of the division of labor consists of interpersonal communication issues, such as the previously mentioned obstacles of first-mover bias and social comparison pressures (§4.2). And Kitcher's egoistic pursuit of prizes represents one means of overcoming such obstacles.⁹¹

Shallow consensus straddles the subjective and objective aspects of epistemic divisions of labor. As mentioned previously (§1.6, §1.7, §2.5), shallow consensus describes a

⁹¹ D'Agostino notes other "disinhibitors" that help to overcome communication issues, including "multiple accountabilities," "novelty premiums," "asset ownership," and "intrinsic motivation" (64).

communicative pattern in which agreements provisionally conceal underlying disagreements.⁹²

In this respect, shallow consensus relates to the subjective dimension of divisions of labor, enriching the picture of risk spreading discussed by Kuhn and Kitcher. While useful for communicative purposes, shallow consensus is an artifact of working in complex, multifaceted problem spaces, which corresponds to the division of labor's objective dimension. Shallow consensus is thus a useful point of entry for examining how enquiries fare in such problem spaces.⁹³

D'Agostino (2010) considers how four different types of enquiry fare in complex problem spaces: exhaustive, myopic, parallel, and modular enquiry (127). Exhaustive enquiry guarantees finding a global optimal solution but is inefficient given its time and resource demands; myopic enquiry is efficient but is not effective for finding an optimal solution. Parallel enquiry can be both efficient and effective but presupposes that a problem space is decomposable into independent tasks. Any residual interdependencies between tasks render a parallel enquiry inconclusive, potentially pitting individuals at cross purposes.

The final type of enquiry is a response to the fact that problem spaces often cannot be decomposed without residual interdependencies, especially not without presupposing the very knowledge of the domain that is sought after. Modularity is a design that creates a "high degree of independence or 'loose coupling' between component designs by standardizing component interface specifications" (128). Modularity thus reflects the fact that standards must be constructed for how decomposed components should relate to each other.

⁹² One of shallow consensus's many insights is to help bridge the Kuhnian notions of normal science and scientific revolutions. Shallow consensus suggests that disagreements exist but are minimized in relatively "normal" periods, while during crises people actively seek and draw attention to them. Importantly, there is a range of intermediate situations between these two extremes.

⁹³ To be clear, shallow consensus and the need for a division of labor result from the complexity of problem spaces *and* humans' bounded rationality (20). I discuss bounded rationality below.

To illustrate modularization and how it differs from parallel enquiry, D’Agostino draws on the example of designing sub-systems of a computer from Baldwin and Clark (2000). While it is most efficient to separate the tasks of designing a computer’s hard drive system and its motherboard—the first step of modularization being the division of such tasks—there are still residual interdependencies between the designs. Baldwin and Clark (2000, 49) note that in their case study there were as many possibilities of conflicting designs as there were compatible ones. Modularization is an effort to minimize the possibility of such conflict. The second step—which is the decisive difference between parallel and modular enquiry—consists in building an “interface” between the tasks. The interface assigns parameters for each task in order to ensure consistency, acting as a “design rule” for the separate teams (D’Agostino 2010, 128).

The assurance of consistency comes at a cost: designers lose the ability to explore some possible solutions. This is the basic sense in which fixing parameters imposes a *path dependency* on the eventual solution. Only an exhaustive enquiry can guarantee that there is not a more effective global solution, but using an exhaustive enquiry to validate a modular one would undercut the latter’s efficiency. Modularity thus mitigates though does not eliminate the trade-offs between efficiency, effectiveness, and conclusiveness found in exhaustive, myopic, and parallel enquiry, respectively. Modularization is not by itself a methodological blueprint for enquiry but depends on broader dynamics within epistemic communities.

Given the limitations of modularity, D’Agostino notes the need for “*both division and diversity to pursue enquiry in complex situations*” (134; original emphasis). That is, it is necessary to have different teams create different design rules and then compete, producing knowledge that is less myopic than a single modularization but still much more efficient than an exhaustive enquiry. Having competing teams thereby mitigates but does not eliminate the

myopia that results from modularity's path dependency. D'Agostino writes, "we are stuck with history... 'the explanation of why something [such as a problem-solution] exists rests on how it became what it is'" (135). Solutions to complex problems must be understood in terms of historical paths. The arbitrariness that results from path dependency is independent of the arbitrariness produced by the sordid motives of individual researchers (136). It is important to note that the objective path dependency of enquiry reframes the importance of political interests and the like, since such interests do not veil what would otherwise be value-neutral enquiry.

Furthermore, D'Agostino's account of path dependency augments Kuhn's "evolutionary" epistemology. The basic insight shared by both is that enquiry should be understood in terms of where it has come from rather than as approaching a determinate endpoint. The concept of modularity shows that this is not merely an artifact of the rhetorical demands of winning arguments but intrinsic to the very posing of problems/tasks. With modularity, it is also less tempting to conflate successful solutions/theories with being the most "fit." Although I fully embrace a biological approach to epistemology—as indicated by my endorsement of enactivism—appealing to evolution in the context of social epistemology, as in the case of Kuhn and Wray, is fraught with danger. Even though biological evolution does not entail equating evolution with optimal fitness, the adaptationist intuition is difficult to shake. Modularity clearly illustrates how a superior path may be reasonably set aside because it appeared inferior at an early stage of enquiry.

Building upon the generic notion of an epistemic division of labor, D'Agostino proposes a "federalist model" of enquiry (133). The basic purpose of the model is to conceptualize how intra-research team dynamics relate to the community-level competition of different teams. The federal model demarcates a community of enquiry along a horizontal and a vertical axis

consisting of “jurisdictions” and “levels,” respectively (147). A jurisdiction consists of a deliberative body that partitions a research domain. Some jurisdictions run in parallel to each other, some overlap, while some subsume others. Levels correspond to hierarchies between various jurisdictions, the highest level possessing only one deliberative body. At lower levels, the deliberative bodies are research teams that work roughly in parallel and compete with each other.

Higher levels act as “adjudicators” for lower ones, with the highest level being the final adjudicator. An adjudicator, whether at the highest level or only a relatively higher one, has a privileged perspective on the competition between two teams (149). The adjudicator will have access to more information than either team had when each began, will be less attached to either solution than the respective teams, and can judge each solution in terms of its path as a whole. The federal model, with its horizontal and vertical axes of enquiry, to some extent idealizes the structure of actual epistemic communities.⁹⁴ In a particular community of enquiry, the roles of team member and adjudicator are often mixed up, with individuals serving multiple functions and working on multiple teams.⁹⁵

The federal model of enquiry thus accomplishes the main task of the chapter—namely, to identify cognitive processes that are endogenous to particular epistemic communities. Within a research team, individual cognitive tasks are structured by modular interfaces in order to ensure coordination with other individual efforts. That is to say, individual efforts are *not* isolated activities that are aggregated to form a group product. The significance of modularization is

⁹⁴ Kitcher and Wray both allude to the federal model’s two axes. Wray (2011, 175) states that there is “a sense in which research teams can be said to ‘believe something’ or ‘hold a view,’” which indicates a fundamental difference between team collaboration and the larger community within which it occurs. Of more direct relevance, Kitcher (1990, 17) notes that his model of a purely self-interested community of individuals only excels if it is divided up into “fiefdoms.” Yet neither Wray nor Kitcher explore the general dynamics between these two axes.

⁹⁵ As I detail in §4.5, Rehg’s (2009) Fermilab case study substantiates and also complicates D’Agostino’s model of parallel and hierarchical axes of enquiry.

evident from the need for competition between different groups, a result of the myopia induced by modularization's path dependency. Furthermore, the crucial role of adjudication can only occur within such a community. In sum, the efforts of individuals can be properly understood only in relation to how they complement and compete with others' efforts.⁹⁶ Furthermore, the risk spreading emphasized by Kuhn and Kitcher suggests that, in such communities, individuals themselves understand their efforts in relation to others.

The emergent patterns identified by the federal model of enquiry are, I argue, general types of coordination. On my reading, the federal model identifies patterns of joint sense-making and thereby extends the insights of participatory sense-making to larger social contexts. Yet even though his account complements enactivism and its concept of participatory sense-making, D'Agostino himself falls short of endorsing an embodied epistemology. This is in part because D'Agostino brackets the issue of the fundamental nature of knowledge in order to focus on describing epistemic communities. The upshot of this argumentative strategy is that D'Agostino's account stands on its descriptive merits, which makes the account relevant for a wide range of theorists. The downside, on the other hand, is that the account is open to re-interpretations that cut against the spirit of D'Agostino's model.

Ladyman (2012, 605), for example, faults D'Agostino (2010) for not addressing the question of *what* knowledge and justification are. Such questions must be addressed, Ladyman thinks, in order to *naturalize* epistemology, which the title of D'Agostino's book promises to do. Ladyman thus proposes that a more appropriate title would have been "socializing knowledge" (605). Accordingly, Ladyman suggests an individualistic reconstruction, asking whether it is

⁹⁶ One implication of path dependency is the need for narrative explanations of epistemic communities. This is notable since attempts to show that interactive dynamics are quantifiable at times appear to insinuate that controlled measures and experiments are the ideal form of social analysis. See, for instance, De Jaegher, Di Paolo, and Gallagher (2010).

better to think of Kuhn and D'Agostino's essential tension as existing at the individual level. D'Agostino's description of community dynamics would thus explain how a larger social group balances out each individual's unique blend of conservative and innovative impulses (606). Ladyman's proposal echoes the approach to social epistemology exemplified by the testimony view, in which social aspects are strictly derivative upon individual cognitive mechanisms.

While Ladyman does not himself pursue the following line of reasoning in his brief review, his proposal can cite D'Agostino's own distinction between the objective and subjective dimensions of the division of labor. As noted above, D'Agostino distinguishes between the objective and subjective dimensions of epistemic divisions of labor in order to differentiate between, on the one hand, the decomposition of a problem and, on the other, Kitcher's focus on differing individual attitudes toward problems. For theorists like Ladyman, the objective dimension is more fundamental than the subjective one. It is clear from D'Agostino's own account how the objective dimension shapes and supports the subjective one. For example, he writes: "the very existence of the objective impediments may serve to 'tune' our communities of enquiry to work effectively in overcoming the subjective impediments" (143). It is, in this manner, important that the environment is a complex problem space because this complexity is what elicits and fosters a division of labor and a diversity of approaches.

Yet the reciprocity of the relationship—how the subjective dimension shapes and supports the objective one—is not as apparent. D'Agostino details the significance of having an appropriate culture of enquiry, such as a culture that includes shallow consensus. But such considerations are easily re-situated within Ladyman's individualistic epistemology as mere exigencies and, as such, they would be incidental to the nature of objective knowledge. Such a

reading could even grant that human understanding of objective reality is always tied to and constrained by the historical trajectories of our epistemic communities.

On this Ladyman-inspired reading, enquirers compete with each other to better approximate aspects of the objective world, despite being stuck with inherent arbitrariness. This interpretation is consistent with Kuhn's evolutionary epistemology so long as it is not claimed that all of science's disciplines are approximating the objective world in a unified manner. Hence, the interpretation reads Kuhn's evolutionary epistemology as essentially a contemporary version of the Tower of Babel story. In this context, D'Agostino's mapping of epistemic communities would be a manual for individuals to better navigate the aspects of the objective world that they are specifically concerned with.⁹⁷

In sum, while D'Agostino shows how social epistemology informs cognitive science—counteracting the latter's tendency to treat interactive environments as merely input for cognitive processing—it is also important to recognize how cognitive science informs social epistemology. With regard to the present discussion, embodied cognitive science is useful for showing how the subjective dimension also shapes and constrains the objective one, which thereby counteracts realist re-interpretations of D'Agostino, such as Ladyman's individualistic reconstruction.

The frog example from §3.2, for instance, indicates what is wrong with a realist interpretation. The lesson of the example is that a frog perceives tongue-flicking opportunities (e.g., a flick-at-one-point, a flick-between-two-points, and so on)—as opposed to first identifying an object (e.g., a fly or worm), then determining whether it is edible and, if so, what tongue flick the object demands, and, finally, acting on a desire to eat by actually flicking its tongue. Instead,

⁹⁷ The plausibility of this interpretation stems in part from D'Agostino's (2010) overall task of charting factors that aid and hinder assembly bonuses. While D'Agostino's practical aims are not themselves problematic and, indeed, are of obvious value, I argue in Ch. 5 that a manipulationist interpretation (e.g., the idea that epistemic agency is only a means for acquiring objective knowledge) *is* problematic. D'Agostino himself does not address this question.

objectivity is a function of how an organism partitions its world into interactive possibilities: that is, how it anticipates, acts, and then differentially responds to the feedback from its actions.

Only if there were a dramatic change in a frog's environment—such as people flicking BBs in the frog's visual field—would a frog need to refine its discrimination of tongue-flicking opportunities. But even given such a refinement, perception would still consist of differential contact rather than the reception and subsequent processing of external content.⁹⁸ Because it is a function of an organism's interaction with its particular environment, objectivity is not more fundamental than subjectivity. Given the teleology intrinsic to cognition—that is, the nonrepresentational goal-directedness of self-maintenant processes—it is appropriate to conclude that the subjective aspect of enquiry reciprocally shapes the objective one.

Embodied epistemology, as articulated by enactivism and Bickhard's model of error-guided learning, is thus one way of responding to Ladyman. If knowledge is in its most primitive form skillful interaction with the world, then providing a “how” as opposed to a “what” account does, in fact, help to *naturalize* epistemology. Yet it is not entirely clear how sympathetic D'Agostino is to an embodied epistemology. As evidenced by some early glowing references, D'Agostino's anti-realism is tied to Martin Kusch and Richard Rorty rather than embodied cognitive science (2). The closest that D'Agostino comes to endorsing embodied epistemology is in his discussion of “bounded rationality” (20). Bounded rationality is a formal description of some cognitive limits inherent to finite beings. D'Agostino's goal with the concept is to show that the collectivization of inquiry is a response to “in principle” considerations as opposed to

⁹⁸ It is worth noting that understanding perception as differential contact buttresses social epistemology's concern with the details and dynamics of epistemic communities. It implies that epistemic niches in general—scientific institutions being particularly notable examples—are not merely a means for more efficiently filtering information but rather are an integral part of perception.

merely practical exigencies. That is, epistemic communities are more than the aggregate effort of what would happen if a single individual had more time and resources.

Of the seven factors that D'Agostino notes in connection with bounded rationality, the first and most important one is inexhaustibility. Echoing the familiar Kuhnian theme of nominalism, inexhaustibility refers to the endless ways in which any material state of affairs can be described (21). In this respect, D'Agostino's anti-realism, like Kusch, is closely tied to the philosophy of language. This focus is borne out by the rest of the factors that D'Agostino discusses. Reflexivity, for example, is even more closely related to Kusch: when a claim is made regarding agents and those agents subsequently become aware of it, the claim may be either reflexively undermined or confirmed. Reflexivity thus highlights the performative dimensions of language use.

Rather than linking these factors to the embodied nature of cognition, D'Agostino moves towards more abstract matters. The other factors of bounded rationality explicate the nominalist implications for human deliberation. While bounded rationality thus addresses key issues within the domain of social epistemology, without the explicit link to an embodied epistemology it remains vulnerable to realist re-interpretations.

Demonstrating the pertinence of embodied cognitive science completes the section's second major task; to wit, to show the reciprocity between neo-Kuhnian social epistemology and embodied cognitive science. D'Agostino's federal model of enquiry counteracts cognitive science's tendency to downplay the significance of epistemic niches, while conversely, embodied cognitive science forestalls realist reconstructions of D'Agostino's account. One of the chief obstacles to this connection, however, comes from within neo-Kuhnian social

epistemology; in the next section, I consider a neo-Kuhnian social epistemology that is at odds or, at the very least, in serious tension with embodied epistemology.

4.4 Two Disembodied Social Epistemologies

In the present section, I contrast my proposed connection between social epistemology and enactivism with two varieties of disembodied social epistemology. The hallmark of the latter is the belief that social interaction rests upon mutually recognized truths. Fuller (2002, 2011, 2012) and Kusch (2002a), each in their own distinctive way, endorse such a view. At stake between embodied and disembodied conceptions of social epistemology, I argue, are competing accounts of normativity, with the former providing a more nuanced and heterogeneous account. To motivate an embodied social epistemology, I model its normative dimensions on Bickhard's account of error-guided learning, which entails that normativity is an immanent phenomenon. I introduce this model by way of critiquing Fuller's social epistemology, a critique that provides the resources for then reconstructing Kusch's communitarian epistemology.

Fuller's epistemology is directly informed by political philosophy. Fuller (2002, 6) claims that epistemology is nothing other than political philosophy applied to the community of enquiry. The political nature of epistemology relates to the same basic issue raised by Kitcher (2001, 2011)—namely, scarcity of resources. Fuller (2012) writes:

Much of the ongoing discussion, especially among analytic epistemologists, about whether science “aims for the truth” has been misdirected because, in my view, the interesting disagreements arise less over that question than which truths are worth pursuing and the means by which they are pursued. (269)

The passage is important, in part, because it alludes to Fuller's commitment to a realist metaphysics—in which truth is a correspondence relationship between an idea and a material state of affairs—a commitment that is easy to overlook given his vigorous criticisms of analytic

social epistemology. But the issue with analytic social epistemology, for Fuller, is not its notion of objective truth as such but rather its disregard for the issue of how resources should be allocated in scientific research.⁹⁹

Fuller's distinctive brand of realism, which he labels "realizationist," emphasizes the pliability of material states of affair. Fuller (2012, 272) writes, "my position would be more accurately described as 'realizationist' (rather than, say, 'relativist'). I believe that we increasingly come to turn into reality whatever we conceive." Fuller is optimistic about the potential for science to structure physical reality in order to meet our desired conceptions of it. Given this potential and the wide range of possibilities to pursue, it is essential that research realize the most beneficial desires or values.

More important than any of its specifics, Fuller beckons to a wide and radical range of trans- or post-humanist possibilities. Trans-humanism is founded on the idea that humans are "intellects that happen for now to possess animal bodies" (Fuller 2011, 63). If the human mind is multiply realizable, then future technology could in principle supplant the human body. Fuller's unique twist on this theme is to emphasize its potential for achieving political equality and liberty, so long as the technological advances extend to everyone (57). This relates back to what Fuller (2000, 2002, 2012) has long considered the underlying purpose for acquiring any kind of knowledge—namely, to free oneself from domination by others.

In short, if our epistemic communities are working towards the right ends and working together, then Fuller thinks it possible to *realize* equality between people and thereby address what is perhaps the most basic injustice of contemporary society. The role of social epistemology,

⁹⁹ Fuller lumps Kitcher together with analytic social epistemology based on Kitcher (1993) in which Kitcher defends the objectivity of science. Conspicuously absent, however, is any mention of Kitcher's (2001) notion of well-ordered science, whose central tenet is the need for citizenry to inform the direction of science, a proposal quite similar (even if less flamboyant) to Fuller's own.

in this picture, is to “remov[e] obstacles to both the expression of epistemic interests and knowledge of the results of actions taken on those interests” (2002, xvi). That is to say, social epistemology should improve the efficiency of the public’s communication of its desires and, second, track and communicate the effects of the ensuing epistemic projects. Ideally, social epistemology in this way acts as a shared foundation for our epistemic undertakings by informing the public of the competing values at stake and the most efficient means for realizing these values.

Despite many useful insights and provocations, Fuller’s account founders on his commitment to multiple realizability, which presupposes the associated realist metaphysics. Fuller’s argument for multiple realizability is woefully inadequate, which is to say almost non-existent. The argument rests on a critique of embodied epistemology; while resting one’s case primarily on the demolition of another position often hints at a straw man, the strategy is familiar in the philosophy of mind. What is unpardonable, however, is Fuller’s equivocation of embodiment with Darwinian evolution.

With always a taste for intellectual duels and dichotomies, Fuller (2011, 51) equates the embodiment-multiple realizability issue with an ideological struggle between earth-bound Darwinian biology and a theistically-inspired universal life and intelligent design science. The closest that Fuller comes to acknowledging enactivism or any other flavor of embodied cognitive science is a single reference to evo-devo (evolutionary-development) theory (61). Making matters worse, evo-devo theory is interpreted as part of biology’s more general rediscovery of purpose in nature that—given the dichotomy between Darwinian evolution and intelligent design—appears to support shifting away from the putative chains of earth-bound embodiment thinking.

One result of the Darwinian straw man—in which natural selection is considered the only driving force of evolution—is that Fuller, like Clark (§3.4), fails to heed the role of intrinsic teleology or intentionality within embodied cognitive science and epistemology.¹⁰⁰ This is most evident in the caricature of equality and liberty above, wherein the body appears as an aggregate of material needs that the mind takes care of as if they were menial chores. The body’s physical needs are thus considered incidental to the mind’s powers—as if the body were just another aspect of the material world that biologically dependent minds wisely choose to maintain. According to this picture, a human mind could just as easily perceive potential food as the color of tree bark—at least if the mind were freed from the chains of the biological body and thereby the chore of feeding it. In this manner, values like political equality and liberty appear as objective normative standards, which form the blueprints for how best to engineer material reality. To provide an alternative to this picture of values and the attendant conception of normativity, I first briefly revisit the concept of intrinsic intentionality discussed in Ch. 3.

Intrinsic intentionality is the basis for but does not itself consist of conscious intentions or representational aboutness. This distinction corresponds to Weber and Varela’s (2002, 100) distinction between two senses of teleology: the most common sense of “external seemingly purposeful design” corresponds to representational aboutness while the second sense—“internal purposes immanent to the living”—denotes the biological nature of intrinsic intentionality. While Fuller thinks that the normative dimension of social epistemology is exhausted by conscious design, on the enactivist view, values emerge from and remain intimately linked to the immanent purposiveness of living beings. Cognition, on this latter view, consists of an inherently value-laden perspective on the world; as opposed to being an incidental chore, metabolic constraints

¹⁰⁰ In fairness to Fuller, much of the contemporary teleology literature is geared towards the principle of natural selection, as discussed in §3.2. Yet when taking post-embodiment as a general goal, it is incumbent upon him to reach for fruits higher up on the tree.

are the originary source of meaning and serve as the most basic normative standards for learning. Given this concept of intrinsic intentionality and immanent normativity, I will now articulate the implications in contexts that are thoroughly awash in representations, specifically contexts in the domain of social epistemology, in contrast to cases of minimal cognition in which the role of metabolic constraints are apparent. Bickhard's interactivist model (§3.2, §4.1) has already provided an important first step to achieving this by grounding representations in nonrepresentational goal-directedness. To link Bickhard's model of representation and his related account of error-guided learning to social epistemology, I first connect it to D'Agostino's brief remarks on the normative implications of his book *Naturalizing Epistemology*.

D'Agostino (2010) hints at more nuanced forms of normativity than envisioned by Fuller. Near the end of *Naturalizing Epistemology*, for example, D'Agostino styles his account as possessing "a definite whiff of the normative" (178). While D'Agostino pitches his account on the basis of its descriptive merits, he notes that a map of a terrain—such as his mapping of epistemic communities—has normative implications for any user, even if the mapmaker does not tell the user where to go. When connected to an embodied epistemology, D'Agostino's point is more profound than he himself insinuates. Not just maps but descriptions in general carry normative significance, albeit more indeterminate and inchoate than, for example, a command.¹⁰¹ In this manner, it is profitable to read D'Agostino's comments from the perspective of Bickhard's notion of levels of knowing.

Descriptions repartition an agent's interactive possibilities. To describe something is to abstract from a particular interactive context, with the greater abstractness affording new

¹⁰¹ This is something of a truism for advertisers, for whom brand recognition is the Holy Grail.

interactive possibilities and potential errors, creating a new level of knowing.¹⁰² Descriptions are thus not value-neutral representations of a material state of affairs but, rather, a value-laden perspective on what aspects of an interactive context an agent considers most relevant and salient. This reckoning of descriptions dovetails with the translation view's model of tacit learning (§2.3), in which verbal instructions are improvised in the midst of interaction as opposed to being part of a universal recipe for performing a given skill. By conceptualizing the role of embodied skills, tacit knowledge points to the limits of objectifying the normative dimensions of epistemology.

An embodied epistemology suggests a related though decidedly more modest role for social epistemology than Fuller envisions. Rather than setting the menu for how scientists should cook up a post-human future, social epistemology's focus is more retrospective in identifying interactive patterns within communities, a project exemplified by D'Agostino's federal model. The interactive patterns indicate, in part, inchoate forms of normativity—in particular, social epistemology can point to patterns that do not correlate with and may even be contrary to the conscious intentions and epistemic goals within a community. To flesh out more of the immanent normativity at work in such scenarios, I turn now to reconstructing Kusch's communitarian epistemology within an embodiment framework.

In §2.5, I used the concept of shallow consensus to reconcile Turner's individualism and Kusch's collectivism in order to move beyond the individualist-collectivist dichotomy. Having since introduced the enactivist cognitive subject and D'Agostino's federal model of enquiry, it is

¹⁰² Beliefs are, on this view, an especially important form of abstraction but neither an epistemically nor normatively primitive one. In exploring the nature of perception, Maurice Merleau-Ponty ([1945] 2012, 32) makes a related though more fundamental point. Criticizing both empiricist and intellectualist accounts of perception, he argues that conscious attention does not clarify "preexisting givens" but rather actively constitutes new, determinate objects. Merleau-Ponty thus considers all conscious thought as derivative upon agents' embodied interaction with their environments. In this manner, the epistemic significance of beliefs stems from their determinacy *but not* primitiveness.

possible to build upon this sketch. Specifically, I want to reconstruct Kusch's notion of social identity or status, using my characterization of the normativity of descriptions as a guidepost. Social identities are, I argue, opaque coordinating factors or behavioral attractors that structure interaction.

To briefly recap Kusch's (2002a, 2013) position: social identities are a social kind. Other social kinds include money and marriage; for Kusch, money, marriage, and social identities are all constructed via the medium of language. Drawing upon Craig (1999), Kusch (2013) argues that the concept of knowledge originated due to the need to mark a specific type of social identity—namely, being a reliable informant. As a result, the concept of knowledge, like money and marriage, is itself a social kind. On this picture, social kinds act as a shared foundation for interaction. With regard to being knowledgeable, an interaction involving a person who holds the status will be structured such that he or she will be expected and trusted to provide information. Language provides a stability that imbues social kinds with a meaning that supersedes the particulars of any given social interaction. Social identities thus function as a mutually recognized or shared foundation for interaction.¹⁰³

The dilemma confronting Kusch's position is similar to the one that confronted Searle's concept of we-intentions, as diagnosed by Turner (§2.2). Either the shared social identity is so vague as to be vacuous, or it is so rich that it's wildly implausible that it's actually shared by individuals. My alternative account, following the spirit of Turner's modification of Searle,

¹⁰³ In accordance with this picture, Kusch (2002a, 2) states that the goal of communitarian epistemology "is to understand, rather than change, epistemic communities." Kusch thus sees his communitarian epistemology as *purely* descriptive. In the context of the testimony literature, Kusch is right to distance himself from normative approaches such as Goldman in which the details of interaction are often ignored as extraneous. But Kusch's characterization of communitarian epistemology's goal wrongly presupposes that description and understanding can be value-neutral. Embodied epistemology suggests that any change in understanding has normative implications.

jettisons the notion of a shared identity in favor of a more ephemeral and inchoate notion that is geared towards and also derivative upon actual engagement between individuals.

On my account, social identities are ambiguous descriptors whose meaning depends on the particulars of actual interactions, including each individual's unique experiential history. A social identity is an abstraction, which—like other kinds of descriptions—partition an interactive space, thereby modifying the range of relevant interactive opportunities. The range of recognized interactive opportunities is partly a function of an agent's prior interactions, on the basis of which she has refined her expectations when confronted with negative feedback. It is in this sense that social identities serve as coordinating factors or behavioral attractors: the former relating to the concept of participatory sense-making and the latter to Granic's account of aggressive patterns of behavior (§3.3).¹⁰⁴

Thinking of social identities as coordinating factors or attractors helps to explain a social status's relatively stable meaning without annihilating the idiosyncratic dimensions of social interaction. Salient differences in individuals' respective perceptions of a social identity can be smoothed out (or further entrenched) via the continuous feedback and refinement of encounters. There may even be crucial differences in individuals' understanding of a social identity that go unrecognized because circumstances have not yet brought them to the fore. This is the sense in which social identities are shallow or opaque. Agreement conceals pervasive underlying disagreements, but without undermining the role that social identities play in interaction.

To illustrate how a social identity can structure interaction, I offer the following hypothetical scenario, inspired by D'Agostino's federal model. Consider the effect that learning

¹⁰⁴ While social identities and linguistic descriptions appear discordant with De Jaegher and Di Paolo's (2007, 2008) own examples of coordinating factors, this mostly results from their rhetorical context. De Jaegher and Di Paolo's main objective is to counter cognitivism within the philosophy of mind. Furthermore, I am treating social identities in terms of how they influence social interaction, thus maintaining the priority that De Jaegher and Di Paolo give to the dynamics of engagement.

of social comparison pressures within epistemic communities could have on an individual researcher. She might, for instance, be less likely to remain reticent when another team member voices a point at variance with her own findings. When offering the dissenting viewpoint, she can think of herself as searching for a “hidden profile” as opposed to making trouble or attacking her team member. This difference can be expressed in terms of the contrasting roles of a team helper as compared to a trouble-maker.¹⁰⁵ The difference can have a decisive influence not only on whether the dissenter will share her divergent findings but also—granting that she might voice her concerns without having learned of social comparison pressures—influence whether the contribution will be constructive. The perspective of searching for a hidden profile and helping one’s team is a stronger rhetorical position than challenging a team member.¹⁰⁶

The relevant social identities in my example—“team helper” and “trouble-maker”—are more ephemeral than an idealized status like “being knowledgeable.” The hypothetical scenario thus diverges from Kusch’s discussion, yet I think such fleeting social roles draw attention to the formation of identities and thus the dynamics that underpin more stable ones. Accordingly, I propose that more stable and widely recognizable social identities are a product of the same general dynamics—an opaque abstraction that is continuously refined and reinforced via social encounters.

Rehg’s (2009) concept of cogency dovetails with the present account of social identity. Cogency, for Rehg, represents the many context-specific factors that determine whether an argument is found to be persuasive within a particular epistemic community. Possible factors

¹⁰⁵ This shift in perspective and consequent influence on behavior is a good example of how intellectual understanding possesses normative implications. In Ch. 5, I touch on this theme in relation to Gallagher’s (2005a) distinction between cognitive ownership and agency.

¹⁰⁶ Conversely, Rehg (2009) offers a case of a researcher who lacks sensitivity to the rhetorical dimension of enquiry and, as a result, is excluded from an epistemic community. I detail this case and the rhetorical dimension of enquiry in the next section.

include values such as openness, a person's argumentative temperament, and who receives credit. The indefinite range of possible factors is offset by the need for individuals to demonstrate their relevance to a particular situation. Rehg thus suggests that the normativity of scientific argumentation involves "microdynamics" specific to particular interactive contexts (67). In the next section, I bolster my critique of Fuller and Kusch's disembodied social epistemologies by detailing the contextual normativity of scientific argumentation.

4.5 Cogency and the Immanent Normativity of Epistemic Communities

The guiding theme of Rehg (2009) is "Kuhn's Gap," which refers to occasions when an argument is persuasive but not logically compelling. Rehg's concept of cogency relates to the "microdynamics" of persuasion in comparison to Kuhn's focus on macro-social institutional forces. Kuhn's merely superficial treatment of persuasion is evident from Wray's depiction of the evolutionary aspect of Kuhnian epistemology: when a scientific community confronts a theoretical dispute, either there will ultimately be a clear winner, or the community splits into separate disciplines (§4.2).

D'Agostino improves upon Kuhn by drawing on psychology and organizational studies, as detailed in §4.3. But D'Agostino considers these factors in terms of the institutional dynamics that either hinder or promote the communication of information, with the overarching theme being the discovery of a "hidden profile." The concept of shallow consensus comes closest to Rehg's concern with cogency, but it only denotes a negative dynamic—that persuasion occurs on the basis of very thin agreements—without detailing the process of reaching such agreements. The centerpiece of D'Agostino's account—in keeping with Kuhn's focus on macro-social institutional forces—is the federal model's two-axis model of epistemic communities.

With its treatment of microsocial dynamics, Rehg's (2009) theory of argumentation helps substantiate the link between the institutional forces discussed by Kuhn, Wray, and D'Agostino and the enactivist account of the cognitive subject presented in Ch. 3. Rehg begins with the common idea of cogency, namely cogency as the "convincing quality" of an argument (6). This open-ended definition enables the term to cover a range of different meanings, with the disparate meanings a reflection of cogency's context-dependence.

As one illustration of cogency's context-specificity, Rehg examines the process of collaboration and publication at Fermilab, a particle physics research laboratory. The case study specifically deals with the research and discovery of the top quark, which spanned the years 1993-1995.¹⁰⁷ Fermilab is an institution within which many different research teams collaborate. In connection with the top quark research, there were four separate teams: two teams counted lepton decays, one team counted dilepton decays, and a fourth team reconstructed the kinematics of the decay events (Rehg 2009, 169). To pool and publish the four teams' results, the writing process involved four social roles: convener, godparent, writer, and audience (171). Two people acted as *conveners*, whose task was to convene meetings and record their minutes. The conveners asked each team to appoint a *godparent*, whose task was to provide an independent critique of the paper drafted by the *writers* (the third role). After the draft passed the godparents, it was then meant to be presented to the *audience*, which consisted of all the Fermilab scientists who were not part of the four teams. The presentation to Fermilab as a whole acted as a final check before journal publication.

The case study exhibits the basic features of D'Agostino's federal model while also presenting some complications. The four teams worked in parallel, corresponding to the

¹⁰⁷ For the Fermilab case study, Rehg draws heavily upon Kent Staley's (2004) book on the discovery of the top quark.

horizontal axis of the federal model, yet only two used the same technique. The difference in methodology led to some controversy regarding the kinematics group's claims, which I return to below. There was also a complication with respect to the vertical axis: while the godparents were meant to provide an independent critique of the paper draft, they ended up participating directly in its writing. The break with agreed-upon procedure, while criticized, was justified on the grounds that it improved the quality of the paper (172). Hence, procedures and institutional structures were improvised upon when it was considered beneficial to the research.

The writing procedure—that is, the four social roles and three writing phases of draft, revision, and presentation to Fermilab as a whole—was formulated in the wake of having rejected a “four short papers” proposal. The rejection of this proposal—referred to within Fermilab as the “October massacre”—was due to the view that the papers were being rushed without sufficient vetting and motivated the additional second and third phases in the subsequently agreed-upon writing procedure. The perceived rush in the prior proposal was due largely to the kinematics team, who wanted to publish their evidence for the existence of the top quark before the other teams. Complicating matters, two of the team members (Garry Goldstein and Richard Dalitz) were outsiders brought in because of their expertise in kinematics analysis.

The tense situation brought to the fore the difficulties involved in the collaboration of different types of specialists. The kinematics team claimed to have conclusive evidence for the top quark but none of the Fermilab scientists outside of the kinematics team were fully qualified to judge the claim's technical details. The difficult circumstance highlights the wide range of factors involved in argumentation, including the already discussed dialectics of the writing procedure. The kinematics team's rush to publish conflicted with the values of openness and

thoroughness, which in turn prompted Fermilab scientists to formulate the four social roles and three phases of paper writing.

Rhetorical considerations also played a part in assessing the kinematics team's proposal. Krys Sliwa—the Fermilab scientist in charge of the kinematics team—described Gary Goldstein as “overexcited” (177). Goldstein's temperament—his *ethos*—suggested to Sliwa and some of his other colleagues that he lacked the capacity for responsible judgment, and he subsequently lost access to Fermilab's data (177). There was at least one additional major factor that influenced the assessment of the kinematics team, and it corresponds to the concern for credit attribution stressed by Kitcher (1990). If the kinematics team had published their paper first, then credit for the top quark discovery would have fallen primarily on scientists who were not part of Fermilab. At issue are not merely the selfish interests of Fermilab scientists but the credibility of Fermilab as a whole. Maintaining the credibility of the institution itself is crucial to its continued existence, a concern that also played a part in withholding data from an August 1993 conference presentation.¹⁰⁸

Rehg interprets the Fermilab case study as displaying an immanent contextualism. All three of the major factors—the dialectical, rhetorical, and credit issues—took on their particular significance because of the kinematics controversy. The dialectical values of openness and thoroughness, for instance, only became relevant argumentative considerations when the kinematics team sought to publish their findings before the three other teams. To further conceptualize this immanent dynamic, Rehg draws upon Habermas's notion of transcendental

¹⁰⁸ By withholding data from the 1993 Ithaca conference, the Fermilab scientists avoided prematurely claiming the discovery, which would have opened them up to objections they were not prepared to address. Alternately, if they had presented all of the data without claiming the discovery, they risked other researchers claiming the discovery first and thereby impugning Fermilab's reputation. For Rehg, it's important that Paul Tipton, the Fermilab presenter, both piqued the interest of his audience and demonstrated his competency by acknowledging that he had not shown conclusive data (171).

standards or pragmatic presuppositions of conversation.¹⁰⁹ Yet Rehg dispenses with the notion of a *transcendental* standard and takes them instead as a critical heuristic, thereby inverting Habermas as follows: rather than being regulative ideals for all conversations, values such as openness must be fought for by a particular agent in order for them to be meaningful in a particular context (e.g., the value of openness prompting the creation of a new writing procedure) (160).

Rehg's immanent contextualism complements embodied epistemology. The above controversy, for example, can be modeled using Bickhard's terminology. The kinematics team, from this perspective, created a situation in which the abstract values of openness and thoroughness were recognized by Fermilab scientists as relevant tools for restructuring the interactive space (i.e., the creation of the four social roles and three writing phases). In this sense, the values were not determinate blueprints but rather a means for differentially responding to a particular interactive context. In this regard, the values, like the temperament of the scientists, acted as coordinating factors.

Rehg's major contribution to embodied epistemology lies in his reconstruction of a tremendously complex form of normativity. Scientific argumentation is about as far from minimal cases of cognition as one can get, yet the contextualism of scientific argumentation suggests that normativity retains an immanent character. Rehg's theory of argumentation unpacks a kind of rationality long held dear by philosophers of mind and, in doing so, supplants the notion of objective rationality in favor of the context-specific notion of cogency. Ironically,

¹⁰⁹ See Habermas (1979) for his initial formulation of a "universal pragmatics" for communication. Habermas's (1984, 1987) Theory of Communicative Action builds upon this picture by arguing against narrow conceptions of rationality in the social sciences; one of Habermas's most important claims, in this regard, is that rationality is responsive to moral and evaluative claims, not just empirical facts. Although it departs from Habermas's notion of transcendental ideals and counterfactual mode of analysis, Rehg's immanent contextualism dovetails with the pragmatic current in Habermas's work.

one possible response to this interpretation of Rehg's theory is to claim that the Fermilab case study is *too* complex. That is to say, sub-atomic particles are such an extraordinary subject matter with only an extremely attenuated connection to humans' perceptual capacities that the normativity exhibited in the corresponding enquiries is not representative of more common epistemic endeavors and interactions. The idea here—to paraphrase Oliver Wendell Holmes, Jr.—is that tough cases make bad laws.

While I concede that the top quark discovery is an extraordinary case, it is *not* extraordinary to take the collaboration of diverse specialists as a fitting exemplar for epistemic interaction. Even though most people have not had education comparable to the training of a particle physicist, the experience of meeting and working with people whose skills are dramatically different from our own—and also to a large extent unknown—is a familiar and formative one. Analogous to the situation of Fermilab's scientists vis-à-vis the kinematics experts, it is common for individuals to deal with people whose claims they cannot directly assess. Yet it is crucial to negotiate such situations, to negotiate the pervasive and persistent differences between others and us in time- and resource-sensitive circumstances.

Taking the collaboration of diverse specialists as an exemplar for epistemic interaction signals a critical departure from both the testimony and translation views. The latter takes apprenticeship as a universal exemplar, wherein there is a succession of individuals whose skills resemble but productively differ from each other. On this picture, our agency is largely determined by how individuals have been sculpted by sensory stimuli. In terms of the present discussion, the basic issue with this picture is that all interactions are grafted onto the structure of mirroring, which introduces a superfluous and thus costly step to social interactions that are dissimilar to apprentice training such as the collaboration of diverse specialists.

For the testimony view, the exemplary aspect of epistemic interaction is, of course, a speech act—specifically one that involves the faithful reporting of objective reality. Oftentimes, testimony theorists begin with the most mundane, trivial, and dehydrated scenarios with the hopes of establishing a foundation for the interesting stuff. Alvin Goldman (1999, 23), for instance, establishes the objective character of knowledge using the example of whether milk is in the fridge. Quite simply, a person’s belief about whether there is milk and the fact of the matter are logically independent. Accordingly, knowledge is a correspondence relation between a person’s belief and an objective state of affairs. With this conception of knowledge in hand, Goldman casts the goal of social epistemology as increasing people’s stock of true beliefs. Knowledge is thus dehydrated in the sense that its practical connotations are taken to be of an entirely incidental and derivative nature. Likewise, social interaction is epistemically significant only as a means for passing beliefs along.

For the transformation view, by contrast, the contextual nature of knowledge entails that knowledge is intrinsically linked to practical connotations and ramifications. Knowledge is, first and foremost, skillful engagement with the world. With regard to social epistemology, one crucial implication is that the toy examples that are especially pervasive within the testimony literature are false starting points. The underlying methodology of extrapolating from clear-cut examples to the interesting stuff is misguided because it is the interesting and messy stuff that *transforms* us.

It should be stressed, however, that the transformation view does not discount the everyday lives of cognitive agents. Indeed, agents’ everyday lives are interesting precisely because of the endless ways in which our superficially routine behaviors are shaped by transformative epistemic experiences. Consider, for instance, the following odd but eminently

plausible scenarios: the accountant who maintains a spreadsheet that tracks his mate compatibilities, as if the value of a romantic partner could be inventoried like the property of a company; or the former army sergeant who still keeps a weekly “load list” for groceries, as if she were still requisitioning goods for a platoon; or the professor who itemizes his thoughts into numbered bullet points, as if he were in a classroom rather than a casual conversation. In keeping with these sorts of strange yet common examples, the Fermilab case study represents a more general shift in priority—favoring the analysis of innovative cutting-edge epistemic pursuits over trivial examples. Since many of our cutting-edge epistemic pursuits are massively complex community endeavors, one important consequence of the shift is to marginalize single mechanism or single purpose characterizations of social interaction. Such characterizations depend largely on extrapolating from the misleading simplicity of toy examples. In Ch. 5, I challenge another closely related obstacle to appreciating the multifaceted role of community-level cognitive processes—namely, manipulationist reductions of social interaction.

The present chapter has canvassed neo-Kuhnian social epistemology in order to identify salient community-level cognitive processes. Two of the most important dynamics have been the modular epistemic division of labor conceptualized by D’Agostino’s federal model of enquiry and, second, the immanent normativity conceptualized by Rehg’s argumentation theory. The former suggests that individual cognitive activity is intimately tied to larger community processes, while the latter suggests that enquiry involves an indefinite range of potential factors whose ultimate relevance are determined by the interaction of particular cognitive agents. Together, the two dynamics show that individuals’ epistemic agency is intimately bound up with though not determined by larger social contexts. In Ch. 5, I round this picture out by challenging the egoist presuppositions pervasive within cognitive science and social epistemology.

Chapter 5: The Encultured Mind

5.0 Introduction

Dominant within cognitive science and social epistemology is a largely assumed utility or manipulationist viewpoint regarding the subject's relationship with objects in its environment.¹¹⁰

The present chapter examines this manipulationist presupposition in the context of the so-called “four E’s” of cognition. The term 4E cognition refers to embodied, embedded, extended, and enacted cognition and the corresponding theoretical frameworks, which have all leveled trenchant attacks on traditional cognitivism and methodological individualism (Menary 2010b). One indication of the general progress made is the noticeable shift in polemical targets, with many 4E critiques now directed at other 4E theories. My critique of the manipulationist presupposition focuses on the extended view and two competing versions of the enactive view.

The manipulationist perspective underpins one of the most general pictures of the cognitive subject. John Sutton (2010) poses the problem as follows:

Are culture, artifice, and moral practice optional extras, merely dispensable surrogates which ride on top of the brain's own unchanged tendencies? Or are they instead in one form or another inevitable, structuring supplements which construct and maintain the biological process that they simultaneously and deeply transform? (212)

Sutton contrasts two basic ways of understanding how socialization and experience more generally impact the human mind: a core view in which socialization only augments innate capacities, and an encultured view in which a person's mind depends on experience—especially

¹¹⁰ The term “manipulationism” stems in part from Menary's (2007, 2010a) eponymous thesis, which I address in §5.5. Menary's thesis exemplifies a much broader approach in cognitive science and social epistemology, in which other people appear as just another epistemic resource to be exploited.

socialization—for structuring plastic innate capacities. Directly challenging manipulationism first requires, I argue, undermining the notion that a cognitive subject possesses an innate *and* fixed core. Once this is accomplished, I then examine the temporal directionality of first-personal agency.

The chapter contains two major parts: a critique of Andy Clark’s Extended Mind hypothesis (EM) and an analysis of three varieties of enactivism. In §5.1-5.2, I use EM and its associated functionalist framework as a symptomatic case of manipulationism and an exemplar of the core view. By shrinking the core down to a bundle of multiply realizable processes, functionalism represents what is perhaps the most plausible version of the core view but ultimately, as Di Paolo (2009) suggests, undermines the very notion of a mind. There have been a number of attempts—contrary to Clark—to apply EM to the social dimensions of cognition. In §5.3, I consider two of the most prominent attempts in this regard: Gallagher’s proposal for a critical cognitive science and Sutton’s historical cognitive science. While both offer insights and advances, I argue that overcoming manipulationism requires completely abandoning functionalism. Toward this end, I resituate Gallagher and Sutton’s proposals within the enactive movement.

Before directly linking my interpretation of enactivism (Chs. 3 and 4) to Gallagher and Sutton’s proposals, I canvas two other varieties of enactivism. I explicate Daniel Hutto’s Radical Enactive Cognition hypothesis (REC) and Richard Menary’s Cognitive Integrationism (CI) in sections §5.4 and §5.5, respectively. Although Hutto’s and Menary’s projects ultimately dovetail with each other, they reflect importantly different priorities: REC is an account of basic cognition founded on Hutto’s critique of representationalism, whereas CI is centered on Menary’s concept of cognitive practices. Both Hutto and Menary discuss the epistemic

importance of enculturation, yet I argue their accounts are hindered by manipulationism. Treating enculturation in terms of historical causal chains, Hutto and Menary conceive of the cognitive subject in strictly retrospective terms, as if the only epistemically significant aspects of the agent were those sculpted from evolved mechanisms and objective norms. In the final section, I argue that recognizing the prospective nature of agency opens up new dimensions of epistemic significance, including the productivity of *inchoate* goals that help to explain the profound epistemic importance of social interaction. The task of the chapter is thus not only to draw attention to the transformative dimensions of cognition, but to do so while not annihilating the intrinsically first-personal and future-oriented perspective of the cognitive subject.

5.1 The Extended Mind, Functionalism, and their Vanishing Cognitive Subject

Andy Clark's Extended Mind hypothesis represents one of the most plausible though still flawed articulations of the core view. In presenting its vision of the cognitive subject as a biotechnological hybrid, the Extended Mind hypothesis (EM) both extends the boundaries of the cognitive self while shrinking the innate and fixed core that it entails. One of my primary tasks in the next two sections is to identify the essential core implied by EM despite Clark's own at times conflicting characterizations.

The general framework of EM was discussed previously in §3.4 in relation to Clark's critique of enactivism. There I criticized Clark's functionalism for failing to account for intrinsic normativity and consequently perceptual salience. In the present section, I focus on three other aspects: on identifying the basics tenets of EM, on Clark's careful rhetorical framing of EM, and on explaining how EM relates to the functionalist movement in the philosophy of mind.

Examining the latter relationship raises the issue of the fundamental nature of the cognitive

subject; in the next section, I consider how Clark can respond to this issue and argue that it results in an unwelcome manipulationism.

The argumentative strategy motivating EM is driven by an intuition pump. Clark (2008b, xxvi) presents EM as an attack on the “bioprejudices” of the “BRAINBOUND” view, which holds that “the (nonneural) body is just the sensor and effector systems of the brain” (original emphasis). In their introductory article, Clark and Chalmers (1998) formulate the parity principle as the chief means for challenging brainbound intuitions:

If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process. (8; original emphasis)

If two operations serve the same functional role, then, the authors suggest, the only reason to count the neural manifestation as cognitive but not the external one would be the aforementioned “bioprejudice.” The two by now well-rehearsed examples of extension from the introductory article are of playing Tetris and the hypothetical Otto who suffers from Alzheimer’s and, as a result, depends on his notebook as a memory aid. In keeping with the rhetorical thrust of the intuition pump strategy, the Tetris and Otto examples serve less to illustrate the dynamics of extended processes than to enable Clark and Chalmers to dispatch some preliminary objections to EM.

Most pertinent to the present discussion is the portability objection, which stems from what Clark and Chalmers refer to as the vision of the “Naked Mind.” The objection consists of claiming that what counts as cognitive is strictly the “package of resources” that a subject can always bring to bear on a task, irrespective of the local environment (Clark and Chalmers 1998, 10). This picture of the Naked Mind closely relates to what I have labeled the core view. Despite rebutting the objection, Clark and Chalmers endorse the underlying picture: they write, “there is

something to this objection...the brain (or brain and body) comprise a package of basic, portable, cognitive resources” (10). The difference between EM and the Naked Mind position is thus largely terminological: both agree there is a suite of core processes, yet Clark and Chalmers believe that the cognitive label should not be restricted to just these processes. In the next section, I unpack why Clark is committed to the existence of innate and fixed core processes, but for now I want to note EM’s general consonance with the vision of the Naked Mind, specifically the shared view that the individual subject is the endogenous center of cognition. Clark and Chalmers think that the parity principle alone would incorporate too many external resources—would lose focus of the cognitive self—and so they formulate three supplemental criteria for further restricting what external resources count as constituent elements of cognitive processes. The three additional criteria are as follows: the external resource must be (i) a constant in the user’s life, (ii) easily accessible when needed, and (iii) trusted (17). The end result, Clark and Chalmers believe, is a more expansive picture of the cognitive subject that simultaneously avoids the excesses of panpsychism.

In the vast literature that EM has garnered, Clark has stuck closely to the rhetorical contrast between the brainbound and extended views. In terms of a positive argument for EM, Clark has offered a diverse array of extended cognitive processes, including detailed analyses of such coupling dynamics; this positive aspect of Clark’s argument is exemplified by the first part of his 2008 *Supersizing the Mind*. But when faced with criticism, Clark prefers instead a three-pronged strategy linked to his carefully crafted rhetorical position: first, focus on the contrast between EM and the brainbound view; second, note how the parity principle’s supplemental criteria retain an individual center of cognition; and third, attribute any fundamental disagreements to a conflict of intuitions. Missing from this general discussion is a careful

account of the fundamental nature of cognition and the cognitive subject, in particular the relationship between EM and the larger functionalist framework of which it is a part.

For Clark, the functionalist movement within the philosophy of mind addresses the most basic questions concerning the fundamental nature of cognition. Yet upon closer inspection, this argumentative division of labor fails due to the fact that Clark's version of EM is inconsistent with functionalism. To highlight this tension, I turn first to Clark's (2010) method for determining the coarseness of functional roles before assessing the general relationship between EM and functionalism.

After conceding that he lacks a precise standard for differentiating an individual's extended belief system from a "contingent" one, Clark (2010) encourages the reader to follow the intuitive path set by him and Chalmers. Clark writes,

Chalmers and I tend to favor a rather coarse notion of the required functional role in which all that matters is that the information be typically trusted and that it guide gross choice, reason, and behavior in roughly the usual ways ... in ways that would not cause constant misunderstandings and upsets if the agent were somehow able to join with, or communicate with, a human community. (62)

The coarseness that Clark refers to in the above passage is a matter of how tight the requirements should be on counting something as part of an individual's cognitive system. To this question, their intuitions favor the following answer: if an individual could usually succeed in employing a resource within a community, then it seems reasonable for an individual to rely on and trust the resource, which in turn favors treating the resource as more than a contingent coupling relation. To anyone who still resists following EM's path, Clark surrenders any hope of conversion: "nor do I see how to further argue this case with anyone whose intuitions differ" (62). More telling than any of its specifics, the passage encapsulates Clark's rhetorical strategy and also suggests how deeply presupposed the individualism of the core view is within EM.

Regarding EM's individualism, the topic of human community only arises in the above passage due to Kim Sterelny's (2004) critique of EM. Sterelny argues that when individuals employ extended processes, they must guard against informational sabotage by others, and as a result extended processes increase rather than decrease the cognitive load on their individual users. Rather than questioning Sterelny's Hobbesian depiction of human communities, Clark (2010, 61) simply suggests that epistemic agents only protect themselves against an "ecologically normal level of vulnerability." Hence, the "upsets" that Clark refers to above are in part cases where, by trusting an external resource, an individual is taken advantage of by another person. Later, in connection with Clark's 007 principle, I return to the Hobbesian character of EM's individualism. For the moment, I only want to suggest that despite Clark's careful rhetorical framing of EM, the question of how to employ functional analyses is not satisfactorily answered by appealing to intuitions, which brings us back to the issue of EM and functionalism's relationship.

To assess the relationship between EM and functionalism, it is first worth sketching some background to the latter. Common to all functionalist theories of mind is the multiple realizability thesis of mental states. The multiple realizability thesis states that "a single mental kind (property, state, event) can be realized by many distinct physical kinds" (Bickle 2013). Putnam (1967a) uses the experience of pain as an exemplar of a mental kind that is multiply realizable, as it is clear that many different animals share the experience. In presenting the multiple realizability thesis, Putnam's direct critical targets were mind-brain identity theorists and behaviorists; as an alternative, many including Putnam (1960, 1967a, 1967b) have used functionalism as a positive account of mentality that begins with multiple realizability.¹¹¹

¹¹¹ The relationship between Putnam and functionalism is complicated to say the least. Putnam (1988) reverses course, using multiple realizability to argue against functionalism on the grounds that the latter ultimately

In keeping with the multiple realizability thesis, functionalist theories of mind deny that mental states should be identified with a particular physical constitution. Accordingly, the central tenet of functionalism is that a mental state should be understood in terms of the role that it plays within a cognitive system—more precisely, the mental state’s “causal relations to sensory stimulations, other mental states, and behavior” (Levin 2013). As glimpsed above in Clark and Chalmers’ approach to identifying functional roles, one of the central difficulties of functionalism is how exactly to characterize the inputs and outputs of a cognitive system. Paralleling the notions of broad and narrow mental content, Ned Block (1990) characterizes two basic ways of carving up a functionalist cognitive system—namely, in terms of “long-arm” and “short-arm” functional roles, respectively. Long-arm roles correspond to broad mental content, since these functional roles include things from the world when identifying the inputs and outputs of a system, whereas short-arm roles only recognize inputs and outputs internal to an individual organism. The upshot of the short-arm approach is that it can straightforwardly explain how people in different environments can share the same mental state, since the mental inputs are not tied to particular external objects that invariably differ across environments. But the trade-off of the short-arm approach is that it is unclear how to account for subjects with very different physical constitutions, such as the oft-used examples of aliens and androids not to mention the physical differences across humans. Both feats—identifying the same mental state in two individuals who are in different environments and identifying the same mental state across humans or in a human and a very different creature—are significant appeals of functionalism and so it is vital to retain each. The tension is symptomatic, I argue, of the problematic conception of individuality within functionalism in general.

reduces mentality to behavioral dispositions. But even more recently, Putnam (2012) adopts a “liberal functionalism” in contrast to the machine functionalism that he previously defended.

Within functionalism, the problem of how to individuate cognitive agents has received less attention than the question of how functional processes relate to each other. The literature surrounding Fodor's (1983) modularity of mind hypothesis is emblematic of how the existence of a viable concept of individuation is often presupposed. Fodor's modularity hypothesis asserts that the low-level subsystems—putatively responsible for the transduction of sensory stimuli—are relatively automatic functional processes in contrast to higher-level processes, such as those involved in belief-formation. That the cognitive system should be identified with a single human subject is, as Fodor (1986) argues, a metaphysical position justified by the methodology of the various sciences, with psychology being the relevant science in the given context.

The debate over modularity has focused on the question of whether high-level cognitive processes can also be characterized in terms of functional modules. Most advocates of “massive modularity,” such as Carruthers (2006), are invested in evolutionary psychology. One advantage of this general approach is that the principle of natural selection offers an empirical answer, albeit a contentious one, as to why the individual organism should be identified as the basic unit of the cognitive system.¹¹² I turn now to Daniel Dennett, who has elaborated in detail the nature of the functionalist subject and the problem of individuation from the perspective of evolutionary theory.

Dennett has seen through and embraced what for many, including many functionalists, are unwelcome implications regarding the fundamental nature of the cognitive subject. While Dennett (1976) addresses the issue in terms of personhood, I will limit my discussion to Dennett's (1981, 2001) discussion of cognitive subjectivity. Dennett (1981) draws upon the brain-in-a-vat trope in order to illustrate how functionalism undermines some of the most basic

¹¹² Fodor has been highly critical of Darwinian-informed functionalism, which includes EM. Fodor (2001) attacks the massive modularity thesis, while Fodor and Palmarini (2011) challenge the principle of natural selection, the very foundation of Darwinian-inspired functionalism and much else.

intuitions concerning the identity of the thinking subject—specifically, that the subject is a spatially delimited entity and that an individual mind is unique. In the thought experiment, Dennett’s brain is envatted so that he can control his body as it/he retrieves a radioactive warhead. As the scenario becomes more extreme, Dennett comes to find that the technicians have created a functionally identical duplicate of his original brain, which he can switch to and from seamlessly. Tellingly, in reflecting on the thought experiment, Dennett does not consider the duplication of a mind implausible so much as the idea that, for any extended period of time, two minds could be kept in sync by being fed the same set of inputs (323). The issue, as Dennett sees it, is that any slight difference in the timing between the inputs would quickly compound into large computational differences due to the nature of massively parallel computing processes. Hence, the overall moral of the story is that functionalism implies that thinking subjectivity itself is multiply realizable.

For those averse to brain-in-a-vat scenarios, Dennett (2001) takes a more empirically-informed approach to the same question. At issue in this article is the source of creativity; Dennett takes two computer programs—IBM’s Deep Blue chess program and David Cope’s EMI musical composition program—as a bellwether for how artificial intelligence is going to thoroughly undermine the mystification of creative genius. Dennett sees all creativity, and by extension intelligence, as the exploration of a vastly large design space. Learning is simply the acquisition of ways for more efficiently exploring the design space, which in the case of chess consists of memorizing patterns that have been culled from previously played games (16). One of the most important heuristics, on this view, is the very concept of self, which Dennett characterizes as a placeholder for the “largely unknown and unknowable microprocesses as well as the history that set them up” (25). That is to say, there is no self, though it is a useful and, at

this point, perhaps indispensable convention for predicting and explaining the various “microprocesses” at work. Part of Dennett’s point echoes though does not explicitly acknowledge EM: Dennett believes that the boundaries of authorship should be extended (23). The connection is more than incidental—Clark (2003) names Dennett as one of the two biggest influences on EM—but Clark resists Dennett’s complete deflation of the cognitive subject, as is clear in Clark’s (2010) concluding remarks.

After characterizing EM’s cognitive subject as a motley assemblage, Clark (2010) raises the question of whether the extended mind debate undermines the very notion of mind. Might the ultimate lesson be that the “idea of the mental is terminally unstable” and that we ought to “eliminate the mind” (63). But rather than taking the Dennettian route of simply accepting that the cognitive self is a useful convention for understanding a tightly interrelated nexus of functional processes, Clark posits EM as the beginning of a new science of the mind that can better illustrate the “biotechnological openness of the very ideas of mind and reason” (64). That is, the mind and the individual self do exist—as opposed to just being useful heuristics—only they are a biotechnological hybrid.

Given the resistance to Dennett’s account by functionalists such as Fodor, I begin the next section by considering Sprevak’s (2009) critique of EM. Sprevak uses the most basic tenets of functionalism to undermine Clark’s restricted conception of cognitive extension. Sprevak’s critique pierces through Clark’s rhetorical framing of EM and thereby helps to draw out some of the position’s more elusive aspects.

5.2 Uncovering the Adaptationist and Manipulationist Core of the Extended Mind

Mark Sprevak (2009) corroborates Dennett’s picture of the vanishing subject, showing how the basic tenets of functionalism entails at the very least a similar picture. The first step of Sprevak’s

argument is to establish that any worthwhile variety of functionalism is committed to extended cognition. A number of EM's critics have endorsed Clark's parity principle yet insist that it is a contingent historical fact that there have been no observed extended processes that satisfy the standard (e.g., Adams and Aizawa 2008, Rupert 2004). The objection raises the previously discussed issue of the coarse-ness of functional analysis: these critics are adopting a very fine-grained, "short-armed" analysis. One of Clark's responses, which Sprevak endorses, is that the critics use so fine-grained an analysis that it undermines functionalism's ability to account for radically alternate forms of intelligence, such as possible Martians or artificial intelligence (Sprevak 2009, 508). Maintaining these intuitions, which are pivotal to functionalism's appeal, requires a coarse enough standard that would also include extended cognitive processes.

Having established that functionalism entails extended cognition, Sprevak turns to Clark's restricted conception of extension. As noted previously, Clark supplements the parity principle with three criteria: reliability, accessibility, and trustworthiness. Turning Clark's own point against himself, Sprevak (2009) notes that any grain coarse enough to would include a possible Martian intelligence entails an unrestricted parity principle, as a Martian's native intelligence might fail on any or all of the supplemental standards (514-15). Furthermore, there are cognitive processes physically internal to humans that already fail the criteria (516). Hence, functionalism does entail extended cognition, but a more radical notion of extension than Clark's preferred version.

Sprevak understands why Clark wants to resist radical cognitive extension: radical extension is so inclusive that mentality loses its meaningfulness, the same issue that Clark (2010) raises when he asks whether the extended mind debate has eliminated the mind. Acknowledging Sprevak's critique, Clark (2011) even concedes that there is little hope of an empirical resolution

to this question. Since the issue applies equally to EM and the aforementioned critics, Clark brackets the issue and simply encourages the reader to continue attending to the tight coupling dynamics between cognitive subject and world in the hopes that the question will be resolved at some future point. While I am sympathetic to Sprevak's assessment of EM, I think that there are alternative accounts of functions—specifically Ruth Millikan's etiological account of *proper* functions—that mitigate his concerns. After articulating Millikan's position and how it addresses radical extension, I will then examine the extent to which the etiological account of functions buttresses EM.

Millikan's (1984, 1989, 2002) account of proper functions concerns the normativity of purposes. Millikan claims that the normativity of all functions depend on a device's *history* as opposed to its present dispositions. In the case of biological devices, the history is an evolutionary one: a biological device's proper function is the product of natural selection, having aided ancestors' survival at some point in the organism's evolutionary past (Millikan 1984, 28). One of the primary insights of the concept, according to Millikan, is a naturalistic framework that accounts for cognitive norms; indeed, Millikan (1993) argues that all cognitive norms are biological norms ultimately rooted in natural selection. While there are a number of theories that offer compelling empirical accounts of cognition, Millikan (1993, 3) notes that these accounts have difficulty handling defective cases, often altogether excluding them from the category of cognition. By contrast, Millikan differentiates between successful and dysfunctional cognitive acts based on which are selected traits. Importantly, not every instance must be directly explained in terms an evolutionary history, as there are “relational proper functions,” which derive from “direct” proper functions but, unlike the latter, do not correspond to an innate controlling mechanism (Millikan 1984, 47; 2002, 125). Millikan's biofunctions thus encompass

adaptive behavior in addition to directly selected traits. With this conceptual apparatus in hand, let us revisit the issue of radical extension.

To illustrate the unpalatable consequences of radical extension, Sprevak (2009) uses the example of a book. When I hold a book, radical extension entails that all of its contents are a constitutive part of my mind; radical extension is simply the unadulterated application of the parity principle. Suspending for the moment the question of whether the etiological account is compatible with EM, Millikan's position offers a promising even if limited response, differentiating between adaptive and non-adaptive coupling relations. A coupling relation, on Millikan's view, possesses a normative function only if it has changed the conditions under which natural selection has operated on humans. From this perspective, the agent-book relation is historically too recent for it to satisfy Millikan's standard. The Millikanian-EM response thus deflects Sprevak's example of radical extension, denying that the agent-book relation possesses a proper function.

Importantly, even though the standard for cognitive extension is considerably more difficult to satisfy, Millikan's position does not rule out all cases. Richard Wrangham's (2009) cooking hypothesis, for instance, suggests a suite of extended cognitive skills that possess proper functions. Wrangham is concerned with the emergence of social structures necessary for cooking—given the many interdependent skills involved and the need to pass them along to new generations—and the dramatic impact that cooking subsequently had on hominid anatomy.¹¹³ Although humans' intestines do not thereby possess a *cognitive* function, there are a number of epistemic skills involved in cooking that do, skills that are extended in the sense of both social

¹¹³ Wrangham (2009, 44) argues that the nutritional gain from cooking enabled a shrinkage in gut size, noting that homo sapiens' gut is only one-third of the surface area and sixty percent of the weight of what should be expected given the relative proportions of other primates. The selective advantage of the gut shrinkage consists in homo sapiens' greater mobility.

structures and physical artifacts. In addition to the proper cooking of food itself and the proper tool use implied therein, there are right and wrong ways to create and maintain fires, store and transport food, and so on. For the Millikanian, the proper function in each case is whatever contributes to nutritional gains, dysfunction whatever detracts. Millikan's concept of proper functions thus blocks *radical* extension, without blocking forms that have reshaped an organism on an evolutionary timescale.

Let us take stock of how Millikan's position measures up to EM and whether it preserves the general appeal of functionalist theories of mind. Beginning with the latter issue, Millikan can easily account for salient differences in standards for extension between humans and Martians. Such differences need only be explicable in terms of the differing evolutionary histories of humans and Martians. It may of course be epistemically difficult to provide the respective histories but this does not impeach the plausibility of their relevant differences. The case of artificial intelligence is more complicated: for Millikan, AI would only possess derived functions until enough time had passed for the process of natural selection to effectively "reshape" it. Hence, Millikan's concept of biofunctions preserves the Martian intuition, avoids undermining the case of AI, while simultaneously defanging the issue of radical extension. Millikan's etiological account does this by constraining the multiple realizability thesis, at least insofar as cognition is understood to entail the normative significance of epistemic error.

Taking up the issue of how EM and etiological biofunctions compare, the most glaring difference concerns the emphasis that Millikan places on the historical dimension of cognition. The scant attention given to history within EM begins with the introductory article, where Clark and Chalmers's (1998) contrast their "active externalism" to Putnam's (1975) and Burge's (1979) passive meaning externalism. For Putnam and Burge, distal historical causes determine

the semantic content of an individual's mental states. Referencing Dennett's (1991) distinction between cognitive vehicles and contents, Clark (2008b, 76) later builds upon the contrast with meaning externalism, clarifying that EM is a hypothesis strictly concerned with cognitive vehicles as opposed to semantic content. Yet Millikan's concept of biofunctions suggests that the historical dimension of cognition is not limited to passive meaning externalism, as evolutionary history on her account grounds all cognitive norms. If a refinement in cognitive vehicles results in the corresponding individuals possessing more true beliefs, as EM emphasizes, then the change is adaptive and would thus constitute a derived proper function. Such details are absent from but not inconsistent with Millikan's framework. And, as illustrated by the response to Sprevak, the historical dimension can serve a crucial role for EM, treating radical forms of cognitive extension as defective cases.

Furthermore, as with Millikan's account of functions, the principle of natural selection is integral to EM. Its role was glimpsed in the Clark and Sterelny exchange (§5.1): their broadly Hobbesian depiction of human community stems from an adaptationist and egoist interpretation of evolution. Nor is Clark's response to Sterelny an off-the-cuff remark; the adaptationist view is sprinkled throughout Clark's writings. In his 1991 book *Microcognition*, for instance, Clark presents the "007 Principle" as a thumbnail of cognitive evolution. Strongly foreshadowing EM, the principle states that in general creatures will not store or process information using internal media when they can more efficiently use the environment (64). In other words, organisms tend to know only as much as they need to. Similarly, Clark (2008b, 13) presents the "Principle of Ecological Assembly" in which the "canny" cognitive subject "tends to recruit, on the spot, whatever problem-solving resources are available with a minimum amount of effort." Throughout Clark's writings, the environment (including other people) only appear as resources

to be exploited; hence, the Hobbesian view of human community. Likewise, Clark's understanding of evolution is adaptationist in the sense that it treats the principle of natural selection as the only epistemically consequential aspect of evolution, as entailed by claiming that increased efficiency is the general evolutionary trend of cognition.

Given the central role of natural selection to EM's general framework, incorporating Millikan's account can be seen as a clarification of Clark's position as much as a substantive revision. The upshot of the synthesis is, of course, that it enables Clark to address Sprevak's (2009) radical extension objection. But, as evidenced by the preceding paragraph, it comes at the cost of positing a manipulationist core underpinning the cognitive subject. Importantly, Clark has resisted explicitly endorsing an essential core, even in places characterizing the cognitive subject as if there were no core. Clark (2001), for instance, embraces neural constructivism in order to account for the distinctive character of human rationality. Clark writes, "it is a mistake to posit a biologically fixed 'human nature' with a simple 'wrap-around' of tools and culture. For the tools and culture are indeed as much determiners of our nature as products of it" (138). Although the passage explicitly denies a fixed human nature, Clark's point is that most particular cognitive skills are not hardwired in at birth.¹¹⁴ The passage continues, "we humans are natural-born cyborgs, factory tweaked and primed so as to participate in cognitive and computational architectures whose bounds far exceed those of skin and skull" (138). As evidenced by the design language—"factory tweaked and primed"—even when Clark is emphasizing neural plasticity, he retains the manipulationist presupposition. This dovetails with Clark's preferred

¹¹⁴ David Hull (1986) argues that Darwinian evolution entails that all biological species lack a species essence, including humans who lack an essential nature.

neural constructivist story in which the substantive aspects of plasticity occur in the newer parts of the brain.¹¹⁵

EM represents perhaps the most abstract and thinnest version of the core view. Rather than a particular cognitive skill, the fixed and innate core consists of a primordial and multiply realizable drive to survive, from which more particular intentions derive and thereby the acquisition of particular skills. EM's core is adaptationist in the sense that it appeals to the principle of natural selection as its historical origin and sole evolutionary mechanism, while manipulationism more precisely designates particular manifestations of adaptationism—the subject-object relationship of tool exploitation—within EM's synchronic analyses. The significance of the adaptationist and manipulationist core can be seen by revisiting the issue of perceptual salience discussed in §3.4. For Clark (2008a), understanding sensory stimulation in the brute mechanistic terms of “pressing the flesh” is unproblematic precisely because the adaptationist drive to survive and its experiential derivatives ensure that the cognitive subject is not overwhelmed or even paralyzed by raw sensory input. Clark, of course, stresses an individual's development of extended processes for filtering information but this only pushes the issue of salience back one step: what initially motivated or mattered to the organism that spurred the development of the extended process?

The manipulationist viewpoint is not unique to EM but is rather a largely unargued presupposition within cognitive science and social epistemology. The Sterelny-Clark exchange is representative in this regard of how conflicting theories often battle over which framework provides the most compelling manipulationist or resource-utility account. Another instance is Clark's exchange with John Haugeland concerning the issue of social normativity (Clark 2002,

¹¹⁵ Clark cites Quartz and Sejnowski (1997), who describe the neocortex and prefrontal cortex as an “organ of plasticity.” These are the more evolutionarily recent neural structures.

Haugeland 2002). Advocating a position similar to Steiner and Stewart's structural normativism (§3.5), Haugeland criticizes Clark's cognitive incrementalism, arguing that humans' concern for the norms of truth and goodness marks a divide from other animals (30).¹¹⁶ Despite thus asserting that social relations between people are qualitatively different, Haugeland's case rests on avowedly manipulationist considerations. Contrasting social norms to Richard Dawkins's notion of a *meme*, Haugeland writes:

A signal advantage of conformism-enforced norms is that the structures of a community can rely on the fact that almost all its members will abide by almost all its norms. ... So norms, unlike memes, can sustain an elaborate, interdependent social structure, with all of the obvious advantages that it makes possible. (32)

The qualitative difference between norm-governed human communities and other types of relationships, according to Haugeland, is not that the latter is only reducible to manipulationist considerations but, on the contrary, that the former (social norms) possess greater instrumental value. While Haugeland's account suffers from the same issues as Steiner and Stewart's structural normativity—namely, the two-horned dilemma of vacuity or extreme implausibility—Clark's adaptationist core is a more elusive target. After all, it could be the case that other people *are* only epistemically significant as tools to be exploited. Hence, the reason I chose EM as exemplar of the core view and the manipulationist presupposition is because it offers perhaps their most compelling defense, one that ultimately pushes the former to its breaking point. In the next section, I examine a group of theorists that push EM past its breaking point, appropriating it for analyses of social cognition that surpass Clark's own version.

¹¹⁶ It is important to note that Haugeland's conception of norms differs greatly from Millikan's etiological account. Like Steiner and Stewart, Millikan would consider Haugeland's account an instance of "meaning rationalism," which her etiological account serves as an alternative to.

5.3 Socializing EM: Insights, Advances, and Some Intractable Difficulties

In their introductory article, Clark and Chalmers (1998) acknowledge the in-principle possibility of social extension but consider it only a peripheral phenomenon. In a cursory paragraph near the article's end, Clark and Chalmers list an “unusually interdependent couple” and various contractual relationships—such as having a secretary or an accountant—as possible instances of social extension (17). Among the vast literature that EM has engendered, a number of theorists have challenged this narrow view of social epistemic relationships. In the present section, I begin with two separate but similar critiques of EM: both Shaun Gallagher and John Sutton argue that EM entails a much richer account of social interaction than that proffered by Clark and Chalmers. While Gallagher and Sutton both offer valuable insights, I argue their positions remain too closely tied to EM, as evidenced by Sutton leaving the door open for what he calls “third wave” EM. In closing the section, I compare the second and third waves of EM, noting some of the remaining difficulties confronting Gallagher and Sutton's accounts, which I tackle later in the chapter. To begin, I consider Gallagher's “overextension” of EM.

Over a series of articles, Gallagher presents the concept of a mental institution, drawing heavily on EM's parity principle (Gallagher and Crisafi 2009, Gallagher 2011, 2013). In a recurring example concerning the legal system, Gallagher has the reader consider three different problem-solving scenarios: (1) a person named Alexis formulates a decision after having been given a set of facts and a collection of evidence, (2) in addition to the facts and evidence, experts give Alexis three questions and a set of possible answers that she may choose from; (3) in addition to the preceding, the experts instruct Alexis to follow a set of pre-established rules for how to answer the questions (2009, 47; 2011, 114; 2013, 6). Despite the drastically different decision processes, Gallagher estimates the cognitive effort to be roughly equal for each case. As

Alexis works less, the labor is accomplished elsewhere: in the second case, by Alexis and the experts; in the third, by Alexis, the experts, and also the people who created the rules. Shorn of its supplemental criteria (§5.1), the parity principle suggests that the legal system should be considered part of an extended cognitive process. The legal system is thus a mental institution, other examples including museums and universities.

The primary obstacle to a properly socialized EM, from Gallagher's perspective, are the criteria of reliability, access, and trust, though Gallagher also believes that the parity principle needs to be constrained. As an alternative to Clark's criteria, Gallagher (2013) proposes a Lockean notion of ownership. Based on this standard, cognitive extension is a matter of how much work an individual has invested in a system and whether the individual is engaged in "the right way" (9). Hence, while no one "owns" a legal system, individuals are a part of the mental institution directly proportional to the extent of their participation in it, with experienced professionals (in this case, lawyers and judges) being at the extreme end of involvement. One of Gallagher's key points is that there are a number of cognitive skills that legal professionals have mastered that ought to be understood in the context of the legal system and may even be useless and meaningless outside of it. In a later section, I return to Gallagher's conception of cognitive ownership in order to delineate and assess different varieties of enactivism.

Despite being the primary basis of the mental institution concept, Gallagher (2013) does have some reservations about the parity principle itself. Gallagher notes, for instance, that the principle retains a Cartesian element by basing the standard for cognition on whether a process could occur within the head (5). Gallagher's thoughts on EM, in fact, have become progressively more critical, to the point that his use of the parity principle in developing the concept of mental institutions appears in retrospect to be more an immanent critique of EM than a positive

argument. Later, I return to Gallagher's development of an enactive account of intentionality; but, for the moment, I want to focus on Clark's response to the claim that the parity principle is Cartesian. To this specific criticism, Clark has clarified that the parity principle was not intended to establish causal identity but rather to "engage our rough sense of what might belong to the domain of the cognitive" (Kiverstein and Clark 2009. See also Clark 2007, Clark 2010). John Sutton (2010) masterfully traces this dialectic, designating Clark's use of the parity principle as "first wave" EM and contrasting it to a "second wave."

The definitive principle of the second wave, as Sutton defines it, is the complementarity principle. In his analyses of extended coupling processes, Clark has drawn attention to the very different but complementary roles that external and neural resources can serve; Clark notes that in such cases the brain need not "waste its time replicating such capacities" (Clark 1997, 220; Sutton 2010, 205). The external resources in these cases are important precisely because of the different causal roles they serve, roles that may never have been neurally feasible. While parity and complementarity are not inconsistent, Sutton argues that the two waves of EM are in tension. Pushing back against Clark's clarification, Sutton (2010) observes the general effect that the parity principle has on analyses of cognition: "when certain criteria are met, parity suggests, we shouldn't care if exograms rather than engrams are involved in the production of intelligent behavior" (198). The abstract perspective of the parity principle thus tends to "flatten" out the differences between internal and external cognitive resources. Even in light of this criticism, like Gallagher, Sutton is more charitable to EM than he ought to be given that his proposal for a historical cognitive science is at odds with the functionalist framework of EM, especially Clark's adaptationist version.

To highlight the tension between EM and Sutton's historical cognitive science, despite Sutton's mostly conciliatory stance, I turn now to Sutton's own memory arts case study. The memory arts process of memorization involves constructing an elaborate visual memory of an artificial structure—whether an architectural structure like a palace or a symbolic structure like an alphabet—then filling the structure with memories. Because the artificial structure provides discrete locations, locations with no intrinsic relation to the semantic content of the memories, the memories are stable and can be recalled at will (Sutton 2010, 208-11). While this mnemonic practice was found in antiquity and continues into the present day, Sutton is concerned with its prevalence in medieval Europe, the specifics of which problematize EM.

The medieval practice of *ars memoria* does not fit within Clark's adaptationism. As noted above, Clark recognizes that there are significant forms of complementarity—as opposed to merely causal isomorphism—between internal and external epistemic resources, but he understands complementarity strictly in terms of resource maximization. By contrast, in the most literal terms, *ars memoria* is radically inefficient, as it involves the laborious internalization of a text rather than resting content with it being offloaded onto the environment. Clark could stick to his functionalist guns, claiming either that there is an overall gain in cognitive ability or that the practice's popularity was an aberration that eventually died off due to inefficiency. But such a saving move is fanciful when compared to Sutton's own explanation.

As quoted in the chapter introduction, Sutton (2010) argues the mind should not be understood in isolation from culture. With regard to the present example, the memory arts gained prominence in medieval Europe due to a confluence of religious, technological, and scientific factors. The religious context laid great emphasis on maintaining the integrity of Scripture. Due to technological limitations texts were scarce, and contemporary science attributed humans'

fallible memory to animal spirits. Within this historical context, building memory palaces of Scripture was a pressing moral endeavor, not a means for maximizing cognitive resources. Yet Sutton thinks of his historical cognitive science as deepening EM's most basic insight, not as a contradictory framework.

Sutton does not directly address EM's functionalist framework, focusing instead on specific examples of extended coupling processes. Pushing EM further than Clark, Sutton (2010) argues that the pervasiveness of extended processes suggests that the brain is essentially a "leaky associative engine" (205). The brain is so leaky that it needs socialization in order to mold the brain's plasticity into useful cognitive skills, the socialization in the memory arts example corresponding to the aforementioned religious, technological, and cultural factors. In this manner, Sutton minimizes the tension between his picture of the encultured mind and EM. It would seem all that is missing from EM is attention to how outer media reciprocally shape inner biological mechanisms.¹¹⁷ It thus appears as though Sutton's conception of enculturation strengthens EM, filling in more of its picture of the cognitive subject.

In order to counter Sutton's conciliatory stance, I now return to my claim in the previous section that a viable functionalist theory of mind entails a manipulationist core. This core is antithetical to treating the social as anything other than a contingent means for objective epistemic ends. This is nowhere more evident than in places where Clark concedes that he has not given as much attention to social factors as they deserve. Clark (2011), for instance, discusses Edward Hutchins's (2008) characterization of cultural practices as distributed cognitive systems. Clark admits that for humans most cases of ecological assembly involve "long, hard-won, chains of cultural innovation and transmission" *but* Clark goes on to claim that EM

¹¹⁷ As noted in the previous section, Clark (2001) endorses neural constructivism, though, as I argued there, his appeal to constructivism still presupposes an innate and fixed manipulationist core.

concerns cognitive processes operating in the “here-and-now” (459). Similar to his dismissal of Putnam and Burge’s semantic externalism (§5.1), Clark believes that EM targets a different timescale than cultural practices. By treating synchronic and diachronic factors as isolable in this way, he severely limits the possible role of social factors. Likewise, Clark can bracket Sutton’s historical cognitive science, treating it as peripheral to his vision of a science of the mind. But while Clark rightly notes the role of multiple timescales, he neglects the intersections and interactions between them, which are pivotal for understanding what the manipulationist presupposition fails to capture. In the chapter’s closing section, I link the notion of complementarity and cognitive ownership to Di Paolo’s enactivist account of agency (§3.1), arguing that epistemic agency inherently involves diachronic and first-personal dimensions. Without an enactivist account of agency, both Gallagher and Sutton’s appropriations of EM remain prisoner to EM’s functionalist framework. The instability of Gallagher and Sutton’s positions and the robustness of the functionalist framework is evidenced by what Sutton (2010) refers to as “third wave” EM.

After having presented second wave complementarity, Sutton leaves the door open for a third wave of EM. Sutton characterizes this possible third wave as “a deterritorialized cognitive science which deals with the propagation of deformed and reformatted representations and which dissolves individuals into peculiar loci of coordination and coalescence among multiple structured media” (Sutton 2010, 213). In short, rather than positing a very leaky cognitive subject, the third wave flips the perspective, treating the cognitive subject as nothing more than a locus of propagating representations. Mason Cash and Michael Kirchhoff both adopt the third wave proposal, arguing that it provides a richer framework for addressing the cultural and ethical dimensions of cognitive processes (Cash 2013, Kirchhoff 2012, 2013).

Cash's (2013) main thesis consists of an account of responsibility, which he offers as an alternative to Gallagher's aforementioned notion of cognitive ownership. Drawing upon Daniel Dennett's (2003) account of intentionality, Cash argues that "a socially and physically distributed cognitive process counts as mine if it is appropriate to hold me responsible—to blame me or praise me, punish me or credit me—for the ideas or actions produced by this process" (Cash 2013, 62). A cognitive process, on Cash's account, counts as an individual process based on how people hold each other responsible, a process tied to communal norms. Like Dennett, Cash believes that this approach renders superfluous investigations into the "true nature of agency" (68). The immediate problem, however, with Cash's proposal is that it is founded on an erroneous critique of Gallagher's notion of Lockean ownership.

Gallagher's (2013) conception of cognitive ownership represents, for Cash, the limitation of second wave EM. As noted above, Gallagher's conception of cognitive ownership measures cognitive extension in terms of an individual's active involvement, stressing how the individual and her environment reciprocally shape each other. The issue, as Cash sees it, is that a reciprocally transformative relationship does not explain why a process is "owned" by a particular individual when the process is distributed over a network of institutions, practices, and other agents. Cash argues that the idea of ownership is fundamentally flawed, as it begins with individuals and works its way outwards. Cash's model of responsibility, by contrast, suggests beginning with social norms of accountability and then treating individuals as the product of these norms and associated social practices. In terms of the general third wave EM template, social norms of accountability are the deterritorialized processes that produce the individual loci.

There are two interrelated issues with Cash's (2013) critique of cognitive ownership. First, Cash conflates Gallagher's understanding of cognitive ownership with the generic sense of

ownership as an exclusive property relationship. Gallagher's notion of ownership, however, is phenomenological, as evidenced by how he distinguishes it from the first-personal feeling of agency (Gallagher 2005a, Chs. 2 and 8). In this context, both ownership and agency refer to a first-order pre-reflective awareness (174). In normal cases of willed action, the experience of ownership and agency coincide, whereas with involuntary movement, an individual "owns" the movement but does not feel responsible for initiating it (Gallagher 2005a, 56).¹¹⁸ On this phenomenological rendering, ownership refers to the feeling of own-ness. In the context of EM, Gallagher's concept of Lockean ownership thus does not entail an exclusive sense of property ownership but instead the first-personal experience of being heavily involved in a particular cognitive process. The second issue with Cash's critique is another perspective on the first: Cash equates the primitive significance of the first-personal perspective with a restrictive or substantial notion of individuality. As argued in §3.4, the first-personal perspective—when understood in terms of the enactivist conception of intrinsic normativity—addresses the issue of perceptual salience without creating a spurious gap between individual and environment (Weber and Varela 2002, Di Paolo 2005). Having addressed Cash's critique, I turn now to Michael Kirchhoff's account of third wave EM, which includes a more extensive critique of second wave complementarity.

Kirchhoff's most distinctive contribution to the third wave EM proposal is his emphasis on the diachronic dimension of epistemic agency. Employing Dynamical Systems Theory, Kirchhoff (2013, 20) notes that "all cognitive processes unfold in time continuous dynamical systems." Transactive memory systems, in which two or more individuals articulate a shared

¹¹⁸ The significance of the distinction between the feeling of ownership and agency is apparent in various pathological cases, such as schizophrenia. A schizophrenic, Gallagher (2005a) argues, initiates an action, is aware through sensory feedback of having performed the act—thereby giving them a sense of ownership—yet still lacks the feeling of agency (178).

memory, is a particular example of such a process, given that interrupting it undermines the creation of a shared memory (35). Yet the widespread use of synchronic analyses of cognitive systems has obscured the significance of temporal continuity, a problematic tendency that Kirchhoff attributes in part to the influence of analytic metaphysics (6). Clark (1997), for instance, stresses continuous reciprocal causation, in which the outputs of a cognitive process are fed back into the cognitive system as an input. But, as illustrated above, Clark lumps such processes under the grossly oversimplified category of the “here and now,” contrasting them to such distal historical chains as cultural practices and the evolutionary past. Clark’s contrast is predicated on treating all inputs to a cognitive system as discrete causes, whose distinct efficacy is isolable from larger processes such as continuous reciprocal causation. In opposition to Clark, Kirchhoff’s claim that cognitive processes are *continuous* dynamical systems denotes that the processes cannot be broken up into discrete points.

Kirchhoff (2012) shows how the diachronic character of epistemic agency entails that enculturation is integral to cognitive processes. Using the examples of mathematical reasoning and phoneme perception as exemplars, Kirchhoff (2012) concludes:

Certain neural transformations take the form of a re-shaping or reformatting process, where diachronic development in socio-cultural practices significantly transforms a subset of the brain’s cortical representations from primarily non-linguistic to linguistic. (300)

Comparing Kirchhoff’s position to Clark is again helpful: as noted in §5.2, Clark endorses neural constructivism but in a restricted sense. The added perspective afforded by Kirchhoff’s emphasis on diachronicity is that Clark severs temporally extended processes of neural transformation into discrete causal points, as if the points were only coincidentally related to each other. It is this spurious division that enables Clark to treat culture as if it were only relevant as an indirect, distal causal entity. Kirchhoff’s larger aim in the 2012 article is to argue for third wave EM.

Kirchhoff believes that first and even second wave EM fail to account for the enculturation of cognitive subjects. The issue, as Kirchhoff presents it, hinges on his contrast between the “Fixed Properties View” and the “Dynamic Properties View” (288). Within the context of EM, Kirchhoff identifies two manifestations of the fixed properties view. The first corresponds to what I have called the core view: a theory allows that individuals’ cognitive architectures are extended and also that their cognitive abilities are thereby enhanced but denies that an individuals’ representational capacities are transformed during ontogeny (293). The second type of fixed property view concerns “cognitive assembly,” in which the individual is held as the primary assembler or driver of cognitive processes. Kirchhoff adopts this notion of “cognitive assembly” from Clark’s aforementioned deflection of Hutchins (2008): that is, cognitive assembly occurs in the here-and-now and is primarily driven by the brain.

Kirchhoff offers an extended critique of the complementarity principle but tellingly focuses on Clark’s use of the term (Kirchhoff 2012, 289). As noted above, Clark unduly narrows the complementarity principle, interpreting it strictly in terms of resource maximization. Sutton’s conception, by contrast, emphatically embraces the role of socialization in shaping a cognitive subject’s goals and most basic biological mechanisms. Kirchhoff leans upon the second version of the fixed properties view—in which the individual is the primary driver of cognition—in order to lump together Sutton and Clark’s respective conceptions of complementarity. When Kirchhoff does directly address Sutton, he interprets Sutton’s complementarity principle as if it were reducible to Merlin Donald’s (1991) distinction between engrams and exograms; Kirchhoff writes, “building on Donald’s theory of engrams ... and exograms ... as a signature mark of complementarity, Sutton points out that exograms are often discrete, modality and task independent, and much more durable than engrams” (292). Yet Sutton explicitly rejects treating

Donald's distinction as denoting a fixed barrier:

It's important to resist a tendency within second-wave EM to treat the inside and the outside as distinct realms with fixed properties. ... External representational systems need not be permanent, of unlimited capacity, translatable across media, or endlessly reformattable as Donald's typology suggests. (Sutton 2010, 206)

Rather than adopting Donald's engram-exogram distinction, Sutton employs it as a starting point, using the memory arts example to illustrate how an engram can possess characteristics associated with exograms.

In fairness to Kirchhoff's critique, there are problematic aspects of Sutton's account of complementarity. In §5.6, I address how Sutton (2014) wrongly leaves open the possibility of primitive representational content. Additionally, in personal correspondence with Kirchhoff, Sutton appears to endorse an overly synchronic view of constitution (Kirchhoff 2013, 22). But even with regard to the latter issue, it is clear that Sutton's account should not be equated with a strictly synchronic view of constitution. Evelyn Tribble and John Sutton (2012, 589) write, "the analysis of the mind must be fundamentally historical in character, because changing cultural artefacts, norms, and institutions are not external supplements or cues to cognition, but partly constitute it." Tribble and Sutton jointly insist that viewing the epistemic agent as a discrete time slice is a meaningless abstraction. At the very least, this recognition of diachronicity undermines equating Sutton's conception of complementarity to the second sense of the "Fixed Properties View": Clark's notion of cognitive assembly, which Kirchhoff uses to explicate the second kind of fixed properties view, hinges upon a clean division between "here-and-now" cognitive activity and historical causes. Kirchhoff's sweeping critique of complementarity is indicative of underlying issues with his third-wave EM program.

In the following two sections, I examine two frameworks that Kirchhoff links his account to. When attached to Hutto's Radical Enactivism and Menary's Cognitive Integrationism,

Kirchhoff's emphasis on diachronicity becomes a fully-fledged model of the cognitive subject. The connections offer two valuable insights: first, Kirchhoff's (2012) adoption of the third-wave EM is a misnomer and, second, epistemic agency should not be reduced to a strictly retrospective dimension. Regarding the first, unlike Cash (2013), Kirchhoff never draws upon functionalism as an explanatory resource. Kirchhoff thus either uses the "third wave EM" label in a loose rhetorical sense or wrongly presumes that the functionalism of EM is limited to Clark's preferred individualistic version. With that said, there are important similarities between Kirchhoff's diachronic subject and pure functionalism. Indeed, Kirchhoff's (2012) endorsement of a decentralized agent could be read as an argument for an unconstrained functionalism against Clark's individualistic rendering of EM. This similarity will be clarified in connection with the second insight—namely, the strictly retrospective account of agency entailed by Kirchhoff's diachronic model. While Kirchhoff arguably overcomes the core view, he fails to overcome the manipulationist presupposition. Accordingly, as I detail in the next two sections, the resulting account of the cognitive subject is similarly flawed. The first-personal agent is missing, crowded out by pre-determined ends and objective communal norms.

5.4 Hutto's Radical Enactivism: Eliminating Basic Mental Content But at What Cost?

Kirchhoff positions his diachronic cognitive subject within Daniel Hutto and Erik Myin's Radical Enactive Cognition hypothesis (REC) and Richard Menary's Cognitive Integrationism (CI). In the next two sections, I unpack both frameworks, which have largely converged into a single formidable interpretation of enactivism despite differing priorities and terminology. I begin with Hutto and Myin's (2013) REC since it addresses a lower level of analysis, offering an account of embodied directedness that does not presuppose mental representations. My explication of REC consists of two parts: first, identify Hutto and Myin's polemical target and,

second, present their account of basic cognition. At the end of the section, I raise a preliminary issue with REC but postpone my full critique of its manipulationism until after having presented Menary's CI.

Even within the enactivist branch of 4E (i.e., embodied, embedded, extended, and enactive cognitive theories), there has been a proliferation of differing interpretations. While in Ch. 3, I noted two main strands of enactivism—the Varela-inspired phenomenological strand and Noë's sensorimotor contingency strand—Hutto and Myin (2013) critique both. The Varela strand, they argue, falls prey to what they consider cognitivist terminology, such as “sense-making” (35). Similarly, although Hutto and Myin consider the sensorimotor contingency theory less problematic, they still take issue with Noë's use of such terms as “practical” and “mediating” knowledge in the context of basic perception (26). The underlying issue for both strands, according to Hutto and Myin, concerns the thesis that Cognition necessarily Involves Content (CIC); they claim that 4E theories in general have not been able to fully reject the thesis.

In an attempt to shift the rhetorical landscape within the philosophy of mind, Hutto and Myin (2013, Ch. 4) present “The Hard Problem of Content” as the field's most pressing issue. In this context, the Cognition necessarily Involves Content (CIC) label serves as a tool for targeting more subtle forms of representationalism than traditional cognitivist models. In opposition to this framing of the philosophy of mind, Wheeler (2014) argues that the nature of mental representation is a separate issue from the extent of cognition, noting that the content-vehicle distinction enables different pairings of internalist and externalist positions regarding mental representation. To this, Hutto, Kirchhoff, and Myin (2014) argue that REC undermines the content-vehicle distinction, invoking REC's account of basic cognition as an alternative. This exchange highlights the general argumentative pattern of REC: its positive account of basic

cognition follows directly from critiques of the thesis that Cognition necessarily Involves Cognition. I turn now to the account of basic cognition that drives REC, aiming to show how it obscures the form of enactivism that I espoused in Ch. 3.

One of the most extensive explications of the basic cognition entailed by REC comes when Hutto (2006) addresses a group of critics, offering a number of clarifications and amplifications often missing from his more polemically-oriented work. Hutto, for instance, articulates his indebtedness to Millikan's (1984, 1993, 1995) account of representations. Specifically, Hutto (2006b, 83) draws upon Millikan's notion of tracking sameness through differences. On this view, tracking occurs through placing a "Strawsonian dot" for each distinct substance to be identified, with each sensory modality identifying the dot in its own characteristic way (84). Drawing upon Dretske (1981), Hutto understands Millikan's notion of tracking as an organism being "informationally sensitive to natural signs" (84). Hutto (2006a, 140) uses the connection to Dretske to drain Millikan's dots of content: he writes, "the only legitimate notion that can be derived from 'information theory' is that of covariance relations—not referential or truth relations." Hutto thus treats the dots as purely indexical signs rather than possessing Millikan's (1995) notion of primitive "pushmi-pullyu" representational content. Whereas content implies its own conditions of satisfaction, Hutto's indexicals do not. The difference corresponds in turn to Hutto's (2006d) distinction between propositional attitudes and intentional attitudes, in which the latter are "propositional attitudes *extensionally* rather than *intensionally* construed" (189; original emphasis). Hutto believes that conditions of satisfaction—intensionality—require much more than indexicals; they require, on his view, languages that can focus a creature's attention on propositions. Although he does not have an

extensive account of language's origin, Hutto sketches what he believes are precursor social mechanisms.

Hutto's sketch of social cognition is founded on the broader concept of Action Coordination Routines. Action Coordination Routines apply to all types of worldly engagements, denoting instances in which routines develop between organisms' informational sensitivity and certain indexical signs (Hutto 2006c, 161). Hutto thinks there are an enormous variety of such routines, including simple and complex behavioral patterns, but all share the basic format of being "pre-scripted" (161). Given their automated nature, the coordination routines operate like functional modules, as discussed by Fodor and Carruthers, except without informational content. Turning to the issue of social cognition, Hutto argues there are coordination routines specialized for "basic social responding" (165). That is to say, agents become informationally sensitive to other people's expressions, with common expressions prompting the creation of corresponding automated responses. While far from foolproof, Hutto argues that these socially tuned responses foster "embodied expectations" that serve as "reliable enough guides to 'other minds,' at least in historically normal conditions" (166). Hence according to Hutto, there's nothing qualitatively different—at their most basic level—between the coordination routines related to social interaction and those related to other types of worldly engagements. The special character of social interaction does not appear until the emergence of symbol systems, such as human languages.

Although REC does not have a detailed story regarding language's phylogenetic origin, Hutto and Myin (2013) offer an ontogenetic account of general symbol use. Their account piggybacks Sutton's second wave complementarity interpretation of EM, with one major revision. While adopting the idea that external symbols transform and augment cognition, they

make the additional claim that any decoupled symbol use is derivative in “both a logical and developmental sense” due the fact that “coupled activities are the ultimate basis of the decoupled ones” (153). What Hutto and Myin effectively endorse is a reverse parity principle, wherein outside processes become internalized. The reverse parity model explains the rather striking characterization of agency near the beginning of the book: “mentality-constituting interactions are grounded in, shaped by, and explained by nothing more, or other, than the history of an organism’s previous interactions” (8). Hutto and Myin thus argue that all mental aspects of epistemic agency should be explicated in strictly retrospective terms, as if mentality were just the sedimentation of an agent’s previous interactions. The retrospective orientation holds, of course, below the level of mentality since the mechanisms that enable the acquisition of symbol systems are understood to be, following the lead of Millikan, the product of natural selection. Having thus explicated REC with an emphasis on its account of basic cognition, I turn now to some preliminary issues, beginning with a concern that Hutto himself alludes to.

Identifying the proper functions of basic cognitive mechanisms is only possible in principle. Following Millikan, Hutto (2006a) grounds the function of such mechanisms in natural selection, but notes the inherent difficulty of actually identifying a function:

In principle, to decide this [a cognitive mechanism’s proper function] we would have to determine whether or not the current conditions matched those which actually advantaged the organism’s ancestors when the response systems were first selected. My bet is that in most cases there would be no such match. (146)

Hutto sees himself as only raising an epistemic issue for third-personal observers, but it highlights an inherent flaw with strictly retrospective accounts of epistemic agency. Granted, the mismatch between present and past circumstances does not by itself impeach Hutto’s retrospective account, as evidence for mismatch cases can be found in the existence of dysfunctional mechanisms. While Millikan rightly considers this explanation of dysfunction a

significant advance over most cognitive theories, it still marginalizes the epistemic significance of error by treating it primarily as a structural mismatch between evolved history and present circumstances. The variety of enactivism that I advanced in Ch. 3, by contrast, stresses the productivity of error, error being integral to learning.

The productivity of error is tied to the prospective orientation of Bickhard's interactional model of representation (§3.2). Hutto fails to recognize the model's insight, as evidenced by his discussion of covariance relations. Despite rightly limiting information theory's importance to covariance relations, Hutto wrongly infers that representational content must be absent from an organism's basic cognitive capacities. What Hutto fails to recognize is that an organism acts as its own point of reference: the covariance relation corresponds to Bickhard's notion of *differential* contact. The only information that perceptual differentiators contain about the environment is that it caused a given final state within an organism to be reached, with the internal state thereby *implicitly* defining the organism's environment (§3.2). Using one's internal states as a reference point would be problematic if the primary purpose of representations were to capture a mind-independent aspect of the world. But representations' primary purpose is instead to refine an organism's interactive anticipations, which is an inherently salient task. As discussed in Ch. 3, accounting for perceptual salience is the chief merit of Bickhard's interactivism, but I will postpone returning to this issue until next section, since the issue is exemplified by Menary's appeal to communal norms.

In closing my discussion of REC, it is worth questioning Hutto's "radical" enactivist label. REC leaves a great deal of the traditional cognitivist picture intact, simply anchoring traditional symbolic capacities in primitive contentless cognitive mechanisms and extended coupling activities. Bickhard's interactivist model, by contrast, avoids positing a major leap from

primitive to symbolic capacities, with the latter on Bickhard's view simply consisting of an organism's abstraction from its own anticipations. Despite being a new level of abstraction, learning still occurs through the same basic process of differential contact and implicit definition. With regard to language, for instance, Bickhard and Hutto thus subscribe to fundamentally different pictures. For Bickhard, human language is notable for its high level of abstraction, though it still does not refer to a mind-independent aspect of reality. Hutto, by contrast, subscribes to the classic picture of symbolic languages as possessing referential and truth relations, thereby ultimately committing REC—albeit at a secondary level—to a much more substantive notion of content than the variety of enactivism that I am defending.

5.5 Enculturation via Cognitive Practices: Menary's Erasure of Epistemic Agency

Using Menary's (2007, 2009, 2012) Cognitive Integrationism (CI) as a focal point, I argue in the present section that a strictly retrospective account of the cognitive subject annihilates any meaningful notion of first-personal agency. The critique applies to REC as well, especially given the two frameworks' affinities. Despite some notable differences, REC ultimately dovetails with CI. Their biggest difference concerns their respective starting points: while REC is polemically-oriented and geared towards low-level cognitive mechanisms, Menary builds CI on the concept of cognitive practices.

The primary purpose of the cognitive practice concept is to differentiate between different types of extension. More specifically, Menary wants to differentiate between Enculturated Cognition and mere Artifact Extension: in the former an epistemic agent participates in a cognitive practice as opposed to merely “outsourcing” or “offloading” an epistemic task on another person or thing (Menary 2012, 148). Enculturated cognition implies that the agent exhibits a certain cognitive character, whereas artifact extension does not. For an

activity to be counted as part of an agent's cognitive character requires that it is "part of the core set of processing routines that directly lead to completing a task" (151). The activity has to be *integrated* into the agent's core set of cognitive processes. According to Menary, the core set of cognitive processes are those specifically responsible for "reliable" belief formation (152). Menary's manipulation thesis more precisely clarifies the distinction between enculturated cognition and artifact extension.

The manipulation thesis links the concepts of cognitive practice and character to communal norms. Menary (2010a, 563) writes, "we create, maintain and manipulate cognitive niches. ... Cognitive manipulations are bodily engagements with the niche that are regulated by norms." The defining feature of cognitive practices—enculturation—is thus learning how to skillfully *manipulate* artifacts in accord with norms, as opposed to identifying practices with the artifacts themselves. Menary (2012, 161) uses the example of an airport security officer using a baggage scanner: "The ability to manipulate the tool is part of my cognitive character; it is an ability that I have to maintain by diligence. The processes inside the tool are not part of my cognitive character ... they are not abilities of mine." While a baggage scanner produces an image irrespective of its user's abilities, security personnel need to be trained in order to read the images properly and must maintain the skill through practice. Hence, on this view, the acquired skill is part of the security officer's cognitive character whereas the CT scanner is not. The manipulation thesis highlights the integral role of norms within CI: norms are precisely what differentiate cognitive practices from inert physical artifacts. Furthermore, the communal nature of norms highlights an important affinity between CI and REC.

As with REC, the CI framework models cognitive development in terms of outside processes becoming internalized. Using one of his favorite examples of a cognitive practice,

Menary (2012, 152) describes learning a written language as follows: “the inside is transformed to be more like the outside (as a kind of reverse parity principle).” As in Hutto’s picture, decoupled “internal” skills—in this case, skills related to writing—derive from previous coupled interactions with external artifacts. In relation to a cognitive practice, an individual’s epistemic skills are thus meant to be tied to communal norms. To bolster the link between internal skills and communal norms, Menary and Kirchhoff (2013) draw upon Hubert Dreyfus’s model of expertise.

Menary and Kirchhoff (2013) argue that communal norms are most apparent for novices but their visibility recedes with performative fluency. The norms shift from being explicit and context-free to being entirely implicit and embodied. Even though embodied expertise cannot be articulated in terms of explicit rules, Dreyfus believes it is a form of “practical objective knowledge” (§2.2). Among the issues I raised in Ch. 2 against Dreyfus’s account, most troublesome is how “objective” embodied knowledge dramatically narrows the epistemic significance of social interaction: other people are at best an opportunity for implicitly copying a piece of objective knowledge. While the extent to which CI is committed to Dreyfus’s flawed model of coaching is debatable, his notion of objective embodied knowledge is a useful point of entry for assessing the unwelcome ramifications that the cognitive practice concept has on first-personal agency.

Similar to Dreyfus’s notion of objective practical knowledge, Menary understands cognitive practices in terms of stable patterns. The stability of cognitive practices explains how communal norms, whether they are explicit rules or implicit embodied knowledge, are impressed upon individuals. This portrayal of cognitive practices rests in part on Roepstorff et al.’s (2010) work in cognitive anthropology, as cited in Menary and Kirchhoff (2013). Roepstorff et al.

(2010) use the concept of patterned practices to supplant the more homogeneous notion of culture common within neuroscientific studies:

Everyday life is continuously ordered into more and less stable patterns that are specific to particular types of situations. ... A patterned practice approach assumes that regular, patterned activities shape the human mind and body through embodiment and internalization. Vice versa, enacting practices shape and re-shape norms, processes, institutions, and forms of sociality. (1051-52)

Roepstorff et al. thus show that characterizing practices as stable patterns does not entail that practices are static inert entities. Even though, on this view, an activity that falls outside of a given pattern is an aberration with regard to the corresponding practice, the authors note that an agent can reshape the practice. Such a reshaping, however, presupposes that the practice first got “under the skin” of the agent (1052). Despite thus pointing to a significant role for individuals, picturing practices as stable patterns impressed onto individuals ultimately undermines first-personal agency, as evidenced by how this picture manifests itself in particular examples of cognitive practices.

Three examples of cognitive practices from Menary and Kirchhoff’s writings are useful illustrations of the approach’s ramifications for first-personal agency. First is Menary’s (2012) previously discussed example of written language, whose allusion to a reverse parity principle links CI to REC. Reverse parity corresponds to the notion of practices getting “under the skin” in Roepstorff et al. (2010), an image that Menary and Kirchhoff (2013) reiterate in examining mathematical reasoning. In the case of mathematical reasoning, the second example, Menary and Kirchhoff substantiate the notion of reverse parity by appealing to Dehaene’s (2005) neuronal recycling hypothesis, wherein cultural inventions “invade” evolutionarily older brain circuits and

change their functions.¹¹⁹ The rhetoric of “invasion” is apropos—indeed, symptomatic—of the agent’s stature vis-à-vis cognitive practices.

As a third and final example, Kirchhoff (2012) considers culturally attuned phoneme perception, a case also discussed by Roepstorff et al. (2010). While the norms of a written language can be explicated in terms of rules, phoneme perception is important since it serves as evidence for the more contentious notion of objective implicit embodied norms. In a study performed by Näätänen et al. (1997, 434), speakers of Estonian and Finnish languages were found to have significantly higher sensitivity to non-prototypical vowel sounds when the sound was paired with a vowel from their native language. As interpreted by Kirchhoff (2012, 298), “patterns of practices at the socio-cultural level reshape the patterning of cortical connectivity and activity.” The point of using the term “practice” is to suggest that language possesses a stability—in this example, a specific pattern of phonemes—that transcends its particular manifestations. The process’s stability corresponds to its causal efficacy: it is the practice “at the socio-cultural level” that changes an individual’s neural structure, a process that only incidentally—on this view—occurs through exchanges with particular language speakers. At this point, given that enculturated imprinting encompasses both explicit rules and embodied skills, it is worth asking whether anything is left of the first-personal agent.

It is helpful to consider the relationship between agents and cognitive practices in terms of temporal directionality. As noted in the previous section, Hutto and Myin (2013) present a strictly retrospective account of mental thought. By claiming that all mental thought is not only “grounded in” but also “explained by” an organism’s past interactions (8), Hutto and Myin

¹¹⁹ Dehaene’s neuronal recycling hypothesis is an alternative to Quartz and Sejnowski’s (1997) neural constructivism. In contrast to the latter’s plasticity, Dehaene believes that cultural inventions are constrained by having to repurpose fairly rigid neuronal structures. Yet Menary and Kirchhoff are not particularly concerned with the mechanisms of internalization so long as the broader picture of the outside becoming internalized is retained, whether via structural or functional transformations.

marginalize if not totally eliminate the spontaneity of first-personal agency. This issue is not obviated by the fact that the passage addresses mental thought as opposed to basic cognition: spontaneity is not going to be found in basic cognition's "pre-scripted" Action Coordination Routines either, since the evolutionarily and developmentally selected routines are even more impersonal and automated. Given that the retrospective construal of the agent follows from the reverse parity principle, Hutto and Myin's characterization applies to Menary's CI as well. For both frameworks, the epistemic agent is sculpted by its history of communal norm-guided interactions. The variety of enactivism that I advanced in Ch. 3, by contrast, treats agency in prospective terms. The perceptual salience issue that it addressed manifests itself in the present context in terms of an individuation problem.

On a retrospective account, it is unclear how to individuate cognitive practices and epistemic agents. Seen from another angle, it is unclear how to individuate the salient aspects of an agent's interactive history with regard to a particular action. To address the problem by referring to statistical regularities only highlights the complete absence of first-personal agency, as the practice's patterns are doing all of the explanatory work. As an alternative to retrospective accounts, I turn now to my "team helper" example from Ch. 4. In the example, I understand social identities in terms of Granic's notion of behavioral attractors (§3.3, §4.4). As a behavioral attractor, a social identity is not a causal force underpinning or driving an agent's actions but rather corresponds to what an agent recognizes as interactive opportunities.

My specific illustration of a social identity concerned social comparison pressures within epistemic communities. The social comparison pressure involved instances in which an agent remained reticent when a team member voices a point at variance with her own findings (§4.3). Learning of and subsequently being able to recognize the behavioral pattern can empower the

agent, enabling her to frame her actions as being a “team helper” searching for a “hidden profile” as opposed to a “trouble maker” attacking a team member. In this case, the significance of the behavioral pattern consists precisely in overcoming or mitigating it. The retrospective account has its own way of explaining such cases, but, as argued in the next section, it relies on a manipulationist presupposition that narrows the epistemic import of social interaction. For the present moment, I only want to stress that it is the prospective orientation of the first-personal agent that determines what particular aspects of her interactive history appear as relevant. While an agent’s interactive history constrains its range of interactive opportunities, it is the agent’s prospective directedness that determines in a given situation which interactions actually appear as opportunities. The inchoateness of an agent’s interactive history thus appears as a virtue—instead of a problematic indeterminacy—since it enables the agent to exploit its history in more novel ways.

In the present section, I have raised a general concern with retrospective accounts of agency. Even though Menary and Hutto rightly emphasize the social dimensions of cognitive development, their retrospective accounts reify the patterns of cognitive practices and norms into causally efficacious entities. As a result, the cognitive subject is annihilated, as similarly witnessed with regard to functionalism. Furthermore, like functionalism, Menary and Hutto fall prey to a manipulationist presupposition, which is needed in order to glue together what appear to otherwise be historically contingent processes. As an alternative, in the chapter’s closing section, I present a complementarity-based interpretation of enactivism.

5.6 The Encultured Mind: An Enactive Account of Complementarity

Sutton’s complementarity principle holds an argumentative advantage over Hutto and Menary’s retrospective accounts. Complementarity is not saddled with the strong reverse parity claim and

thus has a lower explanatory burden. Yet complementarity does not claim that the historical dimension is irrelevant—far from it. The difference, as I detail in the present section, hinges upon moving beyond manipulationism. After formulating Hutto and Menary’s manipulationist response to the team helper example from last section, I present an enactivist interpretation of the complementarity principle, drawing upon Gallagher and Miyahara’s (2012) enactive account of intentionality and Di Paolo’s (2009, 2010) more general account of agency. In this context, complementarity is understood in terms of reflective equilibrium. To begin, I want to consider from a broader perspective Hutto and Menary’s models of the cognitive subject in order to highlight its manipulationism.

Hutto and Menary’s account of the cognitive subject consists of two general layers: basic cognitive mechanisms and public representational systems. Exemplified by Hutto’s notion of Action Coordination Routines, the former fulfill pre-determined ends as ultimately set by natural selection. Public representational systems provide novel means for accomplishing those basic ends. In terms of the team helper example, the more basic layer can be seen as a leverage point for counteracting social comparison pressures. The agent desires accomplishing a certain task, the desire is undergirded by more basic ones, the most basic ones ultimately bottom out in matters of survival (e.g., earning a paycheck to feed oneself). Hence, upon learning of the social comparison pattern, the agent is driven by one of the more basic ends to counteract the pattern rather than being enchained by it. The notion of the basic layer consisting of “pre-scripted” ends is one of the clearest forms of the manipulationist presupposition in cognitive science and social epistemology. Similar to Clark’s neural constructivism (§5.2), Menary’s discussion of cognitive transformation is limited to finding novel ways to accomplish pre-established ends. But, as I first

argued in Ch. 3, the creation and transformation of epistemic aims is one of the most vital dimensions of social interaction.

To expand on the idea of generating epistemic ends through social interaction—first discussed in relation to participatory sense-making (§3.3)—I turn now to Gallagher and Miyahara’s (2012) enactive account of intentionality. To illustrate the difference between the enactive and cognitivist perspectives, they invite the reader to imagine the following scenario: you are driving a car along a road when you notice a person on the edge of the street looking back and forth; you slow down or ready yourself to do so. In considering the scenario, Gallagher and Miyahara argue that rather than forming a belief that the person might cross the road, which subsequently led you to brake, the placing of your foot on the break pedal accounts for all of your intentionality towards the pedestrian. Now if there were another person in the car who asked why you had slowed down, then—in responding that you think the person might cross—you would have formed a representation of the pedestrian. That is to say, reflecting on one’s actions, as prompted on some occasions by another individual’s query, is not simply the articulation of latent intentions. While the point is most pertinent for challenging cognitivist theories, it also cuts against strictly retrospective accounts of agency.

The epistemic agent is not strictly a product of its past interactions. In the driving scenario, the social interaction generates a new intention in part because of the *inchoateness* of what Gallagher and Miyahara call “operative intentionality”: in the first scenario, the driver only possesses an embodied directedness towards the pedestrian, while in the second—by simply being prompted to reflect on the braking action—there is also a representation of the pedestrian. Likewise in the team helper case: the behavioral pattern corresponding to social comparison pressures is pliable, which explains why simply reflecting upon and recognizing the pattern can

have a meaningful impact. The inchoateness of embodied directedness is a corollary to the indeterminacy of an agent's interactive history. In this way, Gallagher and Miyahara's account of enactive intentionality offers another perspective on why the indeterminacy of one's interactive history is an epistemic virtue: it is an inchoate resource to be exploited rather than a mass of patterns that have been impressed onto the agent.

Di Paolo (2009, 2010) offers a more general perspective on the complementary relationship between the agent and its interactive history. Di Paolo's goal in the two essays is to clarify why cognition is not reducible to autopoietic conservation. In Di Paolo (2009), this involves conceptualizing how the sense-making activities of an organism can radically transform its metabolic substrate. Part of this conceptual groundwork consists in using the notion of operational closure as an alternative way of determining whether something is a constituent part of a living system: a process counts as part of a living system based on whether it is conditioned by another process of the system, rather than whether it is on the inside or outside of a physical boundary (Di Paolo 2009, 16). Conceived in this way, it is possible to recognize cases in which an animal's behavior fundamentally reshapes the network of processes that constitute the living system and thereby becomes itself a constituent process.

Di Paolo (2010) offers the water boatman insect (*corixidae*) as a concrete example of how agency can transform an organism's metabolic substrate. By trapping air bubbles on the hairs of its abdomen, the water boatman is able to live most of its life underwater, where it feeds mostly on algae (63). The insect's preference for eating underwater plants leads it to live in conditions more precarious than necessary for its survival and reproduction, a development that does not cohere with a strictly autopoietic viewpoint. Agency opens up previously non-viable domains and with it greater precariousness (66). By illustrating how agency intervenes in its

metabolic substrate, Di Paolo explains how cognition is intimately related with though not strictly an extension of autopoietic self-maintenance. Agency is not simply a means of accomplishing pre-established ends but rather is capable—at the most fundamental level of its own being—of creating new ends. Di Paolo (2010, 66) describes the reversal as a shift from a “life-support of agency” (i.e., the organism acts to satisfy its autopoietic ends) to an “agency of life-support” (i.e., the actions of an organism transforms its metabolic ends). Hence, while Di Paolo stresses the importance of an organism’s interactive history, the agent is irreducible to it.

Having used two enactivist accounts to highlight the pliability of an epistemic agent’s embodied directedness and its interactive history, I will now link them to the notion of complementarity. The complementarity principle helps to clarify the ontological implications of my general interpretation of enactivism—that is, my synthesis of Di Paolo, Gallagher, and Bickhard’s respective works in Chs. 3 – 5. My understanding of complementarity cuts slightly against Sutton (2014), since he leaves open the possibility of primitive representational content. From this angle, Sutton treats the relationship of complementarity as holding between inner representations and extended processes of engaging the world. On my usage, complementarity refers to the relationship of differential contact between agent and environment. There is no internalization of outer content, whether at a basic cognitive or symbolic level. Founded upon differential contact with one’s environment, representational content is instead endogenously produced by an agent, the content indexed to its own internal states.

My reading of complementarity finds common cause with and deepens the significance of the most noteworthy innovation in twentieth century ethical and political philosophy: the method of reflective equilibrium. For Rawls (1999) and Goodman (1983), reflective equilibrium is a deliberative method crucial to negotiating conflicting demands, whether concerning

induction or the political interests of disparate groups of people. Goodman's (1983) original aim was to shift away from the old problem of attempting to justify the act of induction to the new problem of better understanding how acts of induction are actually legitimated. The shift is based on how Goodman believes we already assess inductive inferences: we balance our endorsement of general inferential rules with our assent to specific inductions (121). A rule may be revised, for instance, if it is found to block a prediction we want to preserve upon reflection and vice versa.

The enactivist version of the complementarity principle goes further by suggesting why the old problem of induction is intractable. If mental content is intrinsically indexed to the internal states of individual epistemic agents—as opposed to referring to objective mind-independent features of the world—then there is a basic reason why our epistemic methods lack an objective foundation. For this reason, complementarity suggests that our practices of justification, as described by Goodman (1983), is not a contingent or temporary limitation of our cognitive abilities.

This rendering of complementarity also buttresses reflective equilibrium's dramatic social implications, beginning with Rawls's application of reflective equilibrium to matters of political justice. By the end of his career, it became increasingly apparent to Rawls that a general convergence on a shared equilibrium amongst disparate groups of people was implausible (Daniels 2013). In this context, Rawls's desire to bracket appeals to moral truths is supported by enactivist epistemology rather than requiring a special exception for ethical matters. With regard to Kusch's social epistemology (§1.7, §2.5, §4.4), the complementarity principle challenges his division between natural, artificial, and social kinds. Despite a number of crucial insights,

especially his critique of testimony theories, Kusch's social epistemology projects too large of a role for language. To recap, social kinds are the products of collective stipulation via language.

In Ch. 1, I noted that the distinction between artificial and social kinds does not hold up to scrutiny, since a typewriter would cease to exist just as much as money would if all humans disappeared. In both cases, all that would be left would be token traces: rectangular cubes of metal, circular discs of metal, and so on. Enactivist complementarity takes this critique further, challenging the distinction between natural and social kinds, which Kusch (2002b) adopted from Barnes (1983). Enactivism undermines the underlying picture of words being pinned onto events or material entities. Much of Kusch's advances over other testimony theories stems from emphasizing the performative dimension of language. But the notion of collective stipulation—the centerpiece of Kusch's account—entails an overly stable picture of the relationship between language and world.

There are more social dimensions of knowledge than acts of collective stipulation. As first presented in Ch. 4, I used the notion of a “team helper” as an example of a social identity in order to move away from the overly stable categories implied by Kusch's privileging of language. The performative nature of language is important not because of the distinctive nature of language but rather because language operates on a more fundamental embodied directedness. Enculturation thus begins at a level prior to symbol systems, one of its most basic roles being the generation of epistemic aims. Enculturation's influence on inchoate embodied directedness is obscured if one anchors the social dimension to language, as in Kusch, or to the underlying manipulationist presupposition, as in the case of Hutto, Menary, and functionalist theories of mind.

For many types of epistemic analysis, the manipulationist presupposition is a useful heuristic. Yet for a wide array of theoretical frameworks, as detailed in the present chapter, manipulationism is an *indispensable* presupposition, necessary for gluing together what would otherwise be a coincidence of processes. From this perspective, the epistemic agent turns into a veritable Frankenstein, its only epistemically consequential features being those sculpted by natural selection and its cultural intermediaries. In presenting the encultured mind, as understood in terms of my enactive complementarity principle, I have sought to illustrate how profoundly shaped we are by our historical, social, and developmental circumstances, influences that because they intersect and interact cannot be fully isolated from each other. This profound influence, as I have warned, should not however be equated with agency. The cognitive subject, ever presumptuous, enacts its history in the process of creating a new one.

Conclusion: Epistemic Enquiry, Encultured Enquirers, and their Future

6.1 From the Problem of Justification to Enculturation: How Much Has Changed?

In the preceding chapters, I have traced the movement in social epistemology from the abstract problem of justification to the epistemic dimensions of embodied skills and enculturation. In Ch. 1, I harnessed the testimony literature—especially Goldman’s (1978) epistemics project—to problematize the traditional picture of knowledge, wherein propositional beliefs are the most fundamental epistemic unit. Tellingly, one of the most compelling presentations of the testimony view—Kusch’s (2002a) communitarian epistemology—is centered on language’s performative dimensions rather than its denotative properties. The insights of Kusch’s communitarian epistemology, I argued, are best understood by abandoning the testimony framework in favor of one that grounds cognition in an agent’s embodied interaction with its environment. Viewed from this perspective, the testimony literature is a Trojan horse for justified true belief accounts of knowledge rather than, as Schmitt (1994) suggests, a litmus test for refining them.

Moving to an embodied account of knowledge opens up new avenues of enquiry for social theory. There are, however, a wide range of competing frameworks for conceptualizing the embodied nature of knowledge, as evidenced by range of theories canvassed in Chs. 2 – 5. Synthesizing the works of Di Paolo (2005, 2009, 2010), Bickhard (1996, 2002), Gallagher (2005a, 2013), and Sutton (2010), I have defended an enactivist framework in which the future-oriented epistemic agent is in reflective equilibrium with her environment. For human epistemic

agents, some of the most important environments are the many overlapping social contexts that the agent is thrown into and subsequently participates within: that is, the embodied human epistemic agent is intrinsically encultured.

In this concluding chapter, I offer a more general comparison of my interpretation of enactivism to some of the competing frameworks. In particular, I address theorists who take the naturalization of cognition seriously—especially those that recognize the primitive epistemic significance of embodied skills—but regard enactivism with skepticism under the belief that adopting it would involve radical changes while only promising nebulous advances. Clarifying the methodological ramifications of enactivism for social theory mitigates some of these concerns. In the second half of the chapter, I focus on what I consider to be the most distinctive implications of the enactive framework—namely, the novel perspectives it affords with regard to ethics. I articulate an important ethical implication of my account and argue that it is, in more general terms, a fruitful framework for tackling many pressing ethical dilemmas.

To return to my skeptical interlocutor, two of the most basic objections that can be leveled against my account is that it does not offer any crucial empirical evidence in its favor and, furthermore, that its emphasis on the first-personal and encultured aspects of cognition would hinder a number of contemporary research projects. To expand upon the latter issue, if the emphasis on the first-personal dimension of cognition entails that each agent is unique, then it would seem to cut against some of the most interesting results in neuroscience—such as the findings concerning culturally specific phoneme perception and mathematical reasoning that were discussed in Ch. 5. That is, if each individual is unique, then we should not expect to find any meaningful identity across cognitive subjects. In raising the question how similar or different

cognitive subjects are, the objection alludes to an important aspect of the relationship between enquiry and enquirer.

The results of enquiry cannot be separated from the research interests or goals that motivate its enquirers. This point does not entail a voluntarist constructivism: just because a researcher may desire a particular result does not mean that she will find it. Instead, with regard to the particular issue of comparing cognitive subjects, the point is that whether one finds identity or difference across subjects depends massively on the level of grain the researcher chooses. At one end of the spectrum are social epistemologists, who track population-level patterns of epistemic communities. One of their main tasks is to identify mechanisms that would otherwise be invisible or transparent if considered in terms of particular subjects. But a shared pattern—while significant—does not prove that it is a self-contained, causally efficacious phenomenon. The same lesson holds, at the other end of the spectrum, in appeals to fMRI scans of individual brains: similarity of neural patterns does not establish that what has been identified is the causally efficacious foundation of a particular epistemic skill. It is important to acknowledge that judgments of identity across cognitive agents necessitate setting aside what are considered negligible differences.

There are ineliminable differences amongst agents. These differences might appear to be idle or negligible when viewed at a very high level of generality or, at the other extreme, in terms of single isolated events. Who cares, for instance, if they are slight differences in neural excitation patterns so long as the similarities appear to correspond to a salient epistemic trait. Or, when viewed in terms of a single isolated event, who cares if there are slight differences in how two agents accomplish a certain task, so long as the task is accomplished. To draw upon an example from the end of Ch. 4, who cares if one person has an exact item-by-item printout for

grocery shopping and calls it a “load list,” whereas another person only has a few items scribbled on a piece of scrap paper. In the end, as this dismissal would have it, each comes home with stuff to eat. But such apparently mundane differences are not always as fleeting and trivial as they initially seem. Sometimes they reflect salient differences in the agents’ interactive histories, reflect the differing epistemic ventures that each has participated in and that have transformed the agent in signature ways. Such cases underpin the methodological proposal made in Ch. 4.

As proposed at the end of Ch. 4, one general methodological shift that would benefit cognitive science and social epistemology is to move away from toy examples in favor of focusing on cutting-edge epistemic pursuits. Using toy examples to establish the basic mechanisms underpinning more complex cases presupposes the core view—that is, that there are underlying mechanisms common to both cases. But it is the complex cases—participation in innovative epistemic enterprises—that transform agents, as evidenced by the pervasive differences that appear in how people accomplish mundane epistemic tasks. The methodological shift does not entail that we should stop analyzing the activities of everyday epistemic agents in favor of, to use one of the case studies from Ch. 4, physics research laboratories. The decisive issue, instead, is that epistemic analyses should avoid searching for the lowest common denominator across agents: whether examining the discovery of the top Quark or the compilation of a shopping list, epistemic analyses should be mindful of the differing histories of the agents involved, and how the differences can subtly manifest themselves with respect to a given task. Accordingly, the general methodological shift that I am calling for casts into doubt “one-factor explanations” of social phenomena, as touched on in Ch. 1. The sweeping enormity of one-factor explanations depends heavily on positing the existence of some fundamental cognitive

foundation—whether a neural core or a set of communal norms. The enactivist framework, by contrast, involves multiple levels of analysis.

In addition to the first-personal and population-level perspectives, there is an institutional level—what Shaun Gallagher and Anthony Crisafi (2009) call “mental institutions.” As evidenced by my conception of social identity (Ch. 4), I think that institutions should be understood in terms of opaque coordination mechanisms rather than as a codified set of norms or a collection of physical artifacts. The latter two conceptions reify institutions, turning them into pseudo-agents. The proviso follows in the spirit of Turner’s critique of Searle and normativism (Ch. 2): epistemically productive social interaction does not presuppose a playbook, a codified set of rules, or any other kind of formal organization (institutionalized or otherwise). Formalized institutions might be necessary for certain kinds of tasks—e.g., modern scientific research (Kitcher 1990)—but not for epistemically productive social coordination *per se*. And, following Bickhard’s (2002) levels of knowing model, I understand the formal modes of social collaboration as a more abstract and refined but still opaque mode of coordination.

Understanding the codified rules of social institutions in terms of opaque coordination mechanisms links social institutions to more ephemeral though nonetheless formative epistemic pursuits. From hobbies and other recreational interests pursued through clubs to even more transitory interactions—such as helping a friend or family member with schoolwork or how to use a new phone—there are many gradations in how formalized a social interaction is. Accordingly, for social theory, there are many different possible levels of analysis. The more ephemeral forms of social collaboration are not reflections of formal institutions; on the contrary, it is the institutions that reflect the more ephemeral forms of interaction. Even in the institutional contexts in which there is a codified set of rules that structure interaction, how the rules are

understood and applied is shaped by each individual's unique interactive history, of which institutional training is only a part. It is the particular agents and their particular engagements that make improvised use of the rules and tools of institutions. Hence, even though my framework calls for analyzing a plurality of levels, the first-personal perspective plays a distinctive role.

Having presented enquiry as consisting of multiple levels in order to afford more nuanced comparisons of cognitive subjects, I have come full circle with regard to the objection it was meant to address. To wit, the objection states that stressing the first-personal perspective and uniqueness of particular agents would hinder many contemporary research projects. At this point, it is worth noting that the first-personal perspective and the attendant model of enquiry do not undercut what has been one of the most important perspectives for social epistemology and cognitive science. That is, by denying manipulationism—the view that people's epistemic significance is of *strictly* instrumental value—it might appear as though my enactivist framework undercuts the ability to examine agents' behavior in terms of instrumental benefits, which is an indispensable level of abstraction for cognitive science and social epistemology.

The enactivist framework does not dismiss the importance of instrumental analyses. Instead, it entails that instrumentality should not be treated as the sole fundamental or isolable factor in epistemic analyses. The danger with manipulationism—the dead-end, self-sealing argumentative strategy—is to take cost-benefit analyses and resource maximization as the ultimate aim of all epistemic activity, regardless of evidence to the contrary. Such an argumentative strategy—as considered, for instance, with regard to *ars memoria* case study in Ch. 5—is analogous to Ptolemaic astronomers, adding epicycle upon epicycle in order to save the phenomena. Having thus already presented this critique, I turn my attention to the role that

instrumental analyses still play within my framework. Instead of blocking such analyses, my interpretation of enactivism entails that they should be treated as *heuristic* devices. The difference between manipulationism and heuristic instrumental analyses is that the former squeezes every aberrant activity into one of two results. On the manipulationist view, an activity that does not fit a cost-benefit/resource maximization model is either dismissed as non-epistemic craft (e.g., an irrational agent) or is an indication of an undiscovered instrumental benefit that necessitates tweaking a cost-benefit model.

Treating instrumental analyses as heuristic devices opens up a much greater range of possible outcomes. Take, for instance, Fehr and Gächter's (2002) concept of altruistic punishment. In one of the more novel game theoretic studies, Fehr and Gächter found that participants were willing to pay to punish free riders, despite knowing there would be no future interactions with the person (137).¹²⁰ And even though the groups were shuffled after each punishment, the free riders' contributions increased after having been punished.¹²¹ Hence, the punishment is altruistic since the punisher incurs a cost without deriving a material benefit, though it benefits the participants that subsequently interact with the free rider.

Altruistic punishment is an anomaly with respect to the three most prominent evolutionary explanations of human cooperation. Kin selection does not apply since the group members are unrelated, direct reciprocity does not apply since the group reshuffling prevents a *quid pro quo* incentive, and costly signaling (i.e., reputation building) is impossible given the

¹²⁰ In Fehr and Gächter's (2002) "public goods" experiment, the participants—grouped into fours—had to decide how much money to invest. Investing money came at a personal cost but benefited the other three group members. It was thus in each participant's maximal self-interest to invest nothing, but if no one invested any money, then no one would benefit from the fixed return on investment. The identity of participants were kept anonymous and the groups were shuffled after each investment decision in order to ensure there were no subsidiary benefits for punishing the free rider (137).

¹²¹ The experiment involved punishment and no-punishment conditions. The average contribution was higher in all sessions of the punishment condition. Furthermore, average contributions increased with each subsequent round of the punishment condition (137).

condition of anonymity. Falling outside of these three theories, Fehr and Gächter argue that strong negative emotions were the proximate mechanism that explained people's willingness to punish free riders despite the steep cost (139).¹²²

On a manipulationist view—given its adaptationist underpinnings—it is necessary to present a selection history to explain altruistic punishment. One could argue, for instance, the function of negative emotions in social contexts is to incline agents to punish free riders since it results in greater cooperation, which nets an overall benefit (i.e., a “public good”). There are two basic issues with positing such a selection history that I want to address. First, there is a mismatch between an organism's present environment and that of its ancestors. I raised this issue in Ch. 5 with regard to Hutto's conception of radical enactive cognition. Specifically—having adopted Millikan's account of functions—Hutto (2006a, 146; Ch. 5) concedes that it is practically impossible to identify the proper functions of basic cognitive mechanisms due to the likely mismatch between present and past environments. The mismatch raises the worry—which applies equally to the three aforementioned evolutionary explanations of social cooperation—as to whether the selected mechanism could still play a useful role in contemporary social contexts.

The second issue concerns the high degree of stability implied by a mechanism possessing a selection history. This stability is illustrated by Hutto's (2006c, 161; Ch. 5) concept of Action Coordination Routines and Fodor's (1983a) more well-known concept of functional modules: in both cases, the mechanism is highly automated or, as Hutto puts it, “pre-scripted.” With regard to altruistic punishment, the manipulationist sketch suggests that negative-emotions-in-a-social-context have been selected to serve *the* specific function of inclining an agent to

¹²² Fehr and Gächter's claim that negative emotions were the proximate mechanism rests on two salient punishment patterns: the highest contributors also paid the most to punish others and, second, the greater the deviance from the group norm, the bigger the punishment (139). Surveys of the participants substantiated the patterns: both the punishers reported anger and the free riders reported that they expected to be the object of anger.

punish a free rider, which in turn induces greater communal cooperation. As discussed in Ch. 5 in the context of strictly retrospective accounts of epistemic agency, the danger of positing such highly automated and casually efficacious mechanisms is the risk of erasing any meaningful notion of first-personal agency. It is not the agents doing the work but instead their selection history; as a result, agency is weighed down and non-spontaneous. Given these and other issues, there are incentives for approaching the origins of human cooperation without relying on the principle of natural selection.

My enactive encultured framework enables an alternative approach to altruistic punishment and human cooperation more generally, an approach centered on the concept of participatory sense-making (Ch. 3). Viewing human cooperation in terms of coordination mechanisms shifts the focus away from selection histories and towards the particular agents and dynamics of interaction. This is because coordination mechanisms endogenously emerge when agents interact. On this alternative approach, it is possible to reconstruct the three aforementioned theories of cooperation in terms of coordination mechanisms.

As an example of reconstructing a selectionist account of cooperation, consider the case of selfish reciprocity. Rather than linking selfish reciprocity to natural selection, it is possible to look for context specific features that attract mutual exploitation. In Fehr and Gächter's experiment, for instance, the express goal of the experimental setting—acquire as much money as possible via joint investment—attracts mutual exploitation. It was only another aspect of the experimental setting that blocked the possibility of selfish reciprocity—namely, the shuffling of groups. Shifting the focus to the particular agents and interactive context takes the pressure off of reducing a particular mechanism to a specific function, a suspect methodology exemplified in one-factor explanations.

The above manipulationist sketch of altruistic punishment illustrates well the inherent difficulties of ascribing specific functions to a mechanism. Reconstructing the negative emotions as a coordination mechanism leaves open improvisatory space for the agents: when agents *act* on the negative emotions by punishing the free rider, the negative emotions thereby *serve* the function of inducing punishment in contrast to always possessing the specific function irrespective of the agents' actions. This distinction highlights the study's results concerning the dramatic difference between the punishment and no-punishment conditions: presumably, the negative emotions are present in both, but the increase in cooperation only comes when the individuals can act on the emotions. Alternatively, if the emotions differ in each context—a plausible scenario—then this would strengthen the enactivist point, signaling a qualitative difference attributable to the change in interactive context.

By treating particular agents and interactive contexts each as a primitive datum rather than as manifestations of a more fundamental and discrete mechanism, the concept of coordination mechanisms does not crowd out or weigh down the role of first-personal agents. Furthermore, the concept affords a high level of generality—one can talk about a particular coordination dynamic arising in a wide range of interactive contexts—while avoiding the excessively abstract level of one-factor explanations.

This more context-specific approach to the levels of epistemic analysis dovetails with treating instrumentality as a heuristic level of analysis. While the value of the instrumental level of analysis is evidenced by its role in revealing phenomena such as altruistic punishment, it is not the definitive context-independent standard of epistemic analyses. If a social interaction fails to net a perceptible instrumental benefit—whether at the individual or higher level—it should not be automatically relegated to non-epistemic cruft. The aims of epistemic agents vary across

contexts, as should the ends of epistemic enquiry. While I have thus responded to the second objection regarding the types of analyses my framework permits, the reframing and broadening of perspective noted in relation to altruistic punishment might appear as an idle consideration in light of the first objection—namely, that the enactivist framework does not offer any decisive empirical evidence in its favor but rather only the nebulous promise of broader perspectives.

I am doubtful of there being a single crucial piece of evidence, an empirical result, to decide between my enactive account and the competing theories. It should be noted, however, that this is not uncommon in scientific enquiry. While there will likely be no shortage of neurocognitive reductionist and eliminativist studies, the enactivism that I have advanced promises to incorporate the worthwhile insights of such accounts and present them within a more comprehensive framework. As evidenced by the range of theories that my account has drawn upon, what I have defended is less a single theory of the cognitive subject-environment relation and more an interdisciplinary framework for overcoming manipulationism in its many guises. With that said, the closest thing to a crucial piece of evidence for the enactive view concerns its ethical implications. That is, the broader perspectives of the enactive framework open up new and extremely important dimensions of ethical enquiry.

6.2 The Information Age and Not-Just-Information Problems

By undermining the division between knowledge and its application, the encultured enactive view entails that epistemology is intrinsically linked to ethics. Once the distinction is undermined, manipulationism becomes *ethically* problematic. People are not—even from a strictly epistemic viewpoint—resources to be exploited, yet the belief that they are informs knowledge policy and social organizations. Recall, for instance, Fodor’s no-nonsense take on epistemology: “What

philosophers call ‘linguistic deference’ is actually the use of experts as instruments. ...

Epistemology ... is mostly about money; e.g., about what I’m prepared to pay to insure that one or another of my beliefs is true” (Fodor 1995, 36-37; Ch. 3). Fodor wants to dismiss the significance of epistemology by treating it as merely a practical exigency. While Fodor’s point is largely rhetorical, the economic incentives to promulgate manipulationism are very real: in the information age, there is no need to pay an arborist to find out whether a tree is an elm. Likewise, never mind paying a tutor or seeking help from a friend—there is a more convenient, oftentimes cheaper, and much vaster store of information found in Google searches, digital assistants, and IBM’s Watson. The information age can provide us with a practically infinite number of beliefs, which, for the purposes of argument, I will even assume are largely true. (Or at least not noticeably more untrue than the claims made by an average person on an average day.) The information age would—from the viewpoint of Fodor’s thumbnail of epistemology—appear to be an epistemic utopia. But the enactive encultured account of the cognitive subject suggests otherwise.

Two questions my encultured account is specially suited to address are: first, how do our acquired epistemic skills impact our treatment of and relationship to other people? And, second, how do our beliefs about the nature of knowledge impact our treatment of and relationship to other people? Beginning with the latter question: notice that although it might seem too abstract a question for it to be of major ethical importance, the trust in internet search engines, digital assistants, and the like depends on at least an implicit belief that there is nothing of consequence lost when acquiring information from these sources as compared to another person. But the information age comes with special dangers, such as the ease with which one can become lost in a personal informational bubble. It is all too easy to go to sources that entertain us and confirm

our views, especially when the search engine has a record of all our previous queries. In addition to confirmation bias, our access to the sources is controlled by what promises to be increasingly few corporate gatekeepers, with effects such as “googlearchy”¹²³ often invisible to end users.

To escape our information bubbles, even if only temporarily, we need encounters with *alterity*—with the Other. We need such encounters in order to jar us, to open us up to new perspectives and new meanings, to change the way we live. What I have in mind builds on Iris Marion Young’s (1990) vision of heterogeneous public discourse. The enactive framework links Young’s notion of discourse to less articulate but no less epistemically significant forms of interaction, enabling more heterogeneous forms of communication. It is up to us to, following Young’s vision, to come to know someone else and, in the process, become ourselves a different person.

The enactive encultured framework, to be sure, is not the first theory to draw attention to the epistemic dimensions of ethics. Miranda Fricker’s (2007) concept of epistemic injustice, as touched on in Ch. 1, is in this regard one of the most important works in social epistemology. Fricker offers an additional perspective on racial and sexist prejudices. Her most extensive analysis concerns testimonial injustice, in which people subject to a racial or sexist stereotype suffer from a “credibility deficit” (17). While valuable, Fricker’s conception of testimonial injustice only scratches the surface of what social epistemology and cognitive science have to offer ethics. Furthermore, testimonial injustice fails to address the novel forms of oppression made possible by the coupling of manipulationism and the information age; in fact, the Internet could be presented as something of a remedy for testimonial injustice, since it enables people to

¹²³ The term “googlearchy” was coined by Hindman et al. (2003) in their assessment of how the Internet had impacted access to political information. The authors found that a few successful sites received a disproportionate amount of traffic.

share their testimony behind avatars. Taken to a greater extreme: if knowledge is nothing but information, then why bother with our bodies, or even with other people. Because she endorses a traditional picture of knowledge in which beliefs are the most fundamental unit, Fricker's conception of epistemic injustice fails to recognize the distinctive ethical dangers posed by the information age.

Recognizing the ethical dangers of the information age depends, in part, on considering the first question above—namely, how our acquired epistemic skills can impact how we treat other people. The prodigious and numerous beneficial achievements of the information age have an intoxicating potential, as we seek to apply the same methods to new domains. But many of the biggest problems that we face—including public health and environmental crises—depend on more than the frictionless dissemination of information and cheap technological solutions. These problems are instances, I argue, of not-just-information problems, which I take to be a subset of what Garrett Hardin (1968) famously called “no technical solution” problems. Towards this end, I offer two cases: Atul Gawande's (2013) discussion of public health initiatives in India and Maria Konnikova's (2015) discussion of clean water technologies.

Noting the various successes and failures of public health campaigns against cholera and infant mortality, Gawande (2013) challenges the conventional wisdom of the information age. The key to success in such efforts, he argues, is a “sandals on the ground” approach, sending massive numbers of people out into areas such as Bangladesh and India in order to directly engage people in small groups. One of the leading causes of infant mortality, for instance, is hypothermia; Gawande laments the wish for a technological panacea, such as cheap incubators, noting that “kangaroo care” in which a mother swaddles her infant directly to her body is sufficient for combatting the vast majority of cases. Furthermore, despite previous awareness of

the medical protocols, Gawande details the sustained mentoring efforts that are necessary in order to implement kangaroo care, higher standards of hygiene, and other crucial reforms in rural hospitals.

More than providing information, Gawande likens the mentoring efforts of the public health campaign in India to the change in social identity witnessed within American and European medicine in the nineteenth century. Specifically, Gawande notes doctors' resistance to hand sanitation: Joseph Lister had published his antiseptic method in a prominent medical journal in 1867, but sanitation procedures did not take hold until the turn of the twentieth century. The decisive change was initiated by German doctors, Gawande argues, when they stopped associating medical practice with battlefield surgeons, associating it instead with laboratory scientists. The dissemination of information and technological advances were beneficial in both cases but presupposed social mentoring, mentoring that creates new social identities and the range of behavioral changes implied therein.

The environmental issue of access to potable water is another case in which social mentoring must play an integral role. The technology exists for converting human waste to drinkable water, yet implementing the technology has been met with near universal resistance from people on all continents. This resistance holds despite the fact that people are informed that the water from the "Omniprocessor," as noted by Konnikova (2015), is more pure than unfiltered tap water. As part of an ongoing effort to change people's attitudes, Bill Gates has shot a promotional video in which he drinks water that had five minutes previously been human waste. The case of the Omniprocessor thus serves as an important illustration that it is not just the poorest and least educated peoples for whom pure information is not enough, for whom mentoring is essential for informing the way we live.

Other people are a never-ending and interminable source of surprises, a source of wonderment and frustration, a potential for liberation and oppression. Other people are a source of superstitions and prejudices—including racial, gender, and cultural prejudices; but other people also help us overcome them. The Other can jar us into recognizing an inarticulate bigotry, such as when George Yancy (2008) describes, as a black man, walking across a street to the sound of car doors locking and walking into an elevator as white women clutch their purses more closely. Others, we hope, can jar us into overcoming our distaste for Omniprocessor water, forcing us to consider what sorts of radical lifestyle changes the future demands of us. It is not enough to hear statistics concerning the scarcity of water or even hearing of widespread droughts; it takes watching one of the richest people in the world drink water that has been distilled from shit to appreciate just how serious and immediate the problem is.

It should come as no surprise that the wonderment and frustration caused by others is in no small part due to their indispensable epistemic roles. But such an obvious fact is easily occluded given the history—both recent and distant—of how knowledge and the cognitive subject have been understood and given the impressive—perhaps intoxicating—accomplishments of the information age. Given this background, understanding the integral epistemic roles of other people—recognizing that social interaction is more than a practical exigency for acquiring information or some inarticulable skill—is a hard fought achievement. Even when considered from a strictly epistemic viewpoint, other people are more than just tools or resources to be exploited, manipulated, and managed. It is difficult to pay more than lip service to this fact, but, as suggested by the Gawande and Gates cases, addressing some of our biggest problems, problems that threaten our very future, depend on it.

References

- Adams, Frederick, and Kenneth Aizawa. (2008). *The bounds of cognition*: John Wiley & Sons.
- Adler, Jonathan. (2010). Epistemological problems of testimony. *The Stanford Encyclopedia of Philosophy*. Spring 2013. from <http://plato.stanford.edu/archives/spr2013/entries/testimony-episprob/>
- Ariew, André. (2002). Platonic and Aristotelian Roots of Teleological Arguments. In A. Ariew, R. Cummins & M. Perlman (Eds.), *Functions: New essays in the philosophy of psychology and biology* (pp. 7-32). New York: Oxford University Press.
- Ariew, André, and Perlman, Mark. (2002). Introduction. In A. Ariew, R. Cummins & M. Perlman (Eds.), *Functions: New essays in the philosophy of psychology and biology* (pp. 1-4). New York: Oxford University Press.
- Barnes, Barry. (1983). Social life as bootstrapped induction. *Sociology*, 17(4), 524-545.
- Barnes, Barry, Bloor, David, and Henry, John. (1996). *Scientific knowledge: A sociological analysis*. Chicago: University of Chicago Press.
- Bickhard, Mark H. (2000). Autonomy, function, and representation. *Communication and Cognition-Artificial Intelligence*, 17(3-4), 111-131.
- Bickhard, Mark H. (2002). Critical principles: On the negative side of rationality. *New Ideas in Psychology*, 20(1), 1-34.
- Bickhard, Mark H, and Terveen, Loren. (1996). *Foundational issues in artificial intelligence and cognitive science: Impasse and solution* (Vol. 109). New York: Elsevier.
- Bickle, John. (2013). Multiple Realizability. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. <<http://plato.stanford.edu/archives/spr2013/entries/multiple-realizability/>>.
- Block, Ned. (1990). "Inverted earth." *Philosophical perspectives*:53-79.
- Bloor, David. ([1976] 1991). *Knowledge and social imagery* (2nd ed.) Chicago: University of Chicago Press.

- Brandom, Robert. (1998). *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.
- Bruner, Jerome S, and Postman, Leo. (1949). On the perception of incongruity: A paradigm. *Journal of personality*, 18(2), 206-223.
- Burge, Tyler. (1979). "Individualism and the Mental." *Midwest studies in philosophy* 4 (1):73-121.
- Carruthers, Peter. (2006). *The architecture of the mind*. New York: Oxford University Press.
- Cartwright, Nancy. (1999). *The dappled world: A study of the boundaries of science*. Cambridge, UK: Cambridge University Press.
- Cash, Mason. (2013). "Cognition without borders: "Third Wave" socially distributed cognition and relational autonomy." *Cognitive Systems Research* 25:61-71.
- Churchland, Paul M. ([1984] 2013). *Matter and Consciousness*. 3rd ed. Cambridge, MA: MIT.
- Clark, Andy. (1991). *Microcognition: Philosophy, cognitive science, and parallel distributed processing*. Cambridge, MA: MIT Press.
- Clark, Andy. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT press.
- Clark, Andy. (2001). "Reasons, robots and the extended mind." *Mind & Language* 16 (2):121-145.
- Clark, Andy. (2002). "The roots of 'norm-hungriness'." In *Philosophy of Mental Representation*, edited by Hugh Clapin, 37-43. New York: Oxford University Press.
- Clark, Andy. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. New York: Oxford University Press.
- Clark, Andy. (2007). "Curing cognitive hiccups: A defense of the extended mind." *The Journal of Philosophy* 104 (4):163-192.
- Clark, Andy. (2008a). Pressing the Flesh: A Tension in the Study of the Embodied, Embedded Mind? *Philosophy and Phenomenological Research*, 76(1), 37-59.
- Clark, Andy. (2008b). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.
- Clark, Andy. (2010). "Memento's revenge: The extended mind, extended." *The extended mind*:43-66.
- Clark, Andy. (2011). "Finding the mind." *Philosophical Studies* 152 (3):447-461.

- Clark, Andy, and David Chalmers. (1998). "The extended mind." *Analysis* 58 (1):7-19.
- Coady, C. A. J. (1992). *Testimony: a philosophical study*. New York: Clarendon Press.
- Collins, Barry E, and Guetzkow, Harold Steere. (1964). *A social psychology of group processes for decision-making*. New York: Wiley & Sons.
- Craig, Edward. (1999). *Knowledge and the state of nature: An essay in conceptual synthesis*. New York: Clarendon Press Oxford.
- Cummins, Robert. (2002). Neo-Teleology. In A. Ariew, R. Cummins & M. Perlman (Eds.), *Functions: New essays in the philosophy of psychology and biology* (pp. 157-171). New York: Oxford University Press.
- D'Agostino, Fred. (2010). *Naturalizing epistemology: Thomas Kuhn and the "essential tension"*. New York: Palgrave Macmillan.
- Daniels, Norman. (2013). Reflective Equilibrium. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. <<http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium/>>.
- Dehaene, Stanislas. (2005). "Evolution of human cortical circuits for reading and arithmetic: the neuronal recycling hypothesis." In *From monkey brain to human brain*, edited by Stanislas Dehaene, Jean-René Duhamel, Marc D. Hauser and Giacomo Rizzolatti, 133-57. Cambridge, MA: MIT Press.
- De Jaegher, Hanne, and Di Paolo, Ezequiel. (2007). Participatory sense-making. *Phenomenology and the cognitive sciences*, 6(4), 485-507.
- De Jaegher, Hanne, and Di Paolo, Ezequiel. (2008). Making sense in participation: An enactive approach to social cognition. *Emerging communication*, 10, 33.
- De Jaegher, Hanne, and Froese, Tom. (2009). On the role of social interaction in individual agency. *Adaptive Behavior*, 17(5), 444-460.
- De Jaegher, Hanne, Di Paolo, Ezequiel., and Gallagher, Shaun. (2010). Can social interaction constitute social cognition? *Trends in cognitive sciences*, 14(10), 441-447.
- Dennett, Daniel. (1976). "Conditions of personhood." In *The identities of persons*, edited by Amelie Oksenberg Rorty. Los Angeles: University of California Press.
- Dennett, Daniel. (1981). "Where Am I?" In *Brainstorms: Philosophical Essays on Mind and Psychology*, 310-23. Cambridge, MA: The MIT Press.
- Dennett, Daniel. (1991). *Consciousness Explained*. New York: Back Bay.

- Dennett, Daniel. (2001). "In Darwin's wake, where am I?" Proceedings and Addresses of the American Philosophical Association.
- Dennett, Daniel. (2003). *Freedom evolves*. New York: Viking Books.
- Di Paolo, Ezequiel. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the cognitive sciences*, 4(4), 429-452.
- Di Paolo, Ezequiel. (2009). "Extended life." *Topoi* 28 (1):9-21.
- Di Paolo, Ezequiel. (2010). "Overcoming autopoiesis: An enactive detour on the way from life to society." *Advanced Series in Management* 6:43-68.
- Di Paolo, Ezequiel, Rohde, Marieke, and De Jaegher, Hanne. (2010a). Horizons for the enactive mind: values, social interaction, and play. In J. Stewart, O. Gapenne & E. Di Paolo (Eds.), *Enaction: Towards a new paradigm for cognitive science* (pp. 33-87). Cambridge, MA: MIT Press.
- Donald, Merlin. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.
- Dretske, Fred. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Dreyfus, Hubert L. (1965). *Alchemy and artificial intelligence*. Paper presented at the RAND Corporation Paper Series.
- Dreyfus, Hubert L. (1972). *What Computers Can't Do: A critique of artificial intelligence*. New York: Harper & Row.
- Dreyfus, Hubert L. (2001a). How Far Is Distance Learning From Education? *Bulletin of Science Technology and Society*, 21(3), 165-174.
- Dreyfus, Hubert L. (2001b). Introduction to *Body and World*, by Samuel Todes (pp. xv-xxvii). Cambridge, MA: MIT Press.
- Dreyfus, Hubert L. (2005). *Overcoming the myth of the mental: how philosophers can profit from the phenomenology of everyday expertise*. Paper presented at the Proceedings and addresses of the American Philosophical Association.
- Dreyfus, Hubert L, and Dreyfus, Stuart E. (1986). *Mind over machine*. New York: Free Press.
- Dreyfus, Hubert L, and Dreyfus, Stuart E. (1991). Towards a phenomenology of ethical expertise. *Human Studies*, 14(4), 229-250.
- Dreyfus, Stuart E, and Dreyfus, Hubert L. (1980). A five-stage model of the mental activities involved in directed skill acquisition: DTIC Document.

- Eriksen, Charles W, and Kuethe, James L. (1956). Avoidance conditioning of verbal behavior without awareness: A paradigm of repression. *The Journal of Abnormal and Social Psychology*, 53(2), 203.
- Fehr, Ernst, and Simon Gächter. (2002). "Altruistic punishment in humans." *Nature* 415 (6868):137-140.
- Fitch, W Tecumseh. (2008). Nano-intentionality: a defense of intrinsic intentionality. *Biology & Philosophy*, 23(2), 157-177.
- Fricker, Elizabeth. (1995). Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony. *Mind*, 104(414), 393-411.
- Fricker, Elizabeth. (2006). Second - Hand Knowledge. *Philosophy and Phenomenological Research*, 73(3), 592-618.
- Fricker, Miranda. (1998). *Rational authority and social power: Towards a truly social epistemology*. Paper presented at the Proceedings of the Aristotelian Society.
- Fricker, Miranda. (2007). *Epistemic injustice: Power and the ethics of knowing*: Oxford University Press Oxford.
- Fodor, Jerry A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: The MIT Press.
- Fodor, Jerry A. (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.
- Fodor, Jerry A. (1995). *The elm and the expert: Mentalese and its semantics*. Cambridge, MA: MIT Press.
- Fodor, Jerry A. (2001). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: The MIT press.
- Fodor, Jerry A., and Martin Davies. (1986). "Individualism and supervenience." *Proceedings of the Aristotelian Society, Supplementary Volumes*:235-283.
- Fodor, Jerry A., and Massimo Piattelli-Palmarini. (2011). *What Darwin got wrong*. New York: Farrar, Straus and Giroux.
- Fuchs, Thomas, and De Jaegher, Hanne. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the cognitive sciences*, 8(4), 465-486.
- Fuller, Steve. (2000). *Governance of Science: Ideology and the Future of the Open Society*. London: Open University Press.
- Fuller, Steve. (2002). *Social epistemology* (2nd ed.). Bloomington, IN: Indiana University Press.

- Fuller, Steve. (2011). *Humanity 2.0: what it means to be human past, present and future*. New York: Palgrave Macmillan.
- Fuller, Steve. (2012). Social epistemology: A quarter-century itinerary. *Social Epistemology*, 26(3-4), 267-283.
- Gallagher, Shaun, and Katsunori Miyahara. (2012). "Neo-pragmatism and enactive intentionality." In *Action, Perception and the Brain: Adaptation and cephalic expression*, edited by Jay Schulkin, 117-146. New York: Palgrave-Macmillan.
- Gallagher, Shaun. (1997). Mutual enlightenment: Recent phenomenology in cognitive science. *Journal of Consciousness Studies*, 4(3), 195-214.
- Gallagher, Shaun. (2005a). *How the body shapes the mind*. Oxford ; New York: Clarendon Press.
- Gallagher, Shaun. (2011). The overextended mind. *Versus: Quaderni di studi semiotici*, 113-115.
- Gallagher, Shaun, and Crisafi, Anthony (2009). Mental institutions. *Topoi*, 28(1), 45-51.
- Gallagher, Shaun. (2013). The socially extended mind. *Cognitive Systems Research*, 25, 4-12.
- Gallese, Vittorio, and Goldman, Alvin. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, 2(12), 493-501.
- Gascoigne, Neil, and Thornton, Tim. (2013). *Tacit Knowledge*. Bristol, CT: Acumen.
- Gawande, Atul. (2013). "Slow Ideas." *The New Yorker*.
- Gelfert, Axel. (2006). Kant on testimony. *British Journal for the History of Philosophy*, 14(4), 627-652.
- Gibson, James Jerome. (1966). *The senses considered as perceptual systems* (L. Carmichael Ed.). Boston: Houghton.
- Gibson, James Jerome. (1986). *The Ecological Approach to Visual Perception*. Hillsdale: Erlbaum.
- Gilligan, Carol, and Richards, David AJ. (2009). *The Deepening Darkness: Patriarchy, Resistance, and Democracy's Future*. New York: Cambridge University Press.
- Godfrey-Smith, Peter. (1994). *Spencer and Dewey on life and mind*. Paper presented at the Artificial life IV.
- Goffman, Erving. (1983). The interaction order: American Sociological Association, 1982 presidential address. *American sociological review*, 1-17.

- Goldman, Alvin I. (1978). Epistemics: The regulative theory of cognition. *The Journal of Philosophy*, 75(10), 509-523.
- Goldman, Alvin I. (1979). What is justified belief? *Justification and knowledge* (pp. 1-23): Springer.
- Goldman, Alvin I. (1986). *Epistemology and Cognition*: Harvard University Press.
- Goldman, Alvin I. (1991). Epistemic paternalism: Communication control in law and society. *The Journal of Philosophy*, 88(3), 113-131.
- Goldman, Alvin I. (1993). Epistemic folkways and scientific epistemology. *Philosophical Issues*, 3, 271-285.
- Goldman, Alvin I. (1999). *Knowledge in a social world*. New York: Clarendon Press Oxford.
- Goldman, Alvin I. (2005). Mirror systems, social understanding and social cognition. *InterDisciplines*. <http://www.interdisciplines.org/mirror/papers/3>
- Goldman, Alvin I. (2010). Social Epistemology. *The Stanford Encyclopedia of Philosophy*. Summer 2010. from <http://plato.stanford.edu/archives/sum2010/entries/epistemology-social/>
- Goldman, Alvin I. (2011). Reliabilism. *The Stanford Encyclopedia of Philosophy*. Spring 2011. from <http://plato.stanford.edu/archives/spr2011/entries/reliabilism/>
- Goldman, Alvin I, and Shaked, Moshe. (1991). An economic model of scientific activity and truth acquisition. *Philosophical Studies*, 63(1), 31-55.
- Goodman, Nelson. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.
- Gould, Stephen Jay, and Lewontin, Richard C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 205(1161), 581-598.
- Graham, Peter J. (2000a). Transferring knowledge. *Noûs*, 34(1), 131-152.
- Graham, Peter J. (2000b). Conveying information. *Synthese*, 123(3), 365-392.
- Graham, Peter J. (2000c). The reliability of testimony. *Philosophy and Phenomenological Research*, 61(3), 695-709.

- Granic, Isabela. (2002). The Self-Organization of Parent-Child Relations: Beyond Bidirectional Models. In M. D. Lewis & I. Granic (Eds.), *Emotion, development, and self-organization: Dynamic systems approaches to emotional development* (pp. 267-297). New York: Cambridge University Press.
- Greco, John. (2007). The nature of ability and the purpose of knowledge. *Philosophical Issues*, 17(1), 57-69.
- Greco, John. (2012). Recent Work on Testimonial Knowledge. *American Philosophical Quarterly*, 49(1), 15-28.
- Habermas, Jurgen. (1987). *The theory of communicative action: Lifeworld and system: A critique of functionalist reason* (T. McCarthy, Trans. Vol. 2). Boston: Beacon.
- Habermas, Jürgen. (1979). *Communication and the evolution of society* (T. McCarthy, Trans.). Boston: Beacon.
- Habermas, Jürgen. (1984). *The theory of communicative action: Reason and the rationalization of society* (T. McCarthy, Trans. Vol. 1). Boston: Beacon.
- Hanson, Norwood Russell. (1958). *Patterns of discovery: an inquiry into the conceptual foundations of science*. New York: Cambridge University Press.
- Hanson, Norwood Russell. (1967). An anatomy of discovery. *The Journal of Philosophy*, 64(11), 321-352.
- Hardin, Garrett. (1968). "The tragedy of the commons." *Science* 162 (3859):1243-1248.
- Haugeland, John. (1985). *Artificial Intelligence: the very idea*. Cambridge, MA: MIT Press.
- Haugeland, John. (1998). *Having thought: Essays in the metaphysics of mind*. Cambridge, MA: Harvard University Press.
- Haugeland, John. (2002). "Andy Clark on cognition and representation." In *Philosophy of Mental Representation*, edited by Hugh Clapin, 24-36. New York: Oxford University Press.
- Hindman, Matthew, Kostas Tsioutsoulis, and Judy A Johnson. (2003). "Googlearchy: How a few heavily-linked sites dominate politics on the web." Annual meeting of the Midwest Political Science Association.
- Hutchins, Edwin. (2008). "The role of cultural practices in the emergence of modern human intelligence." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1499):2011-2019.

- Hutto, Daniel D. (2006a). "Against passive intellectualism: reply to Crane." In *Radical Enactivism: Intentionality, Phenomenology, and Narrative. Focus on the philosophy of Daniel D. Hutto*, edited by Richard Menary, 121-50. Amsterdam: John Benjamin Publishing Company.
- Hutto, Daniel D. (2006b). "Both Bradley and biology: reply to Rudd." In *Radical Enactivism: Intentionality, Phenomenology, and Narrative. Focus on the philosophy of Daniel D. Hutto*, edited by Richard Menary, 81-106. Amsterdam: John Benjamin Publishing Company.
- Hutto, Daniel D. (2006c). "Embodied expectations and extended possibilities: reply to Goldie." In *Radical Enactivism: Intentionality, Phenomenology, and Narrative. Focus on the philosophy of Daniel D. Hutto*, edited by Richard Menary, 157-78. Amsterdam: John Benjamin Publishing Company.
- Hutto, Daniel D. (2006d). "Four Herculean labours: reply to Hobson." In *Radical Enactivism: Intentionality, Phenomenology, and Narrative. Focus on the philosophy of Daniel D. Hutto*, edited by Richard Menary, 185-222. Amsterdam: John Benjamin Publishing Company.
- Hutto, Daniel D, Michael D Kirchhoff, and Erik Myin. (2014). "Extensive enactivism: why keep it all in?" *Frontiers in human neuroscience* 8.
- Hutto, Daniel D, and Erik Myin. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, MA: MIT Press.
- Hull, David L. (1986). "On human nature." Proceedings of the biennial meeting of the philosophy of science association, University of Chicago Press.
- Hull, David L. (1990). *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.
- Hurley, Susan L. (1998). Vehicles, contents, conceptual structure, and externalism. *Analysis*, 1-6.
- Jonas, Hans. (1966). *The Phenomenon of Life: Toward a Philosophical Biology*. New York: Harper & Row.
- Kant, Immanuel. ([1790] 1987). *Critique of judgment* (W. S. Pluhar, Trans.). Indianapolis: Hackett Publishing.
- Kelso, J. A. Scott. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: Bradford.
- Kirchhoff, Michael D. (2012). "Extended cognition and fixed properties: steps to a third-wave version of extended cognition." *Phenomenology and the Cognitive Sciences* 11 (2):287-308.

- Kirchhoff, Michael D. (2013). "Extended Cognition & the Causal - Constitutive Fallacy: In Search for a Diachronic and Dynamical Conception of Constitution." *Philosophy and Phenomenological Research*.
- Kitcher, Philip. (1990). The division of cognitive labor. *The Journal of Philosophy*, 87(1), 5-22.
- Kitcher, Philip. (1993). *The advancement of science: Science without legend, objectivity without illusions*. New York: Oxford University Press
- Kitcher, Philip. (2001). *Science, truth, and democracy*. New York: Oxford University Press.
- Kitcher, Philip. (2011). *Science in a Democratic Society*. New York: Prometheus.
- Kiverstein, Julian, and Andy Clark. (2009). "Introduction: Mind embodied, embedded, enacted: One church or many?" *Topoi* 28 (1):1-7.
- Konnikova, Maria. (2015). "Problems Too Disgusting to Solve." *The New Yorker*.
- Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago.
- Kuhn, Thomas S. (1977). "Objectivity, Value Judgment and Theory Choice," in *The essential tension: Selected studies in scientific tradition and change* (pp. 320-339). Chicago: University of Chicago Press.
- Kuhn, Thomas S. (1992). "The Trouble with the Historical Philosophy of Science", Robert and Maurine Rothschild Distinguished Lecture, 19 November 1991, An Occasional Publication of the Department of the History of Science, Cambridge MA: Harvard University Press.
- Kusch, Martin. (1998). *Psychological knowledge: A social history and philosophy*. New York: Routledge.
- Kusch, Martin. (2002a). *Knowledge by agreement: The programme of communitarian epistemology*. New York: Clarendon Press Oxford.
- Kusch, Martin. (2002b). Testimony in communitarian epistemology. *Studies In History and Philosophy of Science Part A*, 33(2), 335-354.
- Kusch, Martin. (2009). Testimony and the Value of Knowledge. *Epistemic value*, 60-94.
- Kusch, Martin. (2012). *Jennifer Lackey on Non-Reductionism: A Critique*. Paper presented at the Epistemology: Contexts, Values, Disagreement: Proceedings of the 34th International Ludwig Wittgenstein Symposium in Kirchberg Am Wechsel, Austria 2011.

- Kusch, Martin. (2013). Naturalized Epistemology and the Genealogy of Knowledge. In M. Lenz & A. Waldow (Eds.), *Contemporary Perspectives on Early Modern Philosophy: Nature and Norms in Thought* (2013 ed., pp. 87-97). New York: Springer.
- Lackey, Jennifer. (1999). Testimonial knowledge and transmission. *The Philosophical Quarterly*, 49(197), 471-490.
- Lackey, Jennifer. (2003). A Minimal Expression of Non-reductionism in the Epistemology of Testimony. *Noûs*, 37(4), 706-723.
- Lackey, Jennifer. (2006a). It takes two to tango: beyond reductionism and non-reductionism in the epistemology of testimony. *The epistemology of testimony*, 160-189.
- Lackey, Jennifer. (2006b). The nature of testimony. *Pacific Philosophical Quarterly*, 87(2), 177-197.
- Lackey, Jennifer. (2006c). Knowing from Testimony. *Philosophy Compass*, 1(5), 432-448.
- Lackey, Jennifer. (2008). *Learning from words: Testimony as a source of knowledge*. Oxford University Press.
- Ladyman, James. (2012). Review of *Naturalizing Epistemology* by Fred D'Agostino. *Australasian Journal of Philosophy*, 90(3), 605-608.
- Levin, Janet. (2013). Functionalism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. <<http://plato.stanford.edu/archives/fall2013/entries/functionalism/>>.
- Maher, Chauncey. (2012). *The Pittsburgh School of Philosophy: Sellars, McDowell, Brandom* (Vol. 42). New York: Routledge.
- Maturana, Humberto. (2002). Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition. *Cybernetics & Human Knowing*, 9(3-4), 3-4.
- Maturana, Humberto R. (1970). Biology of cognition *Autopoiesis and Cognition: The Realization of the Living* (Vol. 42, pp. 2-62). Reprinted in Maturana and Varela 1980: D. Reidel Publishing.
- Maturana, Humberto, and Varela, Francisco. (1980). *Autopoiesis and cognition: The realization of the living* (Vol. 42). Boston: D. Reidel Publishing.
- Mayr, Ernst. (1974). Teleological and teleonomic, a new analysis *Methodological and historical essays in the natural and social sciences* (pp. 91-117). Boston: D. Reidel Publishing.
- McGann, Marek, and De Jaegher, Hanne. (2009). Self-other contingencies: Enacting social perception. *Phenomenology and the cognitive sciences*, 8(4), 417-437.

- Menary, Richard. (2007). *Cognitive integration: Mind and cognition unbounded*. New York: Palgrave Macmillan.
- Menary, Richard. (2009). "Intentionality, cognitive integration and the continuity thesis." *Topoi* 28 (1):31-43.
- Menary, Richard. (2010a). "Dimensions of mind." *Phenomenology and the Cognitive Sciences* 9 (4):561-578.
- Menary, Richard. (2010b). "Introduction to the special issue on 4E cognition." *Phenomenology and the Cognitive Sciences* 9 (4):459-463.
- Menary, Richard. (2012). "Cognitive practices and cognitive character." *Philosophical Explorations* 15 (2):147-164.
- Menary, Richard, and Michael Kirchhoff. (2013). "Cognitive transformations and extended expertise." *Educational Philosophy and Theory: Incorporating ACCESS* 46 (6):610-623.
- Merleau-Ponty, Maurice. ([1945] 2012). *Phenomenology of Perception* (D. Landes, Trans.). New York: Routledge.
- Millikan, Ruth Garrett. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge, MA: MIT Press.
- Millikan, Ruth Garrett. (1989). In defense of proper functions. *Philosophy of Science*, 56(2), 288.
- Millikan, Ruth Garrett. (1993). *White Queen Psychology and other essays for Alice*. Cambridge, MA: The MIT Press.
- Millikan, Ruth Garrett. (1995). "Pushmi-pullyu representations." *Philosophical perspectives* 9:185-200.
- Millikan, Ruth Garrett. (2002). Biofunctions: Two Paradigms. In A. Ariew, R. Cummins & M. Perlman (Eds.), *Functions: New essays in the philosophy of psychology and biology* (pp. 113-143). New York: Oxford University Press.
- Näätänen, Risto, Anne Lehtokoski, Mietta Lennest, Aavo Luuki, Juri Alliki, Janne Sinkkonen, and Kimmo Alho. (1997). "Language-specific phoneme representations revealed by electric and magnetic brain responses." *Nature* 385:432-434.
- Nagel, Ernest. (1977). Teleology revisited. *Journal of Philosophy*, 74(5), 261-301.
- O'Regan, J Kevin, and Noë, Alva. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(05), 939-973.
- Pittendrigh, Colin S. (1958). Adaptation, natural selection, and behavior. *Behavior and evolution*, 390, 416.

- Polanyi, Michael. (1958). *Personal knowledge: Towards a post-critical philosophy*. Chicago: The University of Chicago Press.
- Polanyi, Michael. (1959). *The study of man: The Lindsay memorial lectures*. Chicago: The University of Chicago Press.
- Polanyi, Michael. (1966a). The logic of tacit inference. *Philosophy*, 41(155), 1-18.
- Polanyi, Michael. (1966b). *The tacit dimension*. New York: Doubleday.
- Polanyi, Michael. ([1946] 1964). *Science, faith, and society*. Chicago: The University of Chicago Press.
- Putnam, Hilary. (1960). "Minds and machines." In *Dimensions of Mind*, edited by Sidney Hook, 57-80. New York: New York University Press.
- Putnam, Hilary. (1967a). "Psychological predicates." In *Art, mind, and religion*, edited by W. H. Capitan and D. D. Merrill, 37-48. Pittsburgh: University of Pittsburgh Press.
- Putnam, Hilary. (1967b). "The mental life of some machines." In *Intentionality, Minds and Perception*, edited by Hector-Neri Castaneda. Detroit Wayne State University Press.
- Putnam, Hilary. (1975). "The meaning of 'meaning'." *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The meaning of 'meaning'"*:3-52.
- Putnam, Hilary. (1988). *Representation and reality*. Vol. 454. Cambridge, MA: Cambridge Univ Press.
- Putnam, Hilary. (2012). *Philosophy in An Age of Science: Physics, Mathematics, and Skepticism*. Cambridge, MA: Harvard University Press.
- Quartz, Steven R, and Terrence J Sejnowski. (1997). "The neural basis of cognitive development: A constructivist manifesto." *Behavioral and brain sciences* 20 (04):537-556.
- Rawls, John. (1955). Two concepts of rules. *The philosophical review*, 3-32.
- Rawls, John. (1999). *A theory of justice* (Revised ed.). Cambridge, MA: Harvard University Press.
- Rehg, William. (2009). *Cogent science in context*. Cambridge, MA: MIT Press.
- Roepstorff, Andreas, Jörg Niewöhner, and Stefan Beck. (2010). "Enculturing brains through patterned practices." *Neural Networks* 23:1051-1059.
- Rupert, Robert D. (2004). "Challenges to the hypothesis of extended cognition." *The Journal of philosophy*:389-428.

- Ryle, Gilbert. (1949). *The concept of mind*. Chicago: The University of Chicago Press.
- Sacks, Harvey. (1992). Lectures on conversation (2 volumes). Edited by G. Jefferson, with an introduction by E. Schegloff. Oxford: Blackwell.
- Schmitt, Frederick. (1994). Socializing epistemology: An introduction through two sample issues *Socializing epistemology: The social dimensions of knowledge* (pp. 1-28). Lanham, MD: Rowman & Littlefield Publishers.
- Searle, John R. (1969). *Speech acts: An essay in the philosophy of language*. New York: Cambridge University Press.
- Searle, John R. (1983). *Intentionality: an essay in the philosophy of mind*. New York: Cambridge University Press.
- Searle, John R. (1990). Collective Intentionality and Action. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.), *Intentions in communication* (pp. 401-416). Cambridge, MA: MIT Press.
- Searle, John R. (1995). *The construction of social reality*. New York: Simon & Schuster.
- Selinger, Evan, and Crease, Robert P. (2006). Dreyfus on Expertise: The Limits of Phenomenological Analysis. In E. Selinger & R. P. Crease (Eds.), *The Philosophy of Expertise* (pp. 213-245). New York: Columbia.
- Sellars, Wilfrid. (1956). Empiricism and the Philosophy of Mind. *Minnesota studies in the philosophy of science*, 1, 253-329.
- Sprevak, Mark. (2009). "Extended cognition and functionalism." *Journal of Philosophy* 106 (9):503.
- Staley, Kent W. (2004). *The evidence for the top quark: objectivity and bias in collaborative experimentation*. New York: Cambridge University Press.
- Stanley, Jason, and Williamson, Timothy. (2001). Knowing how. *The Journal of Philosophy*, 98(8), 411-444.
- Steiner, Pierre, and Stewart, John. (2009). From autonomy to heteronomy (and back): The enaction of social life. *Phenomenology and the cognitive sciences*, 8(4), 527-550.
- Sterelny, Kim. (2004). "Externalism, epistemic artefacts and the extended mind." In *The externalist challenge*, edited by Richard Schantz, 239-254. Göttingen, Germany: de Gruyter.
- Stewart, John. (2010). Foundational Issues in Enaction as a Paradigm for Cognitive Science: From the Origin of Life to Consciousness and Writing. In J. Stewart, O. Gapenne & E. Di Paolo (Eds.), *Enaction: Towards a new paradigm for cognitive science* (pp. 1-32). Cambridge, MA: MIT Press.

- Sutton, John. (2010). Exograms and interdisciplinarity: History, the extended mind, and the civilizing process. In R. Menary (Ed.), *The extended mind* (pp. 189-225). Cambridge, MA: MIT Press.
- Sutton, John. (2014). "Remembering as Public Practice: Wittgenstein, memory, and distributed cognitive ecologies." International Wittgenstein Symposium, Kirchberg Am Wechsel.
- Thompson, Evan. (2004). Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela. *Phenomenology and the cognitive sciences*, 3(4), 381-398.
- Thompson, Evan. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.
- Thompson, Evan, & Stapleton, Mog. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi*, 28(1), 23-30.
- Torrance, Steve. (2006). In search of the enactive: Introduction to special issue on enactive experience. *Phenomenology and the cognitive sciences*, 4(4), 357-368.
- Tribble, Evelyn B, and John Sutton. (2012). "Minds in and out of time: memory, embodied skill, anachronism, and performance." *Textual Practice* 26 (4):587-607.
- Turner, Stephen P. (1994). *The social theory of practices : tradition, tacit knowledge, and presuppositions*. Chicago: The University of Chicago Press.
- Turner, Stephen P. (1999). Searle's Social Reality. *History and Theory*, 38(2), 211-231.
- Turner, Stephen P. (2002). *Brains/practices/relativism : social theory after cognitive science*. Chicago: University of Chicago Press.
- Turner, Stephen P. (2010). *Explaining the normative*. Malden, MA: Polity.
- Turner, Stephen P. (2011). Starting with tacit knowledge, ending with Durkheim? *Studies In History and Philosophy of Science Part A*, 42(3), 472-476.
- Turner, Stephen P. (2012a). Making the Tacit Explicit. *Journal for the Theory of Social Behaviour*, 42(4), 385-402.
- Turner, Stephen P. (2012b). The strength of weak empathy. *Science in Context*, 25(03), 383-399.
- Turner, Stephen P. (2014). *Understanding the Tacit*. New York: Routledge.
- Varela, Francisco J. (1979). *Principles of biological autonomy*. New York: North Holland.
- Varela, Francisco J. (1996). The early days of autopoiesis: Heinz and Chile. *Systems Research*, 13(3), 407-416.

- Varela, Francisco J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and cognition*, 34(1), 72-87.
- Varela, Francisco J, Rosch, Eleanor, and Thompson, Evan. (1993). *The embodied mind: Cognitive science and human experience* (3rd ed.). Cambridge, MA: MIT Press.
- Walzer, Michael. (1983). *Spheres of justice: A defense of pluralism and equality*. New York: Basic Books.
- Weber, A., & Varela, F.J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the cognitive sciences*, 1(2), 97-125.
- Wheeler, Michael. (2014). "The revolution will not be optimised: enactivism, embodiment and relationality." 50th Anniversary Convention of the ASIB.
- Wrangham, Richard. (2009). *Catching Fire: How Cooking Made Us Human*. New York: Basic Books.
- Wray, K Brad. (2011). *Kuhn's evolutionary social epistemology*. New York: Cambridge University Press.
- Yancy, George. (2008). *Black Bodies, White Gazes: The Continuing Significance of Race*. Lanham, MD: Rowman & Littlefield.
- Young, Iris Marion. (1990). "The Ideal of Impartiality and the Civic Public." In *Justice and the Politics of Difference*, 96-121. Princeton, NJ: Princeton University Press.