

EXTENDED VIRTUES

by

JOSHUA AUGUST SKORBURG

A DISSERTATION

Presented to the Department of Philosophy
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2017

DISSERTATION APPROVAL PAGE

Student: Joshua August Skorburg

Title: Extended Virtues

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Philosophy by:

Dr. Mark Johnson	Chairperson
Dr. Nicolae Morar	Core Member
Dr. Colin Koopman	Core Member
Dr. Elliot Berkman	Institutional Representative

and

Sara D. Hodges	Interim Vice Provost and Dean of the Graduate School
----------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2017

© 2017 Joshua August Skorburg

DISSERTATION ABSTRACT

Joshua August Skorburg

Doctor of Philosophy

Department of Philosophy

December 2017

Title: Extended Virtues

The central argument of this dissertation is that virtue ethics is overly individualistic. In response, I develop and defend a more relational, ecological account - what I call *extended virtues*. First, following Andy Clark, Kim Sterelny, and others, I argue that cognition and emotion can be embedded in, scaffolded by, or even extended to include various environmental resources. These arguments undermine default internalism about cognitive and affective processes. Next, I show how recent work in social and personality psychology similarly undermines individualism about the bearers of these cognitive and affective processes.

Taken together, these arguments have significant but heretofore underappreciated implications for virtue ethics. After reviewing the literature which attempts to spell out the ethical implications of embedded, scaffolded, and extended cognition, I conclude that a more substantive engagement with virtue ethics is needed. I then show how plausible, mainstream theories of virtue assume default internalism and individualism, and are thus subject to charges of empirical inadequacy.

Finally, I formulate my account of extended virtues in response to these shortcomings. I begin by making three explicit arguments for why an account of extended virtues is needed. I then develop two further arguments - the process argument and the

bearer argument - which yield the conclusion that the processes relevant to, and the bearers of, moral and intellectual virtues can be embedded, scaffolded, or extended. After providing examples and filling in details about the hypotheses of embedded, scaffolded, and extended virtue, I propose that virtues are less like dispositions and more like relations. I conclude by suggesting that ecological metaphors such as stewardship are more fitting than traditional views of morality as inner strength.

CURRICULUM VITAE

NAME OF AUTHOR: Joshua August Skorburg

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Toledo, Toledo
Southern Illinois University, Carbondale

DEGREES AWARDED:

Doctor of Philosophy, Philosophy, 2017, University of Oregon
Master of Arts, Philosophy, 2012, University of Toledo
Bachelor of Science, Journalism, Philosophy, 2009, Southern Illinois University

AREAS OF SPECIAL INTEREST:

Moral Psychology
Applied Ethics

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, University of Oregon, 2013-2017
Teaching Assistant, University of Toledo, 2010-2012

GRANTS, AWARDS, AND HONORS:

Dissertation Fellowship, "Extended Virtues", Oregon Humanities Center, 2017
Research Grant, "Beyond the moral self", The John Templeton Foundation, 2017
Research Grant, "Giving from the heart", The Templeton Religious Trust, 2015
Philosophy Matters Prize, "Implicit Bias, Epistemic Injustice, and the Epistemology of Ignorance", University of Oregon Dept. of Philosophy, 2015
Ila and John Mellow Prize, "Beyond Embodiment: John Dewey and the Integrated Mind", Society for the Advancement of American Philosophy, 2013

PUBLICATIONS:

Skorburg, J.A. & Alfano, M. (forthcoming). Implications for virtue epistemology from psychological science: intelligence as an interactionist virtue. In H. Battaly (ed.), *Handbook of Virtue Epistemology*. New York: Routledge.

Alfano, M. & Skorburg, J.A. (forthcoming). Extended knowledge, the recognition heuristic, and epistemic injustice. In D. Pritchard et al. (eds.), *Extended Knowledge*. Oxford University Press.

Skorburg, J.A. (2017). Jane Addams as experimental philosopher. *British Journal for the History of Philosophy*. <http://dx.doi.org/10.1080/09608788.2017.1336985>

Skorburg, J.A. (2017). Lessons and new directions for extended cognition from social and personality psychology. *Philosophical Psychology*. <http://dx.doi.org/10.1080/09515089.2017.1282606>

Skorburg, J.A. (2017). Chris Abel, “The extended self: Architecture, memes, and minds.” *Environmental Philosophy*, 14(1), 151-53.

Morar, N. & Skorburg, J.A. (2017). Relational Agency: Yes - but how far? *American Journal of Bioethics Neuroscience*, 8(2), 83-85.

Skorburg, J.A. (2016). Pragmatism, embodiment, extension. In R. Madzia & M. Jung (eds.), *Pragmatism and Embodied Cognitive Science: From bodily intersubjectivity to symbolic articulation*. 33-54. Berlin: De Gruyter.

Morar, N. & Skorburg, J.A. (2016). Toward an ecological bioethics. *American Journal of Bioethics*, 16(5), 35-37.

Alfano, M. & Skorburg, J.A. (2016). The embedded and extended character hypotheses. In J. Kiverstein (ed.), *The Routledge Handbook of Philosophy of the Social Mind*. 465-478. New York: Routledge.

Skorburg, J.A. (2013). Beyond embodiment: John Dewey and the integrated mind. *The Pluralist*, 8(3), 66-78.

ACKNOWLEDGMENTS

From my arrival at the University of Oregon to the completion of this dissertation, Dr. Mark Johnson has been an exemplary mentor, interlocutor, and advisor. His diligence and his commitment to doing meaningful philosophy has provided a model for his students and colleagues. I greatly appreciate his insistence on stating my dissertation's arguments in strong, confident terms, and also his patience with my overly wide-ranging research interests. Dr. Nicolae Morar has helped me turn half-baked ideas into a research program. I am beyond grateful for his personal and professional mentorship, and for our collaborations and friendship. Dr. Colin Koopman sets a high bar for scholarship and professional comportment. I hope someday my students will speak half as highly of me as his speak of him. My arguments are sharper as a result of what I've learned from Dr. Elliot Berkman, who has been a supportive, open-minded, and generous guide for my forays into psychology and neuroscience. Dr. Christina Karns has been a wonderful collaborator, teacher, confidant, and friend. Dr. Anthony Chemero's work has long inspired my own. Much of the present project has grown out of collaborations with Dr. Mark Alfano, from whom I have learned much. None of this would have been possible, however, without the unwavering and loving support of my partner, Kristen Reinhardt. Ever the optimist, she is also the strongest and kindest person I know. Finally, there is a real, material sense in which the completion of this dissertation on a tight deadline was made possible by the support of the Oregon Humanities Center, which generously awarded me a Dissertation Research Fellowship during Fall term, 2017. The OHC staff have been remarkably accommodating, and they have significantly lessened the burden of transitioning from dissertation to post-doctoral research.

For my parents, Mary Ann and Jon.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
I.I Introduction	1
I.II Roadmap	3
I.III A note on Terminology and Scope	6
II. PHILOSOPHY OF MIND	9
II.I Introduction.....	9
II.II Historical Backdrop: Pragmatism and Externalism.....	9
II.III. Clark & Chalmers’s “The Extended Mind” and Aftermath.....	15
II.III.I. The Embedded Mind Hypothesis.....	23
II.III.II. The Scaffolded Mind Hypothesis	25
II.III.III. The Extended Cognition Hypothesis.....	29
II.IV. Template.....	34
II.IV.I. Embedding.....	34
II.IV.II. Scaffolding	36
II.IV.III. Extending	37
II.V. Recent Developments.....	38
II.VI. Conclusion	44

Chapter	Page
III. SOCIAL AND PERSONALITY PSYCHOLOGY	45
III.I Introduction	45
III.II The Person-Situation Debate.....	46
III.III From Interactionism in Psychology to Philosophy.....	51
III.IV The Next Wave.....	60
III.V Conclusion.....	74
IV. REVIEW.....	78
IV.I Introduction	78
IV.II Sneddon's <i>Like Minded</i>	79
IV.III Applying the Extended Mind: Enhancements and Assaults.....	87
IV.IV Extending Autonomy, Identity, and the Self	96
IV.V Extending Self, Identity, and the Bearers of Psychological States.....	102
IV.VI Conclusion	112
V. VIRTUE ETHICS.....	114
IV.I Introduction.....	114
IV.II What are Virtues?.....	115
V.III. Are Dispositions Internalist and Individualist?.....	123
IV.IV Conclusion.....	114
VI. EXTENDED VIRTUES.....	132
VI.I Introduction.....	132
VI.II Why Bother?.....	132
VI.III Two Arguments for Extended Vitues.....	137

Chapter	Page
VI.IV Embedded, scaffolded, and extended virtues	132
VI.IV.I Embedded Virtues	139
VI.IV.II Scaffolded Virtues	143
VI.IV.III Extended Virtues	145
VI.V A New Metaphysics of Virtue	149
VI.VI Toward an Ecological Virtue Ethics.....	160
VI.VII Conclusion.....	168
REFERENCES CITED.....	172

CHAPTER I

INTRODUCTION

I.I Introduction

As of this writing (September, 2017), there are roughly 7.5 billion people on earth and roughly 8.2 billion active mobile devices (smartphones, tablets, etc.). 77% of Americans own smartphones (Pew Research Center, Jan. 12, 2017). U.S. adults and teenagers check their phones up to 150 times a day, or roughly every seven minutes they are awake. Most do not go more than an hour without checking their phone. 75% feel panicked when unable to locate their phone (Gazzaley & Rosen, 2016, p. 11). For many teens, the phone is the last thing they see at night, and the first thing they see in the morning (Twenge, 2017). 46% of smartphone owners say their phone “is something they couldn’t live without” (Pew Research Center, June 28, 2017).

On one account of human agency I find particularly compelling, these statistics are completely unsurprising. Smartphones, on this view, are among the suite of tools and props fluently incorporated into the shifting assemblage that is the human mind:

Different neural circuits provide different capacities, and contribute in different ways to our sense of self, of where we are, of what we can do, and to decision-making and choice. External, nonbiological elements provide still further capacities, and contribute in additional ways to our sense of who we are, where we are, what we can do, and to decision-making and choice. But no single tool among this complex kit is intrinsically thoughtful, ultimately in full control, or plausibly identified as the inner ‘seat of the self’. *We (we human individuals) just are these shifting coalitions of tools. We are ‘soft selves,’ continuously open to*

change and driven to leak through the confines of skin and skull, *annexing more and more nonbiological elements as aspects of the machinery of mind itself*

(Clark, 2007b, p. 112, emphasis mine).

Far from being squishy and empirically intractable, I will show how this ‘soft selves’ view follows very naturally from a consideration of recent work in the philosophy of cognitive science and social and personality psychology. My primary aim is to lay the groundwork for an ethical theory – a virtue ethical theory, more specifically – that does justice to this view of human agency and the human mind, but with the crucial caveat that we are continuously open to influence not just from gadgets like smartphones, but also other agents, and especially our close friends and romantic partners. All of this is at odds with views of monadic, fully autonomous moral agents that have tended to dominate Western ethical theorizing.

While virtue ethics (at least the Aristotelian variety) tends to be individualistic in just this way, I contend that it is nonetheless well-positioned to adapt to the emerging picture of soft selves, with smudged boundaries between mind/world, organism/environment, and person/situation. Thus, in much the same way philosophers like Andy Clark have argued that various cognitive processes might be embedded in, scaffolded by, or extended to include various environmental resources, I will argue that components of virtues, too, can be embedded in, scaffolded by, or extended to include environmental resources like smartphones, but more importantly, also other agents. On this view, we are all entrusted with the moral and intellectual character of our friends and partners - we really are our brothers’ and sisters’ keepers. This line of argumentation will render impotent some traditional ways of thinking about the will, virtues, moral agency,

autonomy, and the role of context. The payoff, however, will be a theoretical framework able to capture what I take to be a particularly compelling insight that “the right unit of attention for ethics is the whole person-in-communal relations,” where ethics (including moral philosophy and moral psychology) is about “persons who seek to live well in relations with other persons in particular natural and social ecologies with histories” (Flanagan, 2016, p. 21).

I.II Roadmap

I have divided the text in two Parts, with three Chapters in each. I begin in Chapter II by tracing the development of a certain strand of externalism in the philosophy of mind and cognitive science which begins with the classical American pragmatists (Dewey, 1896, 1916; Peirce, 1903) runs through the more familiar debates sparked by Clark & Chalmers’s (1998) “The Extended Mind” paper, and then reaches a mature formulation in terms of *interactionism* in the work of Palermos (2014), Heersmink (2015), and others. I mark this principled conclusion by developing a template for thinking about the various dependence relations between agents and environments in terms of embedding, scaffolding, and extending (this template is then used to formulate my account of extended virtues in Chapter VI). I conclude Chapter II by considering recent developments suggesting that not only might cognition be embedded, scaffolded, or extended, but so too for affectivity and the emotions (Colombetti & Roberts, 2015; Carter, Gordon, & Palermos, 2016). Taken together, these arguments aim to unseat default internalism about the cognitive and affective processes relevant to virtue ethics.

In Chapter III, I argue that much of the extended mind and cognition research is overly focused on agent-artifact coupled systems (e.g. smartphones, Tactile Visual

Substitution Systems). I then introduce research in social and personality psychology which I take to be a better fit for the kind of *interactionism* that many extended mind and cognition theorists are after. I consider the research on situation selection (Ickes, Snyder, & Garcia, 1997), transactive memory systems (Wegner, Erber, & Raymond, 1991; Huebner, 2016), the Michelangelo effect (Drigotas, Rusbult, Wieselquist, & Whitton, 1999; Rusbult, Finkel, & Kumashiro, 2009), and transactive goal dynamics (Fitzsimons, Finkel, & van Dellen, 2015) to unseat default individualism about the bearers of psychological states relevant to virtue ethics.

The conclusion of the first half of the dissertation then, is that the views of cognition, emotion, personality, memories, goals, and relationships presented therein have significant but heretofore underappreciated implications for virtue ethics. The second half of the dissertation then aims to articulate and respond to a resultant two-fold problem. On the one hand, extended cognition theorists have not substantively engaged with normative ethical theories such as virtue ethics. On the other hand, despite decades of calls for doing empirically informed moral philosophy, virtue ethicists, on the whole, have not substantively engaged with the empirical research surveyed Chapters II and III.

In Chapter IV, I argue that while proponents of extended cognition have identified some promising lines of inquiry into the ethical implications of their views, the research remains underdeveloped and unsatisfying. I survey this literature, including Andrew Sneddon's (2011) book *Like Minded: Externalism and Moral Psychology*, Neil Levy's (2007a,b) work on the applied ethical dimensions of extended cognition, Mason Cash's (2010, 2013) dialogue with feminist accounts of self, identity, and autonomy, and finally, the arguments for extended persons developed by Robert Howell (2016) – which are

closest to my own. In highlighting the strengths and weaknesses of these accounts, I conclude that while there is much promise in the work surveyed in this chapter, the literature is just that: promissory. Identifying a lack of substantive engagement with virtue ethical traditions then sets the stage for my positive account in the last two chapters.

In Chapter V, I argue that much contemporary virtue theorizing is internalist and individualist in precisely the ways that the arguments from Chapters II and III identify as inadequate. I begin by establishing the standard answer to the question, “What are virtues?” I provide substantial evidence in support of the claim that virtues have three features: (1) virtues are dispositions, (2) virtues have a cognitive component, and (3) virtues have an affective component. I show how this holds true for both historical accounts of ancient theories of virtue (Annas, 1993), as well as contemporary accounts informed by the psychological sciences (Miller, 2013, 2014). The argument here is that the cognitive and affective processes posited by plausible, mainstream virtue theories *just are* what the arguments from Chapters II and III call into question.

Finally, in Chapter VI, I formulate and defend a novel account of extended virtues which aims to address the shortcomings identified in Chapter IV while also responding in a satisfactory and empirically responsible way to the challenge formulated in Chapter V. I begin by making an explicit argument for why an account of extended virtues is needed. I then develop two arguments for extended virtues: The *process* argument and the *bearer* argument. These yield the conclusion that the processes relevant to, and the bearers of, moral and intellectual virtues can be embedded/scaffolded/extended. Then, drawing on the template from Chapter II, I fill in the details about the hypotheses of

embedded/scaffolded/extended virtues. Next, I argue that the standard view of virtues as dispositions is inadequate in light of these arguments, and I propose that virtues are better understood as *relations*. I then conclude by suggesting that ecological metaphors offer the most promising framework for conceptualizing extended virtues.

I.III A note on terminology and scope

It is worthwhile to clarify a few points at the outset. This dissertation is entitled *Extended Virtues*. In the first place, what do I mean by “extended”? I noted above that I will be tracing a strand of externalism in the philosophy of mind in Chapter II. What I have in mind here is primarily the brand of *active* externalism advocated in Clark & Chalmers (1998), and pre-figured in the work of John Dewey, Charles Peirce, and others. Thus, I will have nothing to say about the *content* externalism developed by Tyler Burge and Hilary Putnam and there will be no trips to Twin Earth. Nor will I have anything to say about the internalism/externalism debate in meta-ethics. The argumentative strategy of this dissertation is to apply the insights from the extended mind and cognition literature to recent work in empirically informed virtue theory.

What, then, do I mean by “virtue”? There is a sense in which every ethical tradition has a theory of virtue in the sense of a theory about what makes a good person a good person. For deontologists, the virtuous agent has the strength to do what duty requires. For utilitarians, the virtuous agent maximizes net benefits for the greatest number. For Aristotelians, the virtuous agent cultivates a character from which right action naturally flows. Construed in this way, it would be naïve – or better, imperialistic – to think that the arguments developed here will target all virtue theories in this capacious sense. Out of necessity then, my focus will have to be limited, and so I will be mostly

restricted to recent work in (neo)Aristotelian virtue ethics. There are a few reasons for this.

First, I agree with Anscombe (1958) that an adequate philosophy of psychology is needed in order to do good moral philosophy. I subscribe to what Flanagan (1991) calls the Principle of Minimal Psychological Realism and what Lakoff & Johnson (1999) call empirically responsible philosophy. Or more colloquially, I agree with Doris (2002, p. 1) that “thinking productively about ethics requires thinking realistically about humanity.” I will not so much argue for the value of doing philosophy in this way as I will assume it as my starting point. Because the commitment to doing this kind of philosophy has been fairly reliably enacted in recent work on (neo)Aristotelian virtue ethics, this is where my project is situated. As such, I want to acknowledge that by virtue of this situatedness, I will be attuned to some issues and blind to many others. For example, I doubt the arguments I develop here against internalism and individualism would hit their mark in the same way against say, Buddhist or Confucian views of virtue. It is my hope that the recognition of this limited perspective can help to guard against narrow-minded parochialism.

Seen from the other side, it is also worth noting that the conclusions I arrive at (e.g., failures of individualism in ethics) are by no means novel. What will emerge from these pages, I hope, is an appreciation for the many ways in which moral agents depend on contextual features external to themselves, especially other agents. On the view I am advocating, there is a profound sense in which constitutive features of the moral life *simply are not up to us*, and even the strongest will cannot change this fact. For this to be so, moral agency must be cast in terms of vulnerability and interdependence, with various

kinds of interpersonal processes and social relationships taking center stage. If this sounds like a flat-out rejection of a certain Kantian view¹ and a nutshell summary of various kinds of feminist or relational ethics, this is by design. Insofar as this dissertation makes a novel contribution, it does so by using new premises - built from research in the philosophy of cognitive science and social and personality psychology – to undermine a monadic view of moral agency and support broadly relational approaches to ethics.

With these clarifications and qualifications in place, I will now try to convince readers that traditional internalist and individualist ways of thinking about virtue are empirically inadequate and theoretically unattractive. In their place, I will offer a broadly ecological, relational picture of virtue which casts moral agents not just *in* the world, but *of* the world.

¹ Namely, Kant's view that "even if, by some especially unfortunate fate or by the niggardly provision of stepmotherly nature, this will should be wholly lacking in the power to accomplish its purpose; if with the greatest effort it should yet achieve nothing, and only the good will should remain (not, to be sure, as a mere wish but as the summoning of all the means in our power), yet would it, like a jewel, still shine by its own light" (1994, pp. 7-8).

CHAPTER II

PHILOSOPHY OF MIND

II.I. Introduction

This chapter fashions the lens through which the rest of the material in Part I is examined. The aim here is to trace the development of a certain strand of externalism in the philosophy of mind which begins with the classical American pragmatists, runs through the more familiar debates sparked by Clark & Chalmers's (1998) "The Extended Mind" paper, and then reaches a mature formulation and branches out to other sub-disciplines and research programs.

My central claim here is that the most productive way to understand the trajectory of the extended mind is through what I will call the coupled systems framework, exemplified by Clark & Chalmers's claim that, in cases of extended cognition "the human organism is linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right" (1998, p. 8).

Within this framework, a number of nuanced distinctions emerge which account for the various dependence relations between agents and environmental resources: staunch *internalism* (Adams & Aizawa, 2009), the hypothesis of *embedded cognition* (Rupert, 2009), the hypothesis of *scaffolded cognition* (Sterelny, 2010), the hypothesis of *extended cognition* (Palermos, 2014), and the hypothesis of the *socially extended mind* (Gallagher, 2014). These distinctions will provide a template for exploring research in social and personality psychology in the remainder of Part I, as well as the various dependence relations between virtues and contexts in Part II.

II.II. Historical backdrop: Pragmatism and externalism

The standard story in the philosophy of mind is that the hypothesis of the extended mind was born with Clark & Chalmers's (1998) now (in)famous thought experiment involving Otto and Inga trying to find the Museum of Modern Art. This is, strictly speaking, true, in the sense that Clark & Chalmers's paper was the first to explicitly articulate the hypothesis that "cognitive processes ain't (all) in the head" (p. 8).² I will consider the specifics of this paper in the next section, but my aim in the present section is to illustrate the historical depth of this idea of extended mind and cognition. More specifically, I will try to show how the extended mind has its roots in early 20th century philosophy and psychology generally, and in American pragmatism, more specifically.

To begin, we can look to the following passage from John Dewey which seems to straightforwardly anticipate the hypothesis of extended cognition by nearly a century:

Thinking, or knowledge getting, is far from being the armchair thing it is often supposed to be. The reason it is not an armchair thing is that it is not an event going on exclusively within the cortex.... Hands and feet, apparatus and appliances of all kinds are as much a part of it as changes within the brain (Dewey 1916/2004, p. 9).

Charles Saunders Peirce echoes a similar skepticism about internalist views of mind:

Modern philosophy has never been able to quite shake off the Cartesian idea of the mind, as something that "resides," – such is the term, – in the pineal gland.

Everybody laughs at this nowadays, and yet everybody continues to think of mind

² Though Clark & Chalmers's slogan for *active* externalism here is, of course, paying homage to Hilary Putnam's (1975) brand of *content* externalism: "cut the pie any way you like, meaning just ain't in the head!" (p. 227). For more on the distinction between content and active externalisms, see note 6 below.

in this same general way, as something within this person or that, belonging to him and correlative to the real world (Peirce 1903/1998, p. 199).

Many more passages could be enumerated from both Dewey's and Peirce's work, as well as many of their contemporaries, including William James and George Herbert Mead.³

The point of highlighting these passages is to motivate the present historical detour. If the pragmatists were making claims like these, which sound an awful lot like the claims made by extended mind theorists today, then what else might pragmatism contribute to contemporary debates? What sorts of metaphysical and epistemological views did they hold, such that they thought of the mind in the way that they did?

I will focus here specifically on John Dewey, and especially on his view of organism-environment transactions. This philosophical outlook helped pave the way for much of the theorizing that now falls under the "4E" moniker: Mind as embedded, embodied, enactive, or extended.⁴ I am motivated here by a conviction that in order to understand the arguments for the extended mind, we ought to try to understand the philosophical and psychological lineages that helped to make those arguments viable.

One of Dewey's earliest and most influential articles was the "Reflex Arc Concept in Psychology" (1896/2008) in which he criticized the received model of

³ See Chemero (2011, pp. 18-19) and Wagman & Chemero (2014, pp. 106-09) for a discussion of James's anticipation of and contribution to aspects of the extended mind debates. See Madzia (2013) for a discussion of Mead's relationship to the extended mind.

⁴ Such catchy labels are bound to create as much confusion as clarity. Nonetheless, in the Introduction to a special issue of *Phenomenology and the Cognitive Sciences* on the 4E conception of the mind, Richard Menary writes: "one reason that the four E's are grouped together is that they are all held to reject or at least radically reconfigure traditional cognitivism, coupled with a methodological individualism" (2010a, p. 459).

stimulus-response theories of perception and action.⁵ In the essay, Dewey recalls an example from William James's *The Principles of Psychology* (1890/1981) where a curious young child notices the flame from a candle, reaches for it, burns their hand, and then retracts it. On the received view Dewey is criticizing, an explanation would look something like the following: The light from the candle is the initial visual stimulus and the reaching is the initial motor response. Then, the burning of the hand is the next stimulus which gives rise to the next motor response of withdrawing the hand. Dewey's criticism is that this view construes action and perception as "a patchwork of disjointed parts, a mechanical conjunction of unallied processes" (p. 97). His positive claim is that "both sensation and movement lie inside, not outside the act" (p. 98).

From this vantage point, the explanation begins not with the passive reception of the visual stimulus, but rather with the engaged act of seeing. The prior engagement of various motor processes (e.g. tilting the neck, straining the eyes, etc.) determines which stimuli "get through" in the first place. This information then shapes what it seen, and hence, what will be reached for: "The act is seeing no less than before, but it is now seeing-for-reaching-purposes. There is still a sensori-motor circuit, one with more content or value, not a substitution of a motor response for a sensory stimulus" (p. 98).

This account effectively collapses the dualism between stimulus and response, and replaces it with what Dewey variously calls sensori-motor coordinations, larger organic wholes, or concrete unities. In any case, the implications of this argumentative

⁵ Thomas Alexander and Larry Hickman write, "'The Reflex Arc Concept in Psychology' (1896) is one of Dewey's most famous essays. It signaled the end of introspectionist psychology and the beginning of a new functional, organic, social behaviorism. In 1942, a committee of seventy eminent psychologists polled by the editors of *The Psychological Review* voted this essay the most important contribution to the journal during its first 49 years of publication" (p. ix).

strategy reach all the way to contemporary debates surrounding the so-called 4E conception of the mind. To see this, note how Dewey claims that stimulus and response “mean distinctions of flexible function only, not of fixed existence; that one and the same occurrence plays either or both parts, according to the shift of interest...The fact is that stimulus and response are not distinctions of existence, but teleological distinctions, that is, distinctions of function, or part played, with reference to reaching or maintaining and end” (p. 102-3).

The point is that there is no clean, hard distinction between stimulus and response. Motor activity is always actively sculpting the incoming sensory array and this information is always informing the motor activity, which further refines the sensory input, and so on. The resultant collapse of the dualism between stimulus and responses is a classic Deweyan outcome. He makes this very same argument with respect to the mind/body and individual/society distinctions throughout his corpus. But more importantly for present purposes, Dewey also deploys this argument against the organism/environment distinction. Consider the following passages from his lesser known *Lectures on Ethics* (1901/1991):

So it is not simply that we happen to have an organism drop down into an environment and then these two react upon each other. It is quite the opposite. Organism and environment are the two things which converge in the life process. We do not begin with the two things and have them react and produce the life process (p. 364).

And from the (1898/1976) *Lectures on Psychological and Political Ethics*:

No matter how much one school may insist on the organism, no matter how much another school may insist on the environment, it is certainly true that both are parts of a common world. . . . From the larger standpoint we must have one thing, one reality, the world at large, and the distinction between the organism and the environment and their adjustments to each other must be capable of definition and interpretation from the standpoint of this larger whole (p. 272).

My claim here is that *this line of argumentation paves the way for a whole host of extended mind-style arguments*. That is, without Dewey's argument (or at least an argument in a Deweyan vein) for the flexibility and porosity of the organism/environment interface, it is hard to see how many 4E arguments can get off the ground. In a contemporary context, this Deweyan view is nicely summed up by Tom Alexander:

The organism and its environment are mutually implicated at each moment; they are aspects of one situation fundamentally related through the act. The organism is just this ability to draw on a range of material in the world and transform the energy in that material into an organized pattern of activity. An environment is in turn that range of energy which is available to the organism and necessary for its survival" (1987, p. 135).

It is this very same functional understanding of stimulus/response and by extension, organism/environment which underwrites the arguments for Dawkins's (1999) notion of an extended phenotype, Clark's (1989) 007 principle, Rowlands's (2003) barking dog principle, Clark & Chalmer's (1998) parity principle, and Menary's (2007) notion of cognitive integration.

The take home message here is not just that “Dewey got here first,” but rather that many of the claims made by externalists about the mind have a richer historical lineage than one might be led to believe. My own view is that to ignore this lineage is to miss out on potential resources for advancing contemporary debates. So with this brief historical detour completed, I will now turn to the specifics of the contemporary arguments, which I hope to have shown here, are importantly pre-figured in early 20th century American philosophy and psychology.⁶

II.III. Clark & Chalmers’s “The extended mind” and aftermath

My immediate aim in this section is to clarify my conception of externalism in the philosophy mind by considering the major positions staked out in debates following the publication of Clark & Chalmers’s (1998) “The Extended Mind.” Intermediately, the concepts from this debate will be used to inform structurally similar debates in social and personality psychology in the next chapter. Looking ahead even further, these positions and arguments will inform my account of extended virtues in the dissertation’s second half.

I do not aim to survey the vast body of literature that has developed around the extended mind. Instead, I will offer a selective synthesis which serves the ends just identified. Nevertheless, my story starts where most do, with the introduction of Otto and

⁶ I would be remiss not to note that my transition from classical pragmatism to Clark & Chalmers’s brand of *active* externalism omits the very brand of externalism that most philosophers probably associate with that term. Namely, the veritable cottage industry developed from Hillary Putnam and Tyler Burge’s *Twin Earth* chronicles (e.g. Putnam, 1975). Rowlands (2003) helpfully marks the difference between these kinds of externalism by distinguishing *content* from *vehicle*. Content externalism claims that “the semantic content of mental states that have it is often dependent on factors – objects, properties, events and so on – that are external to the subject of that content” (p. 5). Vehicle externalism, which is more or less synonymous with active externalism, claims that “the structures and mechanisms that allow a creature to possess or undergo various mental states and processes are often structures and mechanisms that extend beyond the skin of that creature” (p. 6). In the present context, my project obviously falls squarely into the vehicle externalist camp and I will not have much to say (directly, anyway) about content. See Hurley (2010) and Sneddon (2008, 2011) for more detailed externalist taxonomies and distinctions.

Inga. Clark & Chalmers ask readers to embark on a thought experiment where Otto and Inga both hear about an exhibit at the Museum of Modern Art that they wish to go see. In just the way you would expect, Inga recalls from her intact, normally functioning biological memory that the museum is on 53rd St. and she goes there.

Otto, however, suffers from Alzheimer's, and for a long time has written down in his notebook the sorts of things that people like Inga would just commit to their biological memory. So when Otto hears about the art exhibit, he checks his notebook, finds the page where it says, "the MoMA is on 53rd St." and he goes there. Insofar as both arrive at the museum, "in relevant respects," according to Clark and Chalmers (1998), "the cases are entirely analogous: the notebook plays for Otto the same role that memory plays for Inga. The information in the notebook functions just like the information constituting an ordinary non-occurrent belief; it just happens that this information lies beyond the skin" (pp. 12-3).

There has been much debate in the secondary literature about precisely what the relevant respects of the analogy are, but what is more important for present purposes are the two strands of argumentation that evolve from this thought experiment: (1) the parity principle and (2) the coupled systems formulation. I will explicate each in turn. Here is how Clark & Chalmers describe (1):

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 (so we claim) is part of the cognitive process. Cognitive processes ain't (all) in the head! (1998, p. 8).

According to Menary (2010b), this is not so much an argument for the extended mind as it is an intuition pump – a way to disabuse readers of residual Cartesian prejudices about the interiority of the mind. The implicit claim in the parity principle is that the location of a process is not sufficient to determine whether that process is a cognitive one.⁷ The parity principle also serves as a hinge to connect the extended mind arguments with the more widely accepted functionalist positions in the philosophy of mind. In this sense, so long as a cognitive process has a defined function, its location and constitution are irrelevant.

Even with functionalist credentials, however, this formulation of the extended mind has come under harsh criticism – and for good reason. By focusing primarily on the similarity of internal and external processes, the parity principle is likely to be undercut by the plethora of counterexamples which posit relevant differences between internal and external processes, not the least of which is the very case of memory employed by Clark & Chalmers. Indeed, one can plausibly argue that what is true of Inga’s memory only holds for Otto’s notebook at such a coarse-grained level of description that it is explanatorily vacuous.

For example, a common objection is that Otto’s “remembering” requires perceiving sentences in his notebook, where normal biological recall requires no such perception. Insofar as it is useful to distinguish between memory and perception, it will

⁷ It is no small task in this debate to define in a non-question-begging way what kinds of processes are cognitive and non-cognitive. But for the narrow purposes of this subsection (e.g., tracing the trajectory of the debates sparked by Clark & Chalmers), by “cognitive processes,” I have in mind the sorts of processes that internalists and externalists alike should be able to agree are paradigmatically cognitive: planning, problem solving, remembering, etc. In this sense, the debates are not so much about whether these processes are cognitive, but rather, whether they are fully realized within the skin and skull of an individual. As we will see below, however, the same argumentative structure deployed by Clark & Chalmers has also been deployed in other domains, such as affectivity and the emotions. While questions about whether or not emotions are properly cognitive may arise in these contexts, such questions are orthogonal to aims of the present section.

be a strike against any theory which requires collapsing across them, as the argument from coarse-grained functional similarity seems to do. Similar counterarguments positing relevant differences between Otto and Inga can be mounted with respect to belief updating, introspection, forgetting, etc. While I think these kinds of criticisms probably hit their mark against coarse-grained parity formulations of the extended mind, it is important to note that the arguments from functional parity are not the only games in town. Sutton (2010), Menary (2007), and others argue that the parity principle should not be seen as the conceptual center of the extended mind, and should instead be eschewed in favor of various forms of *complimentarity*. Here is Sutton's gloss:

first-wave EM [extended mind] is based on the parity principle: cognitive states and processes extend beyond the brain and into the (external) world when the relevant parts of the world function in the same way as do unquestionably cognitive processes in the head...second-wave EM is based on a complementarity principle: in extended cognitive systems, external states and processes need not mimic or replicate the formats, dynamics, or functions of inner states and processes. Rather, different components of the overall (enduring or temporary) system can play quite different roles and have different properties while coupling in collective and complementary contributions to flexible thinking and acting (2010, pp. 193-4).

Proponents of the complimentarity view see this formulation as more plausible than the parity principle in advancing extended mind arguments. This much seems right, but my own view is that any offshoot from the parity principle – complimentarity or otherwise – will not provide the most fruitful way forward for the extended mind hypothesis, as the

emphasis in these formulations still tends to be on the intrinsic properties of the components, rather than the nature of the coupling relation between them. It is for this reason that we need to turn to the related but distinct formulation of the extended mind in terms of (2) coupled systems. Here is the original formulation in Clark & Chalmers (1998):

In these cases [of extended cognition], the human organism is linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right. All the components in the system play an active causal role, and they jointly govern behaviour in the same sort of way that cognition usually does. If we remove the external component the system's behavioural competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head (pp. 8-9).

This view focuses on the precise nature of the relationship between the cognitive agent and the external entities to which the agent is coupled. Thus, the arguments for extension do not trade in terms of parity or complementarity, but rather, in terms of what Clark has elsewhere called "continual reciprocal causation" – or what I will later call symmetry and balance of pathways of influence.

Put another way, on the coupled systems formulation, the focus is on the nature of the *relation* between components of a coupled system rather than on features of the components like similarity. Shifting the emphasis in this way comes with its own set of difficulties, however. Most notably, the arguments for the extended mind in terms of

coupled systems have faced the persistent charge of what Addams & Aizawa (2010) call the coupling-constitution fallacy:

When Clark makes an object cognitive when it is connected to a cognitive agent, he is committing an instance of a coupling-constitution fallacy. This is the most common mistake that extended mind theorists make. The fallacious pattern is to draw attention to cases, real or imagined, in which some object or process is coupled in some fashion to some cognitive agent. From this, one slides to the conclusion that the object or process constitutes part of the agent's cognitive apparatus or cognitive processing[...] Yet coupling relations are distinct from constitutive relations, and the fact that object or process X is coupled to object or process Y does not entail that X is part of Y (pp. 67-8).

Taken in isolation, this criticism is surely a valid one – coupling relations and constitution relations are importantly different kinds of things. The real issue, of course, is whether the criticism hits its mark with respect to specific extended mind arguments. When extended mind proponents present putative cases of extended cognition, the claim is that some external entity is a constitutive component of the agent's cognitive processing. The line of objection presented here by Addams & Aizawa, and others is that many (if not all) of those putative cases involve mere coupling between agents and external entities, not literal constitution.

Seen from the other side, the coupling constitution fallacy highlights the need for a principled set of criteria to distinguish mere coupling from full-blown constitution. Without such criteria, an offshoot of the coupling-constitution charge looms large and that is what Rowlands (2013) and others have called the problem of cognitive bloat:

“Once we permit such [extended cognitive] processes, where do we stop? Our conception of the cognitive will become too permissive, and we will be forced to admit into the category of the cognitive all sort sorts of structures and processes that clearly are not cognitive” (p. 86).⁸

On my reading, responses to some combination of the coupling-constitution and cognitive bloat charges comprise the majority of the substantive work in the extended mind debates. Much ink has been spilled over the last decade in attempts to specify the precise conditions under which a cognitive process can be said to be extended. My next task is to spell out, from within the coupled systems framework, some positions that have been staked out between mere coupling and full-blown constitution.

At the heart of the debate over extended minds are questions about which criteria are used to determine when the components of a coupled system are merely coupled, and when the components are constitutive of an extended system. Though many proposals have been made in the literature (e.g., Clark & Chalmers (1998); Wilson & Clark (2009); Sterelny (2010); Sutton et al. (2010)) the most sophisticated is probably Heersmink’s (2015) dimensional analysis:

By synthesising and building on their [the just-mentioned citations] work, I develop a multidimensional framework for conceptualizing the different kinds of cognitive interactions and complementary integration between agents and

⁸ The force of this criticism is usually cashed out as a *reductio ad absurdum*. Rowlands continues: “If we are willing to allow that the sentences in Otto’s notebook are beliefs, why stop there? Why not also allow that status for the entries in the telephone directory of which Otto makes frequent use? Why cannot these be numbered among Otto’s beliefs? Indeed, why stop even there? Why does Otto not believe everything contained on the Internet, given that he is able to use this in a way akin to which he uses his notebook?” (p. 93). Because we would all agree that beliefs are cognitive, and because we would all agree that it is absurd to hold that the entire Internet stores an individual’s non-occurrent beliefs, the objection goes, any theory which entails that the entire Internet is part of an individual’s cognitive processing must be rejected.

artifacts. This framework consists of the following dimensions: information flow, reliability, durability, trust, procedural transparency, informational transparency, individualization, and transformation. These dimensions are all matters of degree and jointly they constitute a multidimensional space in which situated cognitive systems can be located and have certain dimensional configurations. The higher a particular system scores on these dimensions, the more deeply the functional integration is (p. 579).

The debate surrounding different versions of these criteria have given rise to a set of nuanced distinctions between the poles of mere coupling and full-blown constitution. Elucidating these distinctions is a central aim of the present chapter, as the distinctions will be used as a template to get a better grip on analogous debates in social and personality psychology, and also to develop my account of extended virtues in Part 2.

At first pass, the spectrum looks like this: at the far conservative end we have the kind of staunch internalism defended by Addams & Aizawa (2008) which (roughly) holds that the kind of thing cognition is makes it in principle impossible that it could take place beyond the boundaries of the skull. At the far liberal end we have a highly permissive externalism defended by Gallagher (2013) which (roughly) holds that cognitive processing can extend to include social institutions like the legal system. These positions are subject to many of the criticisms raised above (e.g., question-begging and cognitive bloat, respectively) and rather than re-rehearse those objections and possible replies, I will instead focus my attention on the moderate claims occupying centrist positions on the spectrum. On the conservative centrist side is Rupert's (2000) hypothesis of embedded cognition. In a more centrist position is Sterelny's (2010) hypothesis of

scaffolded cognition. And on the liberal centrist side is Palermos's account of extended cognition. In what follows, I will explicate each of these positions in more detail.

II.III.I. The embedded mind hypothesis

The embedded mind hypothesis has been most thoroughly and sympathetically articulated by Rob Rupert (2004, 2009). In slogan form: "According to the hypothesis of embedded cognition, cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place" (2004, p. 393). A bit more substantively, Rupert (2009) writes:

According to this embedded view, typical cognitive processes depend, in surprising and complex ways, on the organism's use of external resources, but cognition does not literally extend into the environment. The embedded view is radical in its own way, for it recommends a decidedly different approach to the study of human cognition: do not look so much at the internal processes of computation or association but instead at the way the structure of the subject's environment facilitates the success of whatever internal cognitive processes occur (p. 5).

A defining feature of the embedded view is the language of *dependence* (as opposed to something like *constitution*). For Rupert, we can claim that cognitive processes *depend* on environmental resources, perhaps even very heavily, without claiming that such environmental resources are themselves constitutive of cognition. A central theme throughout Rupert's (2009) *Cognitive Systems and the Extended Mind* is that most (if not all) of what is attractive about the extended mind can be captured from within the

framework of *embedding* (again, as opposed to extending). In other words, Rupert concedes that external entities might be more important for cognitive scientific explanations than we would have thought, but this does not entail a commitment to the more radical claim that various props, devices, and structures are *constitutive* of cognitive processes. If the framework of embedding can account for the same range of cases as extension without the tricky inference from coupling to constitution, the argument goes, then we should, *ceteris paribus*, endorse the hypothesis of embedded cognition over the hypothesis of extended cognition by “the methodological principle of conservatism” (Rupert 2004, p. 395).

Given this brief sketch of some commitments of the embedded view, what does it actually look like in practice? Rupert uses a familiar academic example:

In the main text, the reader sees ‘(Meyer, Thomson-Forshay, Gabowitz, Taylor, and Repucci 2004),’ which refers to a work in which the reader is particularly interested. Rather than committing five surnames and a date to short-term memory, the reader simply remembers “Meyer plus a bunch”; given contingent facts about the academic milieu, this information almost certainly corresponds to only one entry (2009, p. 180).

Anyone reading the present text has probably had this experience. And for Rupert, it nicely illustrates three features of the embedded view of mind and cognition: (1) there is no elaborate, detailed representation, (2) what is represented is rather coarse-grained, (3) only the needed information is collected (Rupert, 2009, pp. 180-3). The crucial point here is that *all* of this can be captured without making the claim that something like working memory extends to include the book. Rupert then goes on to show how even these cases -

where environmental information is exploited - can be explained in the orthodox cognitive scientific framework of internalism, representation, and computation - warts and all.

There is much to like about Rupert's approach. His treatment is charitable and fair minded, and his conclusions careful, clear and humbly stated. One takeaway for present purposes is the very real worry that by arguing for something like the extended mind, we are bringing along some clunky theoretical baggage, or perhaps unnecessarily departing from orthodoxy without an obvious benefit of payoff. That is, one has to take seriously the claim that if all of the phenomena of interest can be explained from a simpler, less controversial, and more intuitive framework, then all other things being equal, we ought not needlessly complicate the matter. Put positively, I take Rupert's challenge to be the following: if we want to posit the more radical framework of extending (as opposed to embedding), then we ought to be able to show, in a non-question-begging way, how it makes visible important and interesting phenomena that would otherwise be invisible.

II.III.II. The scaffolded mind hypothesis

The hypothesis of the scaffolded mind is due to Kim Sterelny. In slogan form: "The scaffolded mind hypothesis proposes that human cognitive capacities both depend on and have been transformed by environmental resources. Often these resources have been preserved, built, or modified precisely because they enhance cognitive capacity" (Sterelny, 2010, p. 472). Sterelny formulates this view of cognition as a natural outgrowth of the *niche construction* models in evolutionary biology. Developed by John Odling-Smee and colleagues, (e.g. Odling-Smee, Laland, & Feldman, 2003) these models

emphasize the active role of the agent in explaining the adaptive fit of agent and environment. Over time, agents (or, more exactly, lineages of agents) do indeed adapt to their environments. But they also adapt their environment to them. In classic examples of niche construction, these environmental modifications are directly utilitarian physical interventions. Animals construct nests, burrows, and dams, thus protecting themselves from predators and the violence of the world. Trees alter soils with their leaf litter, discouraging competitors. Earthworms modify the structure and moisture of the soils in which they live, easing the threat desiccation (Sterelny, 2010, p. 470).

Sterelny's insight is to extend, as it were, this way of thinking to the cognitive domain. In much the same way as earthworms modify the soil moisture to protect themselves, so too, do humans modify the contents of their cognitive environments to bolster their intellectual capabilities. Everyday examples of memory scaffolding are ubiquitous. Rather than trying to remember to mail a letter first thing in the morning, I simply place the envelope on top of my keys, so I cannot leave the house without the letter. Rather than memorizing all of my appointments, I set up reminders on my smartphone calendar.⁹ Perhaps the best example of scaffolding is also the folksiest and pithiest: "marking a trail transforms a difficult navigational problem into an easy perceptual problem" (Sterelny, 2006, p. 218). The hypothesis of scaffolded cognition thus calls attention to the many ways in which the burden of a cognitive task can be lessened by actively offloading it to the environment.

⁹ From the perspective of niche construction, however, these one-off cases are of less interest than the kind of cumulative, population-level effects which can shape evolutionary trajectories. This is what proponents call *ecological inheritance*. When earthworms modify the soil, or beavers build dams, these modifications to the environment re-shape the developmental niche of the next generation.

In the previous section, I noted that the hypothesis of the embedded mind is couched in the language of *dependence*. This is largely a passive construct. Returning to Rupert's reference page example, the reader does not play much of an active role in creating the conditions that enable the search heuristic to work properly. Rather, there are contingent but reliable structural features of academic texts (e.g. reference page at the end of the book, low probability of first author and year appearing with more than one entry, etc.) which can be exploited, but the reader does not do much in the way of actively creating those features. What is distinctive about the hypothesis of scaffolded cognition is its emphasis on the ways in which agents can *actively sculpt* the kinds of environmental features that enhance, modify, and bolster - that is, scaffold - cognitive competences. Where the environment is (largely) exogenously determined in the framework of embedding, in the framework of scaffolding, the environment is (relatively more) endogenously determined. Put another way, where the embedded mind politely uses reliable, pre-existing environmental resources, the scaffolded mind assertively builds new and better ones.

Though the hypotheses of embedding and scaffolding do diverge with respect to the role of the agent in sculpting the environment, they do converge (or at least do not obviously diverge) in a kind of methodological conservatism. By Sterelny's lights, putative cases of extended cognition can be more insightfully understood in the relatively more established framework of niche construction. Echoing Rupert's argumentative strategy, Sterelny writes:

Many of the examples that motivate the extended mind picture can be seen as special cases of niche construction, cases in which human competences depend

intimately on the environment being scaffolded to support adaptive decision-making... the canonical extended mind cases are continuous with other cases, cases in which there is environmental support of cognition, but which are not plausibly treated as constituents of agents' minds (2010, p. 466).

Again, the claim here is that a more established framework can account for what is supposedly novel about the extended mind, and can do so without the slippery inference to constitution. In appealing to the niche construction research program, Sterelny also believes that his view is part and parcel of a more general explanatory framework than can be offered by Clark & Chalmers-style active externalism. The scaffolded mind hypothesis thus occupies a centrist position on the spectrum because it can incorporate the insights of the embedded perspective (e.g. recognizing how humans exploit reliable environmental features), without unnecessarily claiming such features are constitutive. However, it also gives voice to the more deliberate ways in which humans non-trivially modify their environments to enhance existing capabilities and make possible new ones. Importantly, the scaffolded mind hypothesis does all of this while still acknowledging the possibility of genuinely extended cognitive processes, even if they are not as widespread as some proponents might have us think.

Rupert's challenge was to justify the utility of the more controversial and less intuitive theoretical apparatus needed to get extended mind arguments off the ground. This is well-taken. For Sterelny's scaffolded mind hypothesis, the challenge is to account for the ways in which organisms are not passive receptacles of environmental influence. So one worry from the perspective of the scaffolded mind is that embedded views do not do enough to foreclose on a view of humans as bowling pins, inertly awaiting incoming

stimuli (recalling the stimulus-response model criticized by Dewey. Or, to foreshadow a bit, vases waiting to be dropped on the floor). The scaffolded mind hypothesis forces us to take seriously the toddlers we all once were (and in many ways still are), incessantly probing the environment and figuring out clever ways to work within it.

II.III.III. The extended cognition hypothesis

The hypothesis of extended cognition goes further still. Not only do we exploit reliable features of the environment, and not only do we actively sculpt our environment to be more amenable, but we do these things so persistently and profoundly, the argument goes, that the environmental resources become integrated as proper constitutive parts of us.

With various aspects of the hypotheses of embedded and scaffolded minds elucidated above with reference to various aspects of the hypothesis of extended cognition, I will now elaborate a more refined account which has evolved from Clark & Chalmers's (1998) coupled systems formulation. In the first place, careful readers may have noticed equivocations in the foregoing between e.g., theories of embedded/scaffolded/extended *mind* and theories of *cognition*. For my part, I have tried to refer more narrowly to embedded/scaffolded/extended *cognition* where possible, though other authors often slip between the two.

Pöyhönen (2014), Bernecker (2014) and others have pointed out that there are important differences between talking about mind and cognition. An emerging consensus in the literature is that the latter is “a claim about cognitive processing,” where the former “concerns the location of human mental states,” and that, crucially, these distinct hypotheses “apply to different domains, address largely different issues, and depend on

different sources of evidence” (Pöyhönen 2014, p. 737). One implication of this distinction is that claims concerning embedded/scaffolded/extended *cognition* are evaluated with respect to the explanatory standards of the cognitive sciences, while claims concerning the mind are evaluated with respect to our folk psychological intuitions about mentality and mindedness. While the distinction may seem pedantic, it is still worthwhile to note, because the most careful and nuanced accounts have been developed in the context of extended *cognition*.

One such account is offered by Palermos (2014), who explicitly eschews talk of the extended mind (which, on his view, involves the extension of mental states like beliefs and desires) in favor of talk of cognitive processing (things like remembering and problem solving). The most important feature of Palermos’s account, though, is its direct and convincing response to what I identified above as the two most formidable challenges to formulations of extended cognition in the coupled systems framework: the coupling-constitution fallacy and the cognitive bloat charge. On Palermos’s telling, both are “interrelated worries whose common point—and the problem they accentuate—is the lack of a principled criterion of constitution” (p. 26). If these criteria can be satisfactorily elucidated, they have the potential to not only silence the coupling-constitution and cognitive bloat objections, but also demonstrate the need for an extended cognition framework above and beyond what is posited by the frameworks of embedding and scaffolding.

Palermos borrows from the terminology of Dynamical Systems Theory to make an argument for the ontological postulation of the kinds of tightly coupled systems which

underwrite arguments for extended cognition. Though decidedly less slogan-able than the previous hypotheses, here is how Palermos characterizes such systems:

(I)n cases where two nonautonomous systems mutually interact on the basis of feedback loops, there is an ongoing causal amalgam between the two units that disallows their decomposition into two separate systems on the basis of distinct inputs and outputs... The reason is that the way each component is affected is not exogenous to the component itself, and so cannot be properly thought of as its input. Likewise, the way each component affects the other is directly and synchronically related to the component to be affected and so cannot be properly conceptualized as output of the affecting component... We can call this the ‘ongoing feedback loops’ argument for the (ontological) postulation of coupled systems... In other words, ongoing mutual interdependence on the basis of feedback loops is the criterion by which we can judge whether two seemingly distinct systems constitute an overall system, consisting of both of them (pp. 33-4).

The idea is that two systems can be so interdependent that it makes less sense to talk about them as two interacting systems than as one, higher-order system. When component *X* affects component *Y* in roughly the same way as component *Y* affects component *X*, then there are grounds for postulating that *X* and *Y* constitute an extended coupled system. I have argued elsewhere that this “ongoing feedback loops criterion” is best understood as “inter-action in a strict sense, characterized by symmetry, balance, and reciprocity of the pathways of influence between components of the coupled system” (Skorburg, 2017, p. 11).

At first pass, here is how I cash out those concepts. Recalling a terrible example from Adams & Aizawa (2010, p. 67) the mathematician-and-pencil coupled system exhibits *asymmetry* because the pencil could never influence the mathematician to the degree that the mathematician could influence the pencil. In this sense, all asymmetric coupled systems also exhibit *imbalance* in the pathways of influence between components and are therefore also not *reciprocal*. An example of a *symmetric* but *imbalanced* coupled system might be illustrated by something like Twitter trends. In principle, all Twitter users exert some influence on what is trending, and so the pathways of influence are (or are at least potentially) *symmetric*. But what Katy Perry tweets (with some 103 million followers as of this writing) has significantly more influence on Twitter trends than what I tweet with my few dozen followers. Thus, there is an *imbalance* in the pathways of influence between components, and so no *reciprocity*. Lastly, *symmetric*, *balanced*, and therefore, *reciprocal* coupled systems are perhaps realized by the Tactile Visual Substitution Systems described by Palermos (2014). (Skorburg, 2017, pp. 18-9, n. 5).

In short, Tactile Visual Substitution Systems (TVSS) are devices which transpose incoming sensory information from one sensory modality to another.¹⁰ With TVSS, images from video cameras (usually implanted in glasses worn by persons with visual impairments) are transposed to tactile ‘images’ (usually via electrode arrays placed on the tongue or skin). These systems allow subjects to navigate their environments, locate and grasp small objects, and even read off body language of interlocutors. According to Palermos, in such systems:

¹⁰ For an accessible review, see Bach y Rita & Kercel (2003).

the mutual interaction between the agent and his tactile visual substitution system gives rise to new systemic properties (such as the new quasi-visual experiences produced, or new possibilities for interaction with the environment) that do not belong to any of the subsystems alone, but to the overall coupled system, ATVSS [“Agent-Tactile Visual Substitution System”]...the postulation of coupled systems is necessary with respect to the explanation of certain systemic properties, which we would otherwise be at a loss how to account for (p. 32).

The crux of the argument is that the frameworks of embedding and scaffolding do not do justice to the agent’s transformed multi-sensory experience of the environment. The substitution system allows the agent to perceive the environment in ways which open up new modes of interaction, which in turn afford new perceptual experiences, which subsequently open different modes of interaction, and so on. Palermos’s point is that the feedback and feedforward loops are so tightly integrated on spatial and temporal scales that they cannot be accurately decomposed into perceptual inputs and behavioral outputs.

Extended cognition so-construed can be seen as a direct response to Rupert’s challenge: what is the payoff for positing metaphysically complicated coupled systems? It makes visible higher-order systemic properties that, without such a framework, would remain invisible.¹¹ The relationship to Sterelny’s account of scaffolding is a bit more complex. On the one hand, the kind of extension represented by TVSS might be seen as an unusually specialized instance of scaffolding - an outlier on the distribution. But on

¹¹ One might also think that Palermos’s account is also more phenomenologically faithful to the experience of sensory substitution. In the same way Roger Federer, in the midst of a rally, probably does not perceive his racket as distinct from his body, it seems likely that the machinery of the substitution systems themselves would be transparent to familiar users. At any rate, it should remain an open question whether TVSS-type cases are closer to exception or rule, and whether they can be used to justify wholesale shifts in cognitive scientific practice.

the other hand, from the perspective of Palermos-style extended cognition, the framework of scaffolding might not go far enough in its recognition of the transformative potential of the kinds of bi-directional, symmetric feedback loops between agents and substitution systems.

The precise nature of the relationship between scaffolded views and extended views need not be settled here, however. The point of this section is not to establish the once-and-for-all superiority of one framework over the others. Rather, the aim is to survey the relevant features of each framework in order to construct a template that can be used to examine the varieties of dependence relations, not just between mind and world, but also between person and situation (Chapter III), and ultimately, virtue and context (Chapter VI). The foregoing survey of positions in the extended mind debate has now provided enough information to construct such a template, and that is the task of the next section.

II.IV. Template

This section serves as a summary of the positions that have been staked out within the coupled systems dimension of the extended mind and cognition debates. The aim here is to abstract away from the particulars of those debates in order to produce a more general framework suitable for eventually constructing my account of extended virtues. Isolating the general features of the embedding, scaffolding, and extending relationships here will - or so I will argue - help to illuminate structurally similar debates in social and personality psychology. This will, in turn, pave the way for a nuanced and fine-grained discussion of the ways in which virtues might also be embedded, scaffolded, or extended.

II.IV.I. Embedding

A revised and generalized hypothesis of embedding looks like the following: “typical processes depend, in surprising and complex ways, on the organism’s use of external resources, but those processes do not literally extend into the environment” (adapted from Rupert 2009, p. 5). The relevant descriptors of this view were *dependence* (as opposed to constitution) and *passivity* (as opposed to active construction). The framework of embedding also maintains (or at least does not dissolve) a rigid distinction between the internal and external, organism and environment. The effects of embedding are not particularly transformative, and have minimal downstream influence. Formally, we can say that *X* is embedded in *Y* when *X* depends on *Y*, but the pathways of influence between *X* and *Y* are mostly asymmetric and imbalanced. That is, the shape of the relationship between *X* and *Y* is such that the pathways from *Y* to *X* are more voluminous than the pathways from *X* to *Y* (this is the asymmetry condition). Further, because the shape of the relationship is asymmetric, the degree of influence is also *imbalanced*. *Y* influences *X* much more reliably, and to a much greater degree than *X* is able to influence *Y*.

Informally, we can say that searching for an entry on the reference page demonstrates an *embedded* cognitive process because the reader *depends* on reliable and entrenched academic conventions which make the search heuristic possible. But the reader stands in an *asymmetric* relationship to these conventions because the conventions are the result of decades of precedent solidified by thousands of scholars. Whatever influence any single reader may have is dwarfed by the inertia of the conventions. And because the structure of the relationship is skewed in this way, the influence of the reader on the conventions cannot but be *imbalanced* as well.

II.IV.II. Scaffolding

A revised and generalized hypothesis of scaffolding looks like the following: “The scaffolded mind hypothesis proposes that organisms’ capacities both depend on and have been transformed by environmental resources. Often these resources have been preserved, built, or modified precisely because they enhance organisms’ capacities” (adapted from Sterelny, 2010, p. 472). The relevant descriptors here were *construction* (as opposed to dependence) and *active sculpting* (as opposed to passive reception). The framework of scaffolding softens the distinction between internal and external, organism and environment. One of its defining features is the possibility of downstream transformative effects (ecological inheritance). Formally, we can say that *Y scaffolds X* when *Y* supports and makes *X* possible, and the pathways of influence between *X* and *Y* are mostly symmetric, but imbalanced. Symmetric, because the shape of the pathways between *X* and *Y* admit of bi-directional influence between them. But in most cases of scaffolding, the degree of influence is still *imbalanced*. While *X* can exert non-trivial and sustained influence on *Y*, it is still the case that *Y* exerts more influence on *X* than *X* exerts on *Y*.

Informally, we can say that my smartphone calendar *scaffolds* my memory because it makes possible more detailed and longer-term scheduling than would be possible if I tried to do it all “in my head.” I can obviously influence my calendar, as its contents are largely (though not entirely) the result of my inputs. And my calendar certainly influences my behavior, dictating where I need to go, which deadlines are approaching, etc. In this way, the relationship is roughly *symmetric*, in that there is non-trivial, bi-directional influence between the calendar and me. But ultimately, the

influence is *imbalanced*, as I am still largely responsible for populating the contents of the calendar.

II.IV.III. Extending

A revised and generalized hypothesis of extending looks like the following: The hypothesis of extending holds that components of a coupled system can be bound together so tightly - on the basis of ongoing feedback loops - that it only makes sense to posit a new, extended system. (adapted from Palermos, 2014, pp. 33-4) The relevant descriptors were *constitution* (as opposed to dependence or support) and *reciprocal interaction* (as opposed to one-sided sculpting). The framework of extending thus collapses the sharp distinction between internal and external, organism and environment. Its defining features are transformative effects, not only downstream, but also in real time. Formally, then, we can say that *X* extends to include *Y* when *Y* is constitutive of some part of *X*, because the pathways of influence between *X* and *Y* are mostly symmetric and balanced. The influence pathways are symmetric, because the shape of the pathways between *X* and *Y* admits of bi-directional influence between them. And in most cases of extending, the degree of influence is distinctively *balanced*. *X* and *Y* exert bi-directional and reciprocal influence on one another: *X* influences *Y* in the same way that *Y* influences *X*.

Informally, we can say that the TVSS *extends* its wearer's cognitive-perceptual apparatus because it *reciprocally* opens new possibilities for engagement with the world. The relationship between TVSS and its wearer is *symmetric* because information flows both from the TVSS to the wearer, and from the wearer to the TVSS. Crucially, the degree of influence is mostly *balanced*: The wearer determines the inputs to the TVSS by

orienting to the world in certain way, and the TVSS - fluidly and in real-time - further informs the wearer's orientation to the world, which in turn changes the inputs to the TVSS, and so on.

So much for the template. I take it that the ability to distill the relevant features from the extended mind/cognition debates into this template are indicative of at least some aspects of those debates yielding mature and refined formulations of the various dependence relations between mind and world. In the next chapter, I will examine the extent to which these same dependence relations are apt to describe the person-situation relationship in psychology. If my arguments go through, we will find evidence that in roughly the same way cognition can be embedded in, scaffolded by, or extended to include external resources, so too, might features of personality be embedded in, scaffolded by, and perhaps even extended to include situational resources. This converging evidence will in turn motivate my claim that components of virtues can be similarly understood as embedded in, scaffolded by, or extended to include context. By constructing a template in this section, I have tried to show how concepts from the extended mind/cognition debates can lay the groundwork to start making good on these claims.

With this summary and template now in place, I will conclude the chapter by briefly surveying some recent developments in the literature which have taken the arguments beyond the scope of mind and cognition and into new domains, many of which are relevant to my account of extended virtues.

II.V. Recent developments

Since the publication of “The Extended Mind” in 1998, philosophers of mind sympathetic to the cause have asked what else - besides the dispositional beliefs in Otto’s notebook - might be similarly extended? I will briefly survey some responses to this question, but I will focus especially on the recent developments which bear on aspects of virtue theory, as those will be integral to my account of extended virtues in Part II.

Perhaps the most controversial suggestion in the literature is that consciousness itself is extended. In this vein Noë (2009) claims “the world itself can be described as belonging to the very machinery of our own consciousness” (p. 65).¹² Loughlin (2013) uses the example of an artist and her sketchpad to argue that the hypothesis of extended cognition entails the hypothesis of extended consciousness, and Zawidzki (2012), drawing on trans-humanist science fiction, makes a phenomenological argument for the possibility of extended consciousness. Clark (2009), however, is quite clear that “nothing in the arguments for EM [Extended Mind] should incline us to accept ECM [Extended Conscious Mind] (p. 968). Wheeler (2015) is similarly pessimistic about an inference from Otto’s notebook to extended consciousness: “if the externalist goal in this arena is to establish that the physical machinery underpinning the what-it’s-like-ness of experience is extended beyond the skin, then different arguments will be needed. The internalist is certainly not home and dry, but the externalist has significant work to do” (p. 173).

In a similar vein, though less well developed, is Froese & Fuchs’s (2012) argument that a concept of the extended *body* is needed to give a robust account of social cognition and embodied subjectivity. DiPaolo’s (2009) “Extended Life” similarly takes up extended mind-style arguments, but in the context of enactivist arguments for the

¹² See Wheeler (2015, pp. 165 ff.) for a helpful clarification of Noë’s argument.

continuity between life and mind, thereby countering the claim that enactivism is a form of internalism incompatible with active externalism.

One of the more substantial spin-offs, however, has been the development of a significant body of literature around the epistemological implications of the extended mind. Duncan Pritchard and colleagues (e.g. Pritchard et al. (forthcoming)) are at the fore of this movement. Much of the work in this vein uses extended mind-style arguments to generate cases of extended *knowledge*, which serve to advance the debates beyond more traditional forms of epistemic externalism, towards a more active externalism in social epistemology and virtue epistemology. In hindsight, this is not a particularly surprising development. If it turns out that Otto's beliefs are extended to include his notebook, this has obvious consequences for all sorts of reliabilist epistemologies which focus on the nature of belief-formation processes. As Palermos & Pritchard (2013) point out:

we can claim that our knowledge-conducive cognitive characters may, and indeed many times do, extend beyond our organismic cognitive faculties... Even though the belief-forming process in virtue of which the subject formed his true belief is for the most part external to his organismic cognitive agency, it still counts as one of his cognitive abilities as it has been appropriately integrated within his cognitive character (p. 113).

Ergo, knowledge ain't (all) in the head! Again, I take these developments as evidence of the maturation of the extended mind and cognition research program. At some point, the arguments shift from proof of concept, to applications, implications, and new directions. For present purposes, though, the most important development out of the extended mind literature has been the formulation of a variety of accounts of extended affectivity and

emotion. The thrust of these arguments is nicely summarized by Colombetti & Roberts (2015):

For proponents of the ExM [Extended Mind] approach to cognition, we submit, this should not be an entirely unexpected result. If, as authors working under this banner have suggested, we must construe cognitive capacities as densely and inextricably embedded within a wider cultural, technological, and social context in which thinkers deploy environmental opportunities to shape and expand their powers for thinking, then it is natural to anticipate that the same may hold true for affective phenomena. Only if one were to endorse a very strict conceptual separation between cognition and emotion, would the claim that one but not the other might be realized within an extended system be a sustainable theoretical position. If we are correct in our analysis, proponents of ExM ought to accept that—just as in the case of cognition—an exhaustive description of the nature of human affectivity, philosophical or empirical, cannot be achieved unless attention is directed outward from the merely neural, and indeed beyond the spatial limits of the living organism (p. 1260-61).¹³

Again, in hindsight, it is not surprising that these sorts of claims would be developed. As Carter, Gordon, & Palermos (2015) aptly demonstrate, if cognitivist theories of emotion (e.g. Nussbaum, 2001) are plausible on independent grounds, then it is not at all implausible to think - following Clark and many others mentioned in this chapter - that if

¹³ See also Stephan, Walter, & Wilutzky (2014): “Situated approaches to cognition are motivated by the insight that we are not isolated quasi-Cartesian minds housed in a container-like body that intelligently navigate through their environment by repeated sense-think-act cycles. This insight applies *mutatis mutandis* to emotions. It is not just that we aren’t isolated thinkers; we aren’t isolated emoters either...If anything, the case for an intimate coupling of brain, body, and environment should have been even more obvious for emotions from the very start. If thinking about cognition had not made us wonder about situatedness, emotions should have” (pp. 66-67).

cognitive processes are sometimes extended, perhaps the relevant processes (e.g. judgments, appraisals) countenanced by the cognitivist about emotions might also be so extended.¹⁴

If cognitivism about the emotions or constitution claims á la Clark are too much to stomach, Colombetti & Krueger (2015) offer a dialed down version of *scaffolded affectivity* which maps directly onto Sterelny's account of scaffolded cognition. Similarly, Griffiths & Scarantino (2005) might be construed as developing the analogously *embedded* account of emotions. Without venturing too far into the intricacies of the philosophy of emotion here, the point is simply to show that there is a comparably rich body of literature which also marks out a number of nuanced distinctions between the poles of internal and externalism about emotions.

Before concluding, it is also worthwhile to note that many extended emotion theorists have begun to hint at the potential ethical implications of their view. Taking stock of these proposals here will help to clear the conceptual space needed for my account of extended virtues in Part II.

In the first place, Colombetti & Roberts (2015) argue that the case for extended cognition (usually centered on beliefs, memories, planning, problem solving, etc.) entails a similar active externalism in the domain of affectivity, including mood, sentiment, temperament, and emotion. We are asked to imagine Eve, who dispositionally resents her parents. In her diary, she tracks the many ways her parents make her angry: lack of warmth, constant criticism, etc. As she ages, she grows out of her resentment and does

¹⁴ This is not to endorse cognitivism about the emotions (which I do not), but only to show how extended cognition arguments can be transposed rather easily to debates in the philosophy of emotion.

not regularly, consciously exhibit resentment toward her parents. Except, that is, when she reads and re-reads her diary. Colombetti & Roberts thus write:

Without the diary, we can imagine, she would not rekindle her resentment that often, in fact she may even be able to forget her negative relationship with her parents, and cultivate more positive memories and feelings toward them. Without the diary, that is, she is not (or need not be) so disposed to manifest parent-directed resentment. This scenario strikes us as plausible, and one in which, if one endorses ExM [Extended Mind], one also ought to say that Eve's diary is part of the supervenience base of the system that realizes her standing, dispositional resentment toward her parents. (2015, p. 1253).

After making similar arguments for the various ways in which wider ranging, multi-track affective states (like moods and temperaments) can be similarly extended, Colombetti & Roberts go on, somewhat speculatively, to claim that:

The fact that some extra-neural artefacts are deployed for their transformative effects only within a certain domain—workplace, home, public transport—*is no barrier to their being true extensions of a person's character*. Moreover, environmental components may lend a degree of stability to a person's dispositions, enabling her to exhibit responses with a greater degree of uniformity across different social contexts than would otherwise be possible (p. 1255).

On their telling, introversion can be extended to include, for example, headphones one always wears in public, so as to effectively discourage small talk. I simply want to flag that it is a short leap from here to the normatively evaluable traits of character (e.g., (un)friendliness) that comprise the currency of virtue ethics. Though Colombetti &

Roberts do not make such a leap here, they pave the way for me to do so in Part II.¹⁵ To foreshadow a bit, this new line of inquiry into extended emotions provides a central motivation for the present project: If there is reason to believe that cognition is sometimes extended, and if there is reason to believe that affectivity is extended, then there is reason to wonder whether *virtues* - characteristic dispositions of thought and feeling - are similarly extended.

II.VI. Conclusion

In this chapter, I aimed to tell a story about externalism in the philosophy of mind which is rooted in early 20th century American pragmatism. From there, the relevant strand runs through contemporary debates in the philosophy of mind about the various dependence relations between agents and resources in their environment. Within these debates, a set of nuanced distinctions have been developed which are apt - or so I will attempt to show - for thinking more generally about the ways that agents depend on context. Equipped with these distinctions, I will attempt in the remainder of Part I to show how they can be deployed in the service of an argument that we are spread out into the world beyond the intuitive boundaries of skin and skull. Part II will then wrestle with the ethical implications of such a view.

¹⁵ Another promising route in this vein is offered by Kruger & Szanto (forthcoming), who discuss how dyads (e.g. parent-child, dance partners, best friends, romantic partners, etc.) can co-regulate each other's emotions and perhaps even create novel, emergent emotional states.

CHAPTER III

SOCIAL AND PERSONALITY PSYCHOLOGY

III.I Introduction

In the previous chapter, I traced the historical lineage of the extended mind from Dewey and the classical pragmatists up to the recent outgrowths of the research program, such as the arguments for extended affectivity advanced by Colombetti & Roberts (2015). Drawing on the more nuanced and mature formulations of the extended cognition, I developed a template for conceptualizing the various dependence relations between agents and environments. I concluded by suggesting that principled formulations of the scaffolded and extended cognition hypotheses reveal interesting and important features of cognitive and affective processing that might not be appreciated in more conservative internalist frameworks. Thus, when taken seriously, these frameworks have the potential to undermine default internalism about cognitive and affective processes.

My aim in the present chapter is to substantiate the arguments of the previous chapter by identifying connections with research in social and personality psychology. The upshot is that in the same way arguments in Chapter II are poised to challenge default internalism about cognitive and affective processes, the arguments in this chapter are poised to challenge default individualism about the bearers of such cognitive and affective processes. Taken together these pose a formidable challenge (or so I will argue in Part II) to foundational concepts in virtue ethics.

Here, then, is the roadmap for the present chapter. In Section II, I will trace the development of the person-situation debate from the history of social and personality psychology, showing how this debate is structurally similar to the history of the extended

mind and cognition debate surveyed in the previous chapter. In Section III, I argue that bringing these two research programs into dialogue is mutually beneficial. I then formulate and respond to potential objections to this comparison. Section IV introduces promising empirical research programs in personality psychology which corroborate and expand the frameworks of scaffolding and extending. Section V then concludes by drawing out the philosophical implications from these empirical research programs, namely, that the bearers of the kinds of states and processes described in the previous chapter might similarly extend beyond the bounds of skin and skull.

III.II The Person-Situation Debate

In Chapter II, Section II, I attempted to show how contemporary arguments for extended mind and cognition owe an intellectual debt to the classical American pragmatists and that a recognition of these historical roots may help to advance contemporary debates. In the present section, I will adopt a similar strategy with respect to the history of social and personality psychology.¹⁶

Many introductory courses and textbooks in social psychology begin with Lewin's (1936/2015) equation: $B = f(P,E)$, where "B" is behavior, "P" is the person and "E" is the environment, such that behavior is a function of the person and the environment. When written in this way, the comma indicates agnosticism on the precise nature of the relationship between Person and Environment. On one reading, the comma separates two independent entities, such that the effect of some aspect of personality on behavior is the same regardless of the environment or situation (I use the two interchangeably), and the effect of some feature of the situation on behavior is the same regardless of personality.

¹⁶ Much of what follows is reproduced and adapted from Skorburg (2017).

John Kihlstrom (2013) notes that this reading characterized traditional approaches in social and personality psychology, and also the resulting division in labor in psychology departments in the 20th century. The traditional approaches in personality psychology, Kihlstrom writes,

construed behavior as a function of person attributes such as traits, attitudes, emotions, motives, and values...in such research, the effects of the environment are generally construed as 'noise.' The canonical method of traditional personality psychology thus exemplifies the Doctrine of Traits, which may be stated as follows: Social behavior varies as a function of internal behavioral dispositions that render it coherent, stable, consistent, and predictable (p. 793).

Traditional approaches in social psychology, on the other hand,

construed behavior as a function of differences in the physical and (especially) social environment...in such research the effects of individual differences in personality are generally construed as 'noise.' This view is captured by what might be called the Doctrine of Situationism: Social behavior varies as a function of features of the external environment, particularly the social situation, that elicit behavior directly, or that communicate social expectations, demands, and incentives (p. 793).

The person-situation debate, then, is about how best to explain and predict behavior.

Proponents of the Doctrine of Traits would think that personality variables explain more of the variance in behavior, while proponents of the Doctrine of Situationism would think that environmental variables do more of the explanatory work. Already, we can see that

when the debate is cast in these terms, it is possible to read it in terms of (crude) internalism and externalism, respectively. I will return to this point below.

Much psychological research in the 20th century was conducted along these person-situation party lines, where the relationship between partisans was described as “benign collegial neglect” (Kihlstrom 2013, p. 794). All of that changed, however, with the publication of Walter Mischel’s seminal (1968) *Personality and Assessment*, which was received at the time as a devastating critique of personality psychology. In an oft-quoted passage Mischel claims:

With the possible exception of intelligence, highly generalized behavioral consistencies have not been demonstrated and the concept of personality traits as broad response predispositions is thus untenable...The initial assumptions of trait-state theory were logical, inherently plausible, and also consistent with common sense and intuitive impressions about personality. Their real limitation turned out to be empirical—they simply have not been supported adequately (pp. 146-147).

Another popular soundbite is that in the personality literature reviewed by Mischel, the correlation between scores on various personality measures and actual behavior was about $r = .3$, meaning that only 10% of the variance in behavior is explained by personality variables. Setting to the side methodological and interpretative qualms with Mischel’s review, *Personality and Assessment* did turn out to be rather devastating to personality psychology at the time. Swann & Seyle (2005, p. 156) show that after the publication of *Personality and Assessment*, there was “a marked decline in the number of research studies, graduate training programs, and dissertations devoted to personality psychology,” and this lasted well into the 1970’s.

Among others, work by David Funder and colleagues helped to reverse this trend. Most notably, Funder and Ozer (1983) showed that once the t and F test statistics often reported in social psychological and situationist studies are converted to the r statistic often reported in personality and individual difference studies:

the effects on behavior of several of the most prominent Situational factors in social psychology seem to average slightly less than .40. Moreover, because in most social psychological experiments only two or three levels of the situational independent variable are studied, and because these levels usually are deliberately chosen to be quite different from each other, situational linear effects such as calculated here quite possibly overestimate the true state of affairs (Funder & Ozer 1983, p. 110).

In other words, once apples are compared to apples, the predictive and explanatory power of the traditional social psychological and situational measures are not much different from personality and individual difference measures. Indeed, in a meta-analysis of more than 25,000 studies involving 8 million participants, Richard et al. (2003, p. 337) find that the mean effect size for person and situation effects are .19 and .22, respectively. It seems, then, that, situations are no better at explaining behavior than are persons.

Recall that all of the preceding discussion has been based on the assumption that the comma in Lewin's equation signified an additive, or independent relationship between persons and environments. As a matter of contingent historical fact, this is how social and personality psychology developed as disciplines, but it is not the only way to interpret the relationship between persons and environments. To cite just one of many possible contemporary examples, Mischel himself notes in a 40-year retrospective on

Personality and Assessment that the most promising approaches in social and personality psychology:

bridge the classic partitioning most unnatural and destructive to the building of a cumulative science of the individual—the one that splits the person apart from the situation, treating each as an independent cause of behavior. At the time of the ‘person versus situation’ debate, this splitting (amazingly) threatened to make personality the discipline that studies people apart from situations, and social psychology the unfriendly neighbor that studies situations apart from people.

(2009, p. 289).

In other words, according to the very researcher whose work helped fuel the division, the Doctrines of Traits and Situationism, taken independently, constitute a false dichotomy. The takeaway is that you can ask any contemporary social or personality psychologist about the person-situation debate and they will likely tell you that it is over, and has been for some time.

But what does any of this history have to do with the debates over extended cognition? I will try to show that there are important similarities between the two, and insofar as this comparison is on the right track, three points will follow. First, the resolution of the person-situation debate can guide us toward a resolution of the debate over extended cognition. Second, the template developed in the previous chapter can characterize relevant features of research in social and personality psychology. Third, and perhaps most importantly, the connection between these research programs will secure the plausibility of talking in an empirically tractable way about cognitive and affective processes and their bearers as extended.

III.III From Interactionism in Psychology to Philosophy

Why think an old debate from the history of psychology has any bearing on the more recent debates over extended cognition? I contend that we can tell a coherent narrative about the development of the person-situation debate and the extended cognition debate which begins in both cases with a dichotomy which is eventually eschewed in favor of an interactionist conclusion with a gradient of positions in between. The thread that ties these narratives together is the common interest in the nature of agents, environments, and their interactions.

In the history of the person-situation debate, the dichotomy is pretty straightforward: adherents to the Doctrine of Traits thought individual difference measures – personality traits that inhered in the agent - were the best way to predict and explain behavior. Adherents to the Doctrine of Situationism thought the same about environmental and situational influences external to the agent. While it is doubtful that few, if any, psychologists were actually staunch adherents of one doctrine while completely neglecting the other, Kihlstrom (2013) does make a convincing case on historical grounds that these doctrines did structure much early research in the development of what we now know as social *and* personality psychology. At any rate, the dichotomy is ultimately resolved in terms of interactionism. The quote above from Mischel illustrates this point, as does the following passage:

The persistence of this [person-situation] debate into the 21st century is something of a mystery. Since at least the 1930s, deep thinkers as diverse as Allport (1937) and Lewin (1951) have argued that invidious comparisons miss the point because behavior is a function of an interaction between the person and the

situation. By the 1980s this recognition had deteriorated into a truism. Nowadays, everybody is an interactionist (Funder 2006, p. 22).

In the extended mind and cognition debates, the relevant dichotomy is less stark, but still apparent. Clark & Chalmers never claimed that *all* cognitive processes or mental states were extended, nor did internalist critics ever claim that features external to the brain had *no* influence on cognition. But still, as Ross & Ladyman (2010, p. 155) point out, Adams & Aizawa's coupling-constitution fallacy isn't just pitched as a criticism of Clark & Chalmers's externalism – it is also a defense of internalism.

Much of what is at stake in these debates is the question of what sorts of structures and processes ought to figure into cognitive scientific theories, predictions, and explanations. In this sense, the early debate surrounding the extended mind and cognition did have a rather internalist *versus* externalist flavor. But as the debate evolved, a number of nuanced positions were staked out between the poles of internalism and externalism about the mind and cognition, and these were the subject of Chapter II, Section IV.

While the nuanced positions do represent some progress in the extended cognition debate, I contend that the person-situation debate can also be used as a guide for how the extended cognition debates might be resolved. Put more forcefully, if my comparison between the two traditions holds, then interactionist conceptions of extended cognition ought to be favored over less interactionist alternatives. I will cash out this claim in what follows.

Sutton (2010) identifies the first wave of extended mind arguments as committed to the parity principle,¹⁷ where the second wave is characterized by a commitment to the

¹⁷ Though I explicated Clark & Chalmers's formulation of the parity principle in Chapter II, here is Sutton's gloss: "first-wave EM [extended mind] is based on the parity principle: cognitive states and processes extend

complementarity principle.¹⁸ It is an open question what constitutes the third wave, but Shaun Gallagher’s aforementioned (2013) “Socially Extended Mind” seems a good candidate, as it pushes extended mind and cognition arguments in a new direction and has generated a number of responses in doing so.¹⁹ I agree with the letter of Gallagher’s proposal: the social domain is the new frontier for extended cognition. I disagree, however, with the operationalization of the socially extended mind in terms of “mental institutions” (e.g., Gallagher & Cristafi 2009). Gallagher’s proposal is that mental institutions like legal systems, academic research, and various cultural practices (Gallagher 2013, p. 6) offer the best prospects for the socially extended mind.

My proposal is that in roughly the same way that the person-situation debate is resolved in terms of interactionism, so too, ought the extended cognition debate be resolved in terms of interactionism. More specifically, we ought to give pride of place to proposals like Palermos’s (2014) ongoing feedback loops criterion, or Heersmink’s (2015) dimensions of reciprocal information flow and mutual transformation, which

beyond the brain and into the (external) world when the relevant parts of the world function in the same way as do unquestionably cognitive processes in the head” (p. 193).

¹⁸ Though it is discussed in more detail in Chapter II, here is how Sutton describes the complementarity principle: “second-wave EM is based on a complementarity principle: in extended cognitive systems, external states and processes need not mimic or replicate the formats, dynamics, or functions of inner states and processes. Rather, different components of the overall (enduring or temporary) system can play quite different roles and have different properties while coupling in collective and complementary contributions to flexible thinking and acting (Sutton 2010, p. 194).

¹⁹ The journal, *Cognitive Systems Research* dedicated a special edition (nos. 25-26) to Gallagher’s formulation of the socially extended mind. Gallagher’s argument is discussed in more detail in the previous chapter, but here is a brief formulation: “I have pursued a liberal interpretation of the extended mind, suggesting that we consider cognitive processes as constituted in various social practices that occur within social and cultural institutions. This idea of the socially extended mind builds on the enactive idea of social affordances. Just as a notebook or a hand-held piece of technology may be viewed as affording a way to enhance or extend our mental possibilities, so our encounters with others, especially in the context of various institutional procedures and social practices may offer structures that support and extend our cognitive abilities. (Gallagher, 2013, p. 4).

emphasize the symmetric, balanced, and reciprocal pathways of influence between components of coupled systems.²⁰

On these strong interactionist views, a number of problems arise with Gallagher's formulation of the socially extended mind. None of the examples Gallagher invokes meet the benchmarks for extension (as opposed to embedding or scaffolding): symmetry, balance, and reciprocity. It is hard to see, for example, how a paralegal assistant could exert the same kind of influence on decades of legal precedent as the precedents exert on them. That is, the pathways of influence between the paralegal and the law are probably best described as asymmetric. Perhaps we could charitably grant that the relationship is symmetric and imbalanced, but in any case, we are nowhere near reciprocity – one of the hallmarks of interactionism. And things do not look much better for Gallagher's even more nebulous examples of research practices and cultural practices.

Insofar as we are interested in resolving debates about extended cognition, then adopting something like Palermos's ongoing feedback loops criterion may come at the cost of relegating some putative cases of extended cognition to the framework of embedding. I agree, then, with Huebner's (2013) assessment that Gallagher's formulation of the socially extended mind is more accurately rendered as the socially embedded mind in just this way. This is not to say that the socially extended mind isn't important in its own right, but rather, that it does not offer a plausible path forward for *extended* (again, as opposed to embedded or scaffolded) cognition research. Insofar as something like the interactionism represented by the ongoing feedback loops criterion sets the bar, then Gallagher's socially extended mind will fall short.

²⁰ These views are discussed in more detail in Chapter II, Section III.

The foregoing illustrates how the debates over extended cognition can be informed by the resolution of the person-situation debate. If extended cognition theorists adopt the kind of interactionism used to resolve the person-situation debate, then this can help to distinguish among various proposals in the literature. But the influence also runs in the other direction: The concepts from the extended mind and cognition literature can be brought to bear on research in social and personality psychology. If mutual influence between traditions can be established, then I take this as a strong argument in favor of the connection between the two research programs.

Recall Sterelny's account of scaffolding: "the scaffolded mind hypothesis proposes that human cognitive capacities both depend on and have been transformed by environmental resources. Often these resources have been preserved, built or modified precisely because they enhance cognitive capacity" (Sterelny 2010, p. 472). In the framework of traditional internalism or even Rupert-style embedding, the environment is exogenously determined. Cognitive agents depend on and learn to exploit external features of their environment, but they may not modify those environments in important or systematic ways. Sterelny-style scaffolding, in contrast, construes the external environment as at least sometimes endogenously determined: it is a partially built environment, and it is often engineered to enhance or scaffold our cognitive capacities. This view cuts across a simple distinction between internal and external resources in the sense that an agent can actively sculpt the environment to determine how and what sorts of external entities she can interact with and come to depend on. The reliable presence of these external entities in turn shapes the internal capacities of the agent in new ways, which changes the sorts of external entities she interacts with, and so on.

There are a number of analogues in social and personality psychology which similarly served to undermine a strict dichotomy between the internalism of personality traits and the externalism of situational influences. To take just one, Ickes, Snyder, & Garcia (1997) review the vast literature in personality psychology which has studied the ways in which the active selection of situations can be understood as a function of personality. They conclude:

That people actively gravitate toward some types of situations and deliberately avoid others, and that their choices of situations are reflections of features of their personal identities (including such dispositions as attitudes, traits, and conceptions of self), may constitute major sources of the regularities, stabilities, and consistencies in behavior that are typically regarded as defining characteristics of personality (p. 186).

A significant component of personality, in other words, is the disposition to seek out certain kinds of situations (perhaps, situations that are not discrepant with one's self-concept) and avoid others (perhaps, situations that are discrepant with one's self concept). Selecting situations in this manner solidifies certain features of personality, which in turn leads to seeking out and avoiding certain kinds of situations, and so on. In much the same way Sterelny talks about cognitive scaffolding, a robust literature in personality psychology suggests that people can and do actively sculpt the situations they will encounter, and in so doing, generate the sorts of consistent behavioral, cognitive, and affective dispositions which characterize personality.

A similar case can be made with respect to Palermos-style extension and research on close relationships, but this will be the subject of the next section. The present analogy

between scaffolding and situation selection is meant to show that the relationship between extended cognition and social and personality psychology is a mutually beneficial one: the resolution of the person-situation debate offers clues about the resolution of the extended cognition debates and the scaffolded mind hypothesis provides a larger theoretical framework in which to situate the work on situation selection.

Still, it might be objected that an old debate about the explanation and prediction of behavior is orthogonal to a contemporary debate about the nature of cognitive and affective processes. Relatedly, one might also object that the sort of internalism and externalism represented by the Doctrines of Traits and Situationism is disanalogous with the debates about extended cognition - that the relation of the person to the situation is different in kind from the relationship of, say, Otto to his notebook. If these objections are on the right track, then the resolution of the person-situation debate would not tell us anything about the extended cognition debate, nor would there be any reason to think that former could help bolster the latter.

Regarding the first objection that the two debates are orthogonal, it should be pointed out that connections between social and personality psychology and the extended mind and cognition debates are already established. They have come largely in the context of the so-called situationist challenges to virtue theory.²¹ John Doris, one the pioneers of the situationist challenge to virtue ethics, briefly addresses the extended cognition debate in his (2015) book which seeks to sort out the implications of a wide range of psychological research on philosophical theories of agency, reflection, and

²¹ In a sentence: the evidence from situationist social psychology (e.g. Darley & Latane's (1968) bystander studies; Darley & Batson's (1973) Good Samaritan studies; Milgram's (1974) obedience studies, etc.) challenges the existence of the stable, consistent character traits countenanced by (neo)Aristotlean virtue ethics. See Doris (2002) and Alfano (2012) for detailed, book length treatments.

responsibility. In discussing the possibility of transactive memory in long term romantic partners (more on this below), Doris refers to the extended cognition debate to stake out his position. He ultimately thinks that transactive memory can be explained in the framework of Rupert-style embedding rather than Clark & Chalmers-style extension (Doris 2015, pp. 190-192). But the use of concepts from the extended cognition debate in interpreting social psychological research is, I think, indicative of an important connection between the two.

More substantively, I have argued elsewhere that extended mind-style arguments could offer a response to the situationist challenges to moral and intellectual character:

If those mental states are sometimes extended, perhaps the dispositions to have them are too. Presumably, the dispositions don't extend in every case, just as the states don't extend in every case. Perhaps some people are honest all on their own. Perhaps some people are intelligent all on their own. But if our suggestion is on the right track, in some cases, a person is honest because (among other things) she is suitably integrated with props, tools, or other people outwith her brain and body. Likewise, if our suggestion is on the right track, in some cases, a person is intelligent because (among other things) he is suitably integrated with props, tools, or other people outwith his brain and body (Alfano & Skorburg, p. 465).

We need not venture any further into the weeds of the situationist challenges to virtue theory here. The point here is just to emphasize the fact that a range of resources from the extended mind and cognition literatures have been brought to bear on aspects of the person-situation debate, and *vice versa*, and this tells against the objection that the two debates are orthogonal.

Regarding the second objection - that persons don't extend to situations in the same way that agents extend to their artifacts – I concede that this is, strictly speaking, true. But all I need to show is that the rhetorical moves in these debates are similar enough that one might profitably inform the other. To that end, it is worth noting how Lewin himself claims that the situation *includes* the person and is not separate from it. This is not obviously a constitution claim, but it is not so foreign to claims made by friends of extended cognition, either: “the psychological environment has to be regarded functionally as a part of one interdependent field, the life space, the other part of which is the person” (Lewin 1939/1951, p. 140, quoted in Kihlstrom 2013, p. 794). And Kihlstrom goes on to elucidate Lewin's view as follows: “*persons are part of the situation to which they respond* – or, put another way, that *persons and situations together constitute a unified field in which behavior takes place*” (pp. 795-96, emphasis in original).

For Lewin, at the level of behavior, there is no strict distinction between persons and situations. On this view, it then makes sense to say - recalling Clark - that behavior is the result of an inextricable tangle of feedback and feedforward loops, promiscuously criss-crossing the boundaries of persons and situations. Here again, we can see how the concepts developed in the extended mind literature can help to clarify the interactionist commitments of post-person-situation-debate psychology. And seen from the other side, Clark & Chalmers's claim that “there is nothing sacred about skull and skin” (1998, p. 14) looks like old Lewinian field theory wine in new functionalist bottles. In (roughly) the same way internal personality traits and external situational influences can conspire to produce behavior, so too can internal neural resources and external scaffolds conspire to produce cognition.

The interchangeability of the argument forms from the person-situation and extended cognition debates suggest that far from being disanalogous, the two can actually inform one another. Not only can the terms from the debates about extended cognition help to clarify the person-situation debate (e.g. Doctrine of Traits as internalism, Doctrine of Situationism as externalism, situation selection as scaffolding, etc.), but the terms from the resolution of the person-situation debate (e.g. “nowadays, everyone is an interactionist”), can help to resolve debates about extended cognition.

I take the foregoing to have established a meaningful connection between social and personality psychology and extended mind and cognition. In the next section, I will show how making this connection then paves the way for the next wave of extended cognition research.

III.IV The Next Wave

I suggested above that an interactionist criterion drawn from the person-situation debate helped to show why something along the lines of Gallagher’s (2013) socially extended mind would be a dead end for extended cognition, but perhaps a better fit in the framework of embedding. Here, I will introduce research programs in social and personality psychology which are better suited as standard bearers for principled accounts of extended cognition.

I suggested in the previous chapter that Palermos’s (2014) discussion of tactile visual substitution systems (TVSS) is indeed a strong candidate for the kind of reciprocal influence characteristic of extended coupled systems. While I think this is correct, I also think that TVSS cases are closer to the exception than the rule when it comes to interactionism.

If we take seriously the interactionist lessons from the person-situation debate, then the standard agent-artifact paradigm (e.g. Otto and his notebook, Dave and his iPhone, TVSS, etc.), operative in nearly all accounts of the extended mind and cognition, may be called into question. This is because many agent-artifact systems do not reliably exhibit the right kind of bi-directional, reciprocal feedback necessary for extension, at least not as reliably as many agent-agent systems will.²²

For present purposes, perhaps the most important connection between the extended cognition debate and the person-situation debate is the striking similarity between the interactionism exemplified in proposals like the ongoing feedback loops criterion and some models of social interaction from the history of social and personality psychology. Consider Darley & Fazio's (1980) model of the social interaction sequence:

- (1) Either because of past observations of the other or because of the categories into which he or she has encoded the other, a perceiver develops a set of expectancies about a target person.
- (2) The perceiver then acts toward the target person in a way that is in accord with his or her expectations of the target person.
- (3) Next, the target interprets the meaning of the perceiver's action.
- (4) Based on the interpretation, the target responds to the perceiver's action, and
- (5) the perceiver interprets the target's action. At this point, the perceiver again acts toward the target person and so can be regarded as reentering the interaction

²² I think that many agent-artifact couplings do, in fact, make the cut now, and given the advance of interactive computational technologies, it is highly likely that many more will in the near future. My claim here is simply that the dynamics of social interactions that exhibit the relevant characteristics (e.g. reciprocity) are more widespread and reliably realized. After all, TVSS and related technologies are often prohibitively expensive and not widely available. By comparison, the expectancy effects I discuss below are ubiquitous, especially in the context of friendships, romantic partnerships, and other close relationships.

sequence loop at Step 2...(6) After acting toward the perceiver, the target person interprets the meaning of his or her own action. (p. 868).²³

Kihlstrom (2013) describes this process – more accurately, I think - as a *cycle* of social interaction rather than a *sequence* (p. 787). Similar accounts can be found in Cantor & Kihlstrom (1987) and Jones (1986). It is important to note that these models are all elaborated in the context of research on expectancy confirmation effects, more colloquially known as self-fulfilling prophecies. My claim is that the interactions modeled in this research program will reliably exemplify the sort of feedback described by Palermos (2014), Heersmink (2015), Huebner (2016), and others.

In these kinds of cases, the ping-ponging between my expectations and subsequent signaling, your interpretation of them and subsequent behavior, my interpretation of your behavior and subsequent updating of my expectations, and so on, is paradigmatic of Palermos’s idea of “ongoing mutual interdependence on the basis of feedback loops” and Heersmink’s requirement of reciprocal information flow. This kind of bi-directional, mutual interdependence is at the core of many kinds of social interactions, and for this reason, offers a plausible paradigm for thinking about extension in terms of interactionism. Put another way, other agents are more likely than (many) artifacts to exhibit the required kind of time sensitive, reliable, and reciprocal influence.

If this is on the right track, then the disciplines that study these agent-agent couplings offer a wealth of untapped resources for extended cognition theorists. In the balance of this section, I will introduce some of the most promising research to this end.

²³ It is worth noting here that Darley & Fazio identify symbolic interactionism as the precursor to this model, and this is another reason why I think it is important to consider the role of the classical pragmatists like Dewey and Mead in the development of the extended mind and cognition. For more on the relationship between the “Chicago School” pragmatists and the extended mind, see Madzia (2013).

Recall that the original formulation of the extended mind involved Otto retrieving addresses from his notebook. The next domain of research in social and personality psychology I will consider also involves memory, but it possesses several advantages Clark & Chalmers's seminal thought experiment.

The concept of a Transactive Memory System is due to a series of papers by Dan Wegner and colleagues (e.g. Wegner (1986); Wegner, Giuliano, & Hertel (1985); Wegner (1995)).²⁴ Perhaps the most well-known is Wegner, Erber, & Raymond's (1991) "Transactive memory in close relationships." Therein, Transactive Memory Systems (TMS) are described as follows:

People in close relationships know many things about each other's memories...

Such knowledge of one another's memory areas takes time and practice to develop, but the result is that close couples have an implicit structure for carrying out the pair's memory tasks. With this structure in place, couples in close relationships have a transactive memory that is greater than either of their individual memories. Transactive memory is a shared system for encoding, storing, and retrieving information (p. 923).

Here we have a claim (with empirical evidence, to boot) that the cognitive processes of remembering, recollecting, cueing, etc. can be realized between two individuals in a romantic relationship. Romantic partners without an assigned structure (e.g. one partner is told to remember items from Categories 1 and 2, the other partner, 3 and 4) exhibited significantly more agreement about which partner was more expert in a given category ($M = 5.52$) than randomly assigned impromptu partners ($M = 4.04$) and romantic partners recalled significantly more items ($M = 31.4$) than did impromptu partners ($M = 27.64$).

²⁴ For helpful reviews of TMS research, see Lewis & Herndon (2011) and Ren & Argote (2011).

Wegner's description of this phenomenon has a clear link to the externalist frameworks described in the previous chapter: the "communication and updating of information each has about the areas of the other's knowledge," means that "*each partner cultivates the other as an external memory aid*" (p. 924, emphasis mine).

Perhaps the most surprising result from these studies is that that when structures were assigned to the partners, not only did the effects just described disappear, they actually flipped: romantic partners with an assigned structure recalled significantly fewer items ($M = 23.75$) than did impromptu partners ($M = 30.14$). In these cases, there seems to be something quite important about each partner cultivating and being cultivated by the other. That is, knowing what the other knows, and knowing that they know that you know it. And indeed, there has been recent uptake of TMS research in the extended cognition literature (e.g., Sutton et al. 2010; Harris et al. 2011; Theiner (2013); Huebner (2016); Kirchhoff (2015); Tollefson, Dale, & Paxton (2013)). Though it might not be construed in precisely these terms, one explanation for why extended cognition theorists can get mileage out of this kind of research is that the dynamics of TMS are inherently and obviously interactionist.

At the very least, TMS seems to provide a better lodestone for extended cognition theories than Otto's notebook. I want to suggest, however, that TMS research is just the tip of the iceberg when it comes to the kind of interactionist psychology of interest to us here. One upshot of TMS research is that it brings into focus a much larger body of work examining the ways that relationship partners sculpt one another, and TMS is but one example of this.

More specifically, research on close relationships provides conceptual and empirical support for the view that skin and skull individuals are not the sole bearers of cognitive and affective processes and states. To see this, I will introduce a body of research pioneered by Stephen Drigotas and colleagues on the *Michelangelo phenomenon*. The Michelangelo phenomenon brings together research from the history of psychology on interdependence, expectancy effects, the social self, and self-discrepancy theory and is commonly defined as: “a congenial pattern of interdependence in which close partners sculpt one another in such a manner as to bring each person closer to his or her ideal self” (Drigotas et al., 1999, p. 293).

The analogy at the heart of the theory is that in much the same way the sculptor “released a hidden figure from the block of stone in which it slumbered” (p. 294), so too do our romantic partners sculpt our actual selves into our ideal selves. Because of their proximity, interdependent relationship partners have a unique ability to exert strong, frequent, and efficacious influence on one another. The idea is that the intimate knowledge of each other’s goals and ideals allows partners to nudge and sculpt one another toward those ideals through channels not available to individuals in isolation.

In the literature, there are two mechanisms thought to underwrite this sculpting: perceptual affirmation and behavioral affirmation. When I perceive and behave toward my partner in ways that are congruent with her ideals for herself, I am slowly sculpting her into that ideal version of herself. And she can do the same for me. That is, we each elicit, or coax out behaviors and dispositions within one another that are congruent with our ideal selves. An oft-used illustration is that if Mary aspires to be more outgoing and conversant, John can direct the banter at a party to set Mary up to tell one of her most

charming stories. Of course, this can all work in the other direction too, such that John can perceptually and behaviorally *disaffirm* Mary such that she moves *away* from her ideal self.

Importantly, this framework yields a number of empirical predictions. For example, Drigotas et al. (1999) hypothesized that perceived partner perceptual affirmation would be positively associated with perceived partner behavioral affirmation. In turn, increased perceived partner behavioral affirmation would be positively associated with movement towards the ideal self, and finally, this movement toward the ideal self would be positively associated with couple well-being. The empirical results are mixed, but partner behavioral affirmation does seem to be linked with couple well-being. In other words, it looks like couple well-being and relationship stability increase the more Mary thinks that John behaves in ways congruent with her ideals, and the more John similarly thinks this about Mary.

In a (2009) review paper, Rusbult, Finkel, & Kumashiro report studies showing that when partners behaviorally affirmed each other's goals (e.g., helping to clarify plans, offering assistance, praising goal pursuit, etc.) each partner showed greater movement toward their ideal self and reported enhanced personal well-being and greater life satisfaction. Moreover, when partners are effectively sculpting one another in this way, they also report enhanced couple well-being, greater adjustment, and they show increased likelihood of relationship persistence and stability. The upshot, according to the authors is that "over time, adaptations that begin as temporary, interaction-specific adjustments become stable components of the self, such that over the course of extended interaction, close partners sculpt one another's selves: People come to reflect what their partners 'see

in them' and 'elicit from them'" (p. 305). This transactive account of relationships will be an important feature of my bearers argument in Chapter VI.

Much more could be said about research on the Michelangelo phenomenon, but this cursory review is only meant to show the empirical plausibility of research programs in the constellation of TMS-type research (some of which has already begun to work its way into the extended cognition program). Having hopefully shown that many concepts from the extended cognition literature have been operationalized in empirical work on close relationships, I want to turn my attention to the theoretical underpinnings of these empirical research programs.

As I showed in the previous chapter, the theories about embedding, scaffolding, and extending are well worked out in the philosophy of mind. Research on TMS and the Michelangelo phenomenon, however, has developed independently of this work in the philosophy of mind, and so it is worthwhile to examine the theories behind the former, and their similarities to the latter. What we will see is that as research on Michelangelo effects has progressed, a sort of anti-individualist, anti-internalist trend has emerged in parallel with these same developments in the philosophy of cognitive science. One such development has been the uptake of Michelangelo phenomenon-style research in the goal and self-regulation literatures. The following passage is representative of the state of play:

From the perspective of the most well-known and well-established models in the psychological literature, people either succeed or fail at self-regulation as individuals, toiling in isolation toward goal achievement. From the four year-olds in the famous delay of gratification studies (Mischel, 1974) who sit by themselves

trying not to eat a marshmallow to the hungry participants in the seminal study of self-regulatory depletion who sit by themselves eating radishes instead of cookies (Baumeister, Bratslavsky, Muraven, & Tice, 1998), psychological studies of self-regulation, and the theories that have inspired those studies, have predominantly focused on individuals pursuing their goals alone (Finkel, Fitzsimmons, & van Dellen, forthcoming, p. 22).

In a series of papers, Gráinne Fitzsimmons and colleagues have developed a new model of goal pursuit and self-regulation which rejects this dominant, individualist approach. Their alternative framework - what they call *transactive goal dynamics* – is developed in detail in a (2015) article. Therein, their approach is introduced as follows:

the theory [of transactive goal dynamics] depicts relationship partners as exerting such a great deal of mutual influence in each other's goals, pursuits, and outcomes that the partners' self-regulatory systems become inextricably linked, part of a complex and messy web of interdependence. Ultimately, we suggest that relationship partners are best conceptualized not as mostly independent goal pursuers who occasionally influence each other, but instead, as interdependent subparts of one self-regulating system (Fitzsimmons, Finkel, & van Dellen, 2015, p. 648).

Part of the motivation for this view is that while the classic experimental paradigms used to study self-regulation often yield clean, readily interpretable results, they do not necessarily reflect the ways most people go about trying to achieve their goals outside of the lab. It is not just that Pat has a goal to lose weight. It is also that Pat's partner Jordan is training for a marathon. So when they go grocery shopping together, each can help the

other achieve their goals by, say, purchasing healthy food. Pat trying to lose weight does not happen in a social vacuum. After all, think about how well things would go over if Jordan was instead training for an eating contest.

It is crucial to see, though, that leveraging a criticism against the dominant individualist views of self-regulation amounts to more than a criticism of ecological validity. The positive claim is that social processes play an important role in self-regulation that has been underappreciated by researchers. Put another way, a narrow focus on intraindividual processes (like willpower: Pat trying to resist the temptation to eat the cream cheese brownie) often renders invisible interindividual processes (like affirmation: Jordan's subtle nod of acknowledgment when Pat passes up the cream cheese brownie) that may turn out to be highly relevant to self-regulation and goal pursuit.

Whether or not these social processes have predictive or explanatory value is, of course, an empirical question and one that cannot be settled here. But I am suggesting that there are good reasons to think that some interindividual processes will turn out to be more important to self-regulation than was previously thought. If this is correct, we are confronted with a version of the same challenge raised by Rupert (2009) in the previous chapter. Why bother with a complicated, messy framework positing higher-order individuals when a simpler, more traditional model is widely available? Or, as Finkel, Fitzsimmons, & van Dellen put it, "Why must we consider a challenge to the pervasive (if implicit) assumption that goal-related processes predominantly reside within a single individual?" The answer, they contend, is that:

If the goal of self-regulation research is to develop models of goal-relevant processes that emerge within prototypical laboratory experiments—those in which individuals are, for example, seated alone at a table with a marshmallow on it or at a computer screen with tempting images on it—then the individual unit of analysis may well be optimal. If, in contrast, self-regulation researchers want to understand how people set, pursue, and achieve goals in their everyday lives, then a predominant focus on individual-level processes is likely to yield an incomplete, perhaps even inaccurate, understanding of goal dynamics. Consequently, we argue, any conceptual model seeking to capture such dynamics must address topics like how pervasively people set and pursue goals for others and have those others set and pursue goals for them, and the circumstances under which social processes promote versus undermine goal achievement. (forthcoming, p. 3)

What is striking in the present context is the emphasis on how traditional models construe the relevant self-regulatory processes as “residing *within* the individual” and how such a view yields, at best, an incomplete and, at worst, an inaccurate account of self-regulation and goal pursuit.

This is, in effect, the same argument deployed by Sterelny, Clark, Palermos, and others in the extended cognition literature. The best of these arguments do not claim that internalism about cognition is flatly false *tout court*. Instead, the claim is that a blinkered focus on internal processes alone yields an incomplete picture, and perhaps one that overlooks some of the most interesting and important features researchers ought to be interested in. I take Fitzsimons and colleagues to be making the same kind of claim. It is not that internal willpower does not matter at all. It is rather that a myopic focus on

willpower - at the expense of the unique interpersonal dynamics in close relationships – will render invisible features of self-regulation that might turn out to be crucial for understanding whether and how people achieve their goals.

It is not a far leap from here to the claim that some process relevant to self-regulation sometimes extend to include interdependent partners – no doubt, an attractive result for the extended cognition theorist. And indeed, the template developed in the previous chapter could provide a set of criteria for more precisely conceptualizing the degrees of influence partners can have. The connection to extended mind-style arguments should be clear enough, but what exactly is the model being proposed by Fitzsimmons and colleagues? And how does it provide novel insights into self-regulation? Their (2015) paper breaks Transactive Goal Dynamics into 6 tenets, but I will simply summarize the relevant dimensions here.

Echoing Palermos's (2014) ongoing feedback loops criterion, the first tenet of TGD is that “relationship partners’ goals and pursuits are so strongly interdependent that they are most accurately characterized as one system,” and that the “complexity of overlap and interaction among the partners’ goal dynamics can be great enough that, in effect, the relationship forms its own system of goal pursuit” (Fitzsimmons, Finkel, & van Dellen 2015, p. 651). From the outset, it is clear that something like the coupled systems perspective from the extended cognition literature is needed to account for some of the intricacies of self-regulatory processes. That is, looking at John as an independent, autonomous goal pursuer and Mary as another independent, autonomous goal pursuer might be sufficient in some cases. But when John and Mary become more and more interdependent as their relationship evolves, treating their goals in isolation from one

another risks ignoring crucial, perhaps decisive, pathways through which each can achieve (or fail to achieve) their individual goals, or through which they can both achieve (or fail to achieve) their joint goals.

Once the relationship is viewed from this coupled systems perspective, it can be further characterized in terms of *transactive density*, or the degree to which partner goals are in fact intertwined (tenet two). The greater the transactive density, the greater the interdependence of goals and hence, greater plausibility of construing two partners as one higher-order goal system. TGD predicts that variability in transactive density is explained by opportunities and motivation for interdependence. In other words, the more partners are motivated to form interdependent goals and the more opportunities there are to do so (such as shopping together), the pathways of influence between them will become increasingly symmetrical, balanced, and reciprocal.

The concepts of *transactive gain* and *transactive loss* then track the degree to which goal interdependence produces positive or negative effects (tenet three). When partner goals are well-coordinated and transactive density is high, partners can pool resources to achieve goals more effectively together than they could independently, thereby experiencing transactive gain. But when goals are poorly-coordinated and transactive density is nonetheless high, partners experience transactive loss, where goals are pursued less effectively together than if they had been pursued independently. In turn, this goal coordination is predicted by shared goal representations and relationship orientation and skills (tenet four). To the extent that couples possess these two variables, their goals will better coordinated, and goal outcomes will increase as density increases.

To the extent that couples lack these orientations and skills, their goals will be less coordinated and goal outcomes may decrease as density increases.

Taken together, these concepts also yield predictions about relationships: greater transactive gain is hypothesized to be positively associated with relationship stability (tenet five). The concepts of transactive gain and loss are also crucial for answering Rupert's challenge. Sure, it might be sweet to think that long-term romantic partners *figuratively* become a part of one another, but why think this is *literally* the case? In other words, why bother positing a complicated theory that the relationship itself (rather than the individuals therein) is the bearer of self-regulatory processes and goal states?

The answer is that transactive gains and losses offer novel predictive and explanatory resources unavailable to narrowly individualist accounts of self-regulation and goal pursuit. While inner fortitude and strength of will might partially explain why Pat passed on the cream cheese brownie, a better explanation might be that her weight loss goals are bound up with her partner's fitness goals, and so the presence of her partner's knowing glance in the bakery aisle helps explain why she declined the brownie.

Because this is a relatively new theory, the jury is still out on the extent to which the transactive goal model enjoys empirical support. It could well turn out that individualist, internalist accounts of willpower are more explanatorily powerful than the interpersonal processes posited by transactive goal dynamics theory. But if the literature on the Michelangelo effect is any guide, then we should expect that these interpersonal sculpting processes will provide unique explanatory resources. Again, these are empirical questions that cannot be settled here. But if we take a step back from the empirical questions, the terminology, and the history of scholarship which gave rise to TGD theory,

it is important to note that what is on offer here is a conceptual and empirically testable framework advancing claims that processes relevant to memory, ideal selves, self-regulation, and goal pursuit are not strictly inside of skin and skull individuals. The interactionism underwriting these kinds of claims in Michelangelo and transactive goal models leads me to conclude that research in social and personality psychology ought to be the next wave of extended cognition research. But it is not just that research on close relationships can corroborate various claims of extended cognition (though it can serve that function too), but that it offers a principled way for talking about the *bearers* of cognitive and affective processes as also extended.

III.V Conclusion

From an empirical perspective, one might think that an unfortunate tendency among philosophers of mind is their tendency to use far-fetched, sci-fi fantasy examples to make arguments or construct counterexamples. One of the primary aims of this chapter has been to show that the arguments for extended cognition need not rely on such thought experiments for their support. The kinds of claims advanced by proponents of extended cognition are strikingly similar to research programs in social and personality psychology, especially those programs focusing on the dynamics of close relationships. In other words, I hope to have shown that far from being radical to think of cognitive and affective processes as extended, these are empirically tractable, and indeed, empirically supported claims. The fact that entire research programs in psychology have developed independently of the externalist arguments in the philosophy of mind also speaks to the plausibility of broadly externalist frameworks.

But how exactly should we think of the relationship between the research on transactive memory and transactive goals one hand, and scaffolded and extended cognition on the other? I contend that the research canvassed here in social and personality psychology shows that it is not just that cognitive and affective processes can be extended, but the very bearers of these states need not be skin and skull individuals, either. Where the previous chapter offered reasons for thinking that cognitive and affective processes might sometimes extend to include various artifacts and gadgets, the research covered in the present chapter offers reasons for thinking that these cognitive and affective processes can extend to include interdependent partners.

To come at this from another angle, consider the subject matter of this chapter: personality traits, memories, and goals. A case can be made that each of these, in some sense, constitutes our deepest sense of who we are. Indeed, there is ample historical precedent for thinking that memories are at the core of personal identity (e.g., Locke, 1690). Or when Nietzsche claims “your true nature lies, not concealed deep within you, but immeasurably high above you, or at least above that which you usually take yourself to be,” (1874/1997, p. 129) this could easily be interpreted along the lines that our goals and projects constitute the self. What the research presented in this chapter has suggested, however, is that there is good reason to question whether our personality traits, memories, or goals actually reside *inside* of us at all. And if there are independent reasons for thinking that these features are especially important for self and identity, then there are reasons to think that individuals might not be as internalistic and individualistic as is often assumed. Or, as Clark & Chalmers put it, “far better to take the broader view, and see *agents themselves* as spread into the world” (1998, p. 18, emphasis mine). Where

Clark & Chalmers, Palermos, and others tend to see our tightly integrated gadgets as the most salient feature of the world into which we sometimes extend, I have argued here that it is actually interpersonal dynamics that are most salient, and that this is where the case for extension is the strongest.

If my arguments have gone through so far, then there are (at least) two interpretations available. The *weak* interpretation is that internalists about cognitive and affective processes and/or individualists about the bearers of these processes are no longer entitled to a default status. I have developed and introduced theoretically sophisticated and empirically plausible externalist and/or anti-individualist alternatives which require that the internalist and/or individualist argue for correctness of their view rather than merely assert it. It could of course turn out that the arguments do cut in favor of internalism and/or individualism, but this cannot be assumed in advance. In other words, I take the foregoing chapters to have established, at the least, that it is reasonable to argue that cognitive and affective processes and their bearers sometimes extend beyond the boundaries of skin and skull. And insofar as that is right, the externalist and/or anti-individualist does not bear the full weight of the argumentative burden.

Now of course, I have attempted to muster evidence suggesting that in at least some, and perhaps in many cases, the arguments actually do cut in favor of the externalist and/or anti-individualist. One way this argument has unfolded is by showing how processes of interest are rendered invisible by overly internalist and/or individualist frameworks, and only made visible through something like a coupled systems perspective. The *strong* interpretation, then, holds that internalism about cognitive and

affective processes and/or individualism about the bearers of these processes is mistaken and ought to be rejected, at least in some, and perhaps in many cases.

The correct interpretation is likely somewhere in the middle, and will likely have to be evaluated on a case-by-case basis. Still, the consequences are far-reaching. After all, the very nature of selves, thinking, and feeling are at stake here. In the second half of the dissertation, I will narrow in on one range of consequences. Namely, how the kind of internalism and individualism discussed in the previous two chapters is also assumed in much virtue ethical theorizing. Thus, many of the arguments deployed in the first half of the dissertation will be applicable in the ethical domain as well. In response, I will develop a range of responses to these externalist challenges, namely, a novel account of extended virtues.

CHAPTER IV

REVIEW

Introduction

The overarching aim of the second half of the dissertation is to show that the claims developed in the first half are consequential for empirically informed theories of virtue, and that there is value added for adopting a broadly externalist framework about virtue. In the previous chapters, I laid out a number of arguments which challenged internalism about cognitive and affective processes and challenged individualism about the bearers of those processes. Taken together, these arguments called into question default assumptions about the nature of cognition, emotion, and self.

There is a broad sense in which such questions about cognition, emotion, self, personality, relationships, etc. are deep questions about what it means to be a human being, and hence, are connected to ethical considerations. It is hardly surprising, then, that proponents of extended mind-style arguments would recognize that their views entail the reconstruction of these concepts not just in the philosophy of cognitive science, but also in moral philosophy. For example, Clark (2001) offhandedly mentions, that “we need (I suspect) an account of personal responsibility and moral agency which respect the thin, decentralized and distributed nature of situated intelligent control (p. 141). And indeed we do.

In this chapter, I will argue that while proponents of extended cognition have identified some promising lines of inquiry into the ethical implications of their views, the literature remains underdeveloped and unsatisfying. I will survey the literature examining

the ethical implications of extended cognition and highlight the shortcomings and gaps which will then be the subject matter of the next two chapters.

Perhaps the most substantive engagement between extended cognition broadly construed and ethics broadly construed is Andrew Sneddon's (2011) book *Like Minded: Externalism and Moral Psychology*, and this will be my focus in Section IV.II. Next, in Section IV.III I will consider some of Neil Levy's work which examines the applied ethical dimensions of extended cognition. In Section IV.IV I examine work by Mason Cash which has brought the extended cognition literature into dialogue with feminist accounts of self, identity, and autonomy. In Section IV.V I will take up the arguments for extended persons developed by Robert Howell (2016) – which are closest to my own – and show that much work remains in formulating a plausible account of extended virtues. I then conclude by highlighting that while there is much promise in the work surveyed in this chapter, the literature is just that: promissory. Identifying this lack of substantive engagement with normative ethical theory then sets the stage for my positive project in the last two chapters.

IV.II Sneddon's *Like Minded*

Andrew Sneddon's (2011) *Like Minded: Externalism and Moral Psychology* is probably the most thorough engagement between the extended mind broadly construed and ethics broadly construed. In a sentence, Sneddon argues for the Wide Moral Systems Hypothesis (WMSH), which aims to give externalist interpretations of the research on four aspects of our moral psychology: (1) moral judgment, (2) moral reasoning, (3) the attribution of moral responsibility, and (4) moral motivation. I will focus my attention on (2) and (4), as they are most germane to the present project.

As a general overview, Sneddon tries to show that much of the philosophical and psychological literature on these four topics in moral psychology takes a narrow individualism about the mind as the default. In the first place, for Sneddon, “externalism” is contrasted with “individualism.” Sneddon’s “primary aim is to make plausible the idea that the psychological foundations of morality should be understood, at least partly, in terms of cognitive systems that extend into the environment beyond the physical bounds of individual agents” (p. 9).

The argumentative strategy of the book is the following: to understand moral judgment, moral reasoning, responsibility attribution, and moral motivation we have to look at the ways in which external outputs from other people (behavior, expectations, etc.) serve as tightly integrated inputs to supposedly internal mechanisms for moral judgment (and reasoning, attribution, and motivation). With respect to (1), Sneddon argues that the mechanisms that realize moral judgments are plural and hybrid. That is, the mechanisms are constituted by a motley of internal and external features. This view is contrasted with what Sneddon calls unity theories of moral judgment such as Shaun Nichols’s (2004) sentimental rules theory, Jonathan Haidt’s (2001) social intuitionism, and Jesse Prinz’s (2007) moral sensibility theory.

A recurring theme in the book is that the plurality and hybridity of moral psychological mechanisms makes them more apt to enlist features of the external environment, especially other people. So with respect to (3) Sneddon claims that “the heterogeneous psychology of the attribution of moral responsibility is partially realized by wide cognitive systems...the resources in question for attributions of responsibility are

those provided by such external symbol systems as language, and most important of all, by the minds of other people” (p. 155).

With this brief preview of the book’s argumentative style in place, I will turn my attention to a more detailed treatment of the externalist aspects of (2) moral reasoning and (4) moral motivation and the production of action.

The argument for (2) rests on the premise that “moral reasoning is realized, first and foremost, in an interpersonal moral reasoning system (or multiple systems)” (p. 72). Because moral reasoning is also plural and hybrid, Sneddon thinks we should not be surprised to find that in reasoning and deliberation, moral agents often rely heavily on others. The externalist angle here is that other people’s expectations, knowledge, beliefs, desires, expertise, etc., can figure so prominently into the reasoning processes of an individual that construing other people as mere background or context undersells and misrepresents their contribution. Sneddon goes so far as to claim that this interpersonal dimension is actually the central feature of reasoning:

The present hypothesis is that moral reasoning systems draw on interpersonal resources that play a replicable causal role in social influence. For any given individual, moral reasoning seems to be primarily a way of interacting with others, and in particular for influencing others. It is only secondarily for bringing about effects on the individual. Influencing one’s processes of attributing responsibility or producing actions fall into the class of secondary jobs of moral reasoning, not into its class of primary jobs. (p. 73).

If this view is correct, then the more appropriate framework for thinking about moral reasoning (though Sneddon does not use this language) is a coupled system. Indeed, at first blush, this seems very much in line with my claims about social interaction as the

best fit for theories of extended cognition. More careful reflection, however, reveals two related problems with Sneddon's account.

First, Sneddon's arguments about externalist moral reasoning are subject to potentially devastating objections. In the absence of principled criteria²⁵ about when the expectations, expertise, etc. of others are so tightly integrated as to be constitutive of the reasoning process, Sneddon's formulation of widely realized moral reasoning systems will be left open to the charges of coupling-constitution and cognitive bloat.²⁶ This worry is all the more pressing in the absence of any specification about who the other people are. I imagine a strong case could be made on behalf of romantic partners or close friends, less so with strangers or online social media interlocutors. The point is, without any of these details worked out in advance, Sneddon lacks a convincing response to a *reductio* argument claiming that my moral reasoning system extends to include everyone on the internet reading my social media posts.

The second problem is that it is unclear whether Sneddon is advancing a particularly externalist claim. In other words, there does not appear to be much in Sneddon's arguments that would block the individualist's contention that indeed, other people play an important role in formulating the content of moral reasoning, but the *real* process of moral reasoning still takes place between the ears of an individual. This is

²⁶ To be fair, Sneddon does offer a formulation of what appear to be such criteria, but it is strange, and it is so general it is hard to see how it would be useful in practice: "_____ systems must be causally and functionally integrated chains of _____ resources, and these, individually and collectively, must play a replicable causal role in _____" (p. 7). The blanks are filled in with, e.g., "moral psychological," "cognitive," and the like. Sneddon notes elsewhere that externalist hypotheses "are warranted when there is evidence of the causal and functional integration characteristic of systematic individual-environment relations... The greater the causal-functional integration between individual and environment with regard to a given psychological phenomenon, the more warrant there is for externalist hypotheses about that phenomenon" (p. 192). Again, this is fine as far as it goes, but the devil is in the details, and many of the relevant details needed to deflect potential objections are simply absent.

related to the first problem because again, in the absence of principled criteria, we are left in the dark about when and how another person's beliefs, expectations, knowledge, etc., can become so tightly integrated that the usual individualist framework is unable to account for it.

Insofar as Sneddon's project is an externalist one, then the inability to convincingly demonstrate the benefits of his framework over its individualist rivals is deeply troubling. Indeed, this very criticism is raised throughout Neil Sinhababu's (2012) review of Sneddon's book. For example, "Nothing here requires a drastic revision of our concepts of belief and desire, or our picture of the mind. Rather than accepting extended mind, we can stick with a compact mind whose states are heavily affected by the world." (Sinhababu, 2012, p. 828). And similarly, "Extended mind again seems irrelevant. That some mental states (perceptual states, for example) are deeply dependent on the environment has always been accepted by the traditional compact mind view." (ibid). Put another way, Sneddon has no answer to the challenge articulated by Rupert (2009) in Chapter II: why posit a radical view when a simpler, more conservative view will do?

I hasten to add that these criticisms are not meant to undermine the possibility of externalist interpretations of moral psychological research on reasoning. On the contrary, I endorse Sneddon's claim that the vast majority of research on moral reasoning is overly individualistic and that broadly externalist alternatives are needed. My claim here is simply that Sneddon's case, as it stands, is not going to compel anyone antecedently hostile toward externalism to join the cause. In the case of moral reasoning specifically, Sneddon's account could be bolstered against the above charges by adopting something like the template I developed in Chapter II to talk more precisely about where and how

moral reasoning extends to include other people. Further, I merely scratched the surface of the literature on close relationships in Chapter III, but this body of work could potentially put some empirical flesh on the theoretical bones of Sneddon's claims about dependence and reliance.

I will turn now to (4), Sneddon's externalist account of moral motivation and the production of action. This section of the book is framed with respect to the person-situation debate, which I argued above offers a fruitful outlet for the next wave of extended cognition theorizing. Here is Sneddon's gloss: "Recent years have seen the development of the implications of the 'person-situation' debate in psychology for philosophical discussions of virtue." He finds that "the reception of this work has been lukewarm at best. I, for one, have been convinced, so I find this a bit puzzling" (p. 157). What is Sneddon convinced of? That "situationists offer a much more context-sensitive account of our psychology" (p. 159). Sneddon takes "such context sensitivity to be a clue that our action-production systems are widely realized" (ibid).

Sneddon equates his brand of externalism with situationism and argues that the person-situation debate ought to be settled in favor of the situationists. He doubles down on this claim, saying "both those making the [situationist] case and those resisting it have underestimated the scope of the implications of this work" (p. 157). Thus, the evidence from situationist social psychology undergirds the argument for the Wide Moral Systems Hypothesis. Unfortunately, this strategy will fail.

In Chapter III, I provided a number of reasons to think that the person-situation debate was settled not in terms of situationism, but interactionism. Perhaps the strongest reason was the empirical finding that situations do not account for significantly more

variance in behavior than do personality variables (Richard et al, 2003). Add this nearly universally accepted conclusion (among psychologists, anyway) to the ongoing replication crisis in social psychology (OSF, 2015) and the prospects for philosophical arguments built on situationism look, at best, shaky, and at worst, wrong-headed.

Returning to the argument, Sneddon claims that the philosophical import of situationism is to be found in the philosophy of action: “To put it as straightforwardly as possible, I will return philosophical discussion of this [person-situation] debate to its original topic: the production of action” (p. 157). And similarly, “the situationist challenge is radical in the sense that it applies to the very root of consideration of the production of actions...Situationism calls for revision of familiar philosophical schemas of the production of actions, such as Davidsonian causalism” (p. 186). I worry that this emphasis mischaracterizes the debate and its uptake in the philosophical literature.²⁷

Closer to the mark, perhaps, is the uptake of the person-situation debate in the context of virtue ethics, which Sneddon does mention, albeit briefly. By this point, the arguments have been played out in the philosophical literature (e.g., Snow, 2010), but the basic idea is that classic situationist social psychological studies (e.g., Milgram 1974) show that most people lack the cross-situationally robust traits of character countenanced by (neo)Aristotelian virtue ethics. Sneddon does not add anything new to this debate and does not treat it in any detail. A less charitable interpretation might suggest that Sneddon is mistaken to not treat the virtue literature in more detail. A more charitable

²⁷ It is not clear that the person-situation debate is, at its root, about the production of action in the sense that philosophers like Davidson talk about the production of action. I think a more accurate historical framing casts the person-situation debate in terms of individual differences and the prediction of behavior. Some of Walter Mischel’s earliest work exemplifies what I have in mind: Mischel (1965) examined which personality variables are relevant for predicting who would succeed as a Peace Corps volunteer. To be sure, the debate that evolved from these kinds of studies might inform the philosophy of action in one way or another, but Sneddon does not offer enough evidence to support the claim that this is at the center of its philosophical import.

interpretation sees this as an opportunity for future work, and indeed, parts of the next the next two chapters will address some of these issues. At first pass, I argued above that the lessons from the person-situation debate have important bearing on the debates over extended cognition. In turn, principled formulations of extended cognition may have important bearing on how we think about the nature of character.²⁸

It is unfortunate that Sneddon's argument for an externalist account of moral motivation rests on a questionable reading of, and dubious inferences from, the person-situation literature. Insofar as the plausibility of his brand of externalist moral psychology relies solely on situationist foundations, I think the argument will fail. I hope to have shown in the previous chapter how research in social and personality psychology might be put to better use for externalists. It is worth emphasizing again that the aim of my criticisms is not to undermine the viability of an externalist moral psychological research program. Rather, the aim is to underscore the fact that while some proponents of extended cognition have begun to take stock of the ethical import of their views, much work remains. I will conclude this section, then, with some more general remarks about Sneddon's book and some of the questions it raises.

Like Minded is, first and foremost, a book which re-describes and offers re-interpretations of classic findings in moral psychology. Because of these largely descriptive aims, Sneddon concludes the book by saying things like: "the WMSH offers us no new options for making the world a better place" (p. 235) and "the Wide Moral Systems Hypothesis has little practical promise for education and therapy at the moment" (p. 248). This all seems correct and there is nothing intrinsically problematic about doing

²⁸ See Alfano & Skorburg (2016) for more on this point.

(re)descriptive work. However, when put in such stark terms, these reflections do highlight the need for a more substantive engagement with normative ethical theories.

In terms of its positive contribution, *Like Minded* can be read as advancing a strong claim and a weak one. The former is something like the view that individualist interpretations of moral psychological research are strictly false, and the Wide Moral Systems Hypothesis provides a more plausible alternative. For the reasons laid out above, I do not believe this claim is adequately supported. A weaker interpretation, which I think Sneddon succeeds at defending, is that there are viable alternatives to standard individualism about moral psychological phenomena. It is not clear that the Wide Moral Systems Hypothesis will be the standard-bearer here, but it starts us down the right track.

In the end, Sneddon makes a convincing, though perhaps not a knockdown case, that individualism need not be the only game in town. In the end, this opens the door for a more substantive engagement between externalism and normative ethical theories like virtue ethics.

IV.III Applying the Extended Mind: Enhancements and Assaults

In this section, I will examine the small literature that has emerged around the applications of the extended mind. Some of this is rather well-developed, but I will nonetheless identify some problems and shortcomings to be addressed. Put another way, much of this work makes important strides in an applied ethics context, but it does not directly address what I take to be the larger issue, namely, the lack of substantive engagement with normative ethical theories.

In his (2007b) *Neuroethics: Challenges for the 21st Century* and his (2007a) “Rethinking neuroethics in light of the extended mind thesis” in *The American Journal of*

Bioethics, Neil Levy has raised the question: What follows for neuroethics debates if the hypothesis of extended cognition is true? The introduction to his book would serve as a fine introduction for the present project:

understanding the mind properly plays a significant role in motivating an important alteration in the way ethics is understood, and in what we come to see as the bearers of moral values. What might be called an externalist ethics gradually emerges from the pages that follow, an ethics in which the boundaries between agents, and between agents and their context, is taken to be much less significant than is traditionally thought” (Levy 2007b, p. xi).

While our emphases differ, Levy’s program nonetheless serves as an important guide. Throughout his work, Levy shows how most neuroethicists unsurprisingly take a narrow internalism about the mind as a default background assumption. But once something like scaffolded or extended cognition is recognized as a viable alternative, neuroethical concerns premised on internalist presumptions are also called into question.

The central argument is that if we seriously entertain the plausibility of extended cognition, then the widespread assumption in neuroethics that internal, neurological interventions are, by virtue of their internality and invasiveness, morally problematic or fraught, no longer holds. If the mind extends beyond the skull into the environment, the argument goes, then interventions which alter those parts of the environment are not relevantly different from directly intervening on the brain. Here is how Levy puts the point:

Consider the growing literature on the topic of neuroenhancement (Buller 2006; Kass 2003; Parens 2006). This debate seems largely to concern whether such

enhancements are permissible. But the question of whether neuroenhancements are permissible can be an ethically pressing one only if we have a real choice about whether or not such enhancements are used. In the light of the extended mind, the implicit presupposition that we have such a choice is, if not clearly false, at least in need of sustained defense...it seems that we are instead forced to conclude that we have no choice about whether to allow interventions into the mind. Instead, the important questions concern *which* interventions are permissible. (Levy 2007a, p. 9).²⁹

By Levy's lights, insofar as the hypothesis of extended cognition is true, we are always already intervening on the mind. To see this from the opposite side, consider a line of argumentation popularized by Julian Savulescu (e.g., Savulescu, 2005). Everyone recognizes that education is permissible, desirable, and perhaps even obligatory. Everyone also recognizes that education – say, learning a new language – changes physical structures and processes in the brain. If, as seems increasingly likely, we are able to alter such structures and processes in the brain via neuroenhancements, do we have the

²⁹ It is worthwhile to note here, if briefly, how Levy's arguments have also spurred debates in the philosophy of enhancement. While a detailed treatment of questions of enhancement are beyond the scope of the present project, this literature has made important strides which do bear on my own project of developing a more substantive engagement between extended cognition and normative ethical theories. For example, Richard Heersmink (forthcoming) highlights how the cognitive enhancement debate largely focuses on internal interventions and enhancements at the expense of external artifacts, and how situated cognition approaches conversely focus largely on external artifacts, at the expense of their moral dimensions. He thus opens the door for a more sustained discussion of the moral dimensions of various mind-extending artifacts, arguing that "authenticity, equal access, and fairness may be at stake when using cognitive artifacts" (p. 8). Heersmink also hints at a few positive claims, suggesting that future work on the moral dimension of cognitive artifacts ought to focus on questions like "(1) what cognitive skills do we want to have in an information society; and (2) how to design these artifacts and systems such that they optimize well-being (broadly construed) and reduce possible negative consequences on brain, cognition, and culture" (p. 11). And he claims that the "degree of dependency and integration are proportional to the artifact's moral status" (p. 12). This is all ripe for a dialogue with any number of normative and political theories.

same obligation to enhance as we do to educate? Savulescu thinks that answering in the negative is little more than a manifestation of internalist prejudice.

These arguments are two sides of the same coin. The extended cognition literature helps us to see that there is nothing sacred about the boundaries of skin and skull, and this is not only true in the context of the philosophy of cognitive science, but also in the context of applied ethics. Levy puts the point succinctly: “It doesn’t really matter what we call mind and what we don’t. We instead need to rethink the significance of the boundary between inner and outer - including, perhaps especially, its ethical significance” (2007b, p. 59).

Levy attempts to formalize this insight into applicable ethical principles, and I believe this is one of the most important positive contributions to the literature on the ethics of extended cognition. There are two versions of what Levy dubs the ethical parity principle (EPP). The first, “EPP Strong” assumes the truth of the hypothesis of extended cognition: “since the mind extends into the external environment, alterations of external props used for thinking are (*ceteris paribus*) ethically on a par with alterations of the brain” (2007b, p. 61).

In practice, this would mean that if we think taking psychotropic drugs to enhance mood or performance is authenticity-undermining or unfair, then we should be equally worried about the external props constituting the extended mind that may similarly enhance performance or mood. And on the flipside, if we do not have qualms with interventions into the extended mind (say, enhancing memory with smartphones), then sufficiently analogous interventions into the brain should be equally permissible.

The real thrust of EPP Strong is that the distinction between inner and outer is not intrinsically morally significant. An intervention cannot be impermissible solely on the grounds that it is done directly on the inside, to the brain. And interventions into the environment cannot be permissible solely on the grounds that they are done on the outside, to the environment.

Levy is quick to point out, however, that “EPP Strong” stands or falls with the hypothesis of extended cognition. To the extent that extended cognition is implausible, so too is the ethical parity principle built upon it. For this reason, Levy also introduces “EPP Weak” which instead relies only on the plausibility of (something like) the hypothesis of embedded cognition: “Alterations of external props are (*ceteris paribus*) ethically on a par with alterations of the brain, to the precise extent to which our reasons for finding alterations of the brain problematic are transferable to alterations of the environment to which it is embedded” (Levy 2007b, p. 61).³⁰

Much of the book is comprised of an application of “EPP Weak” to classical issues in neuroethics (e.g., pharmaceuticals and authenticity, enhancement, mind-reading, etc.). On the one hand, this strategy has the advantage of a wider appeal. One need only accept the comparatively uncontroversial hypothesis of embedded cognition to appreciate the import of the Ethical Parity Principle. But on the other hand, insofar as Levy relies mostly on “EPP Weak,” and thus downplays the importance of the hypothesis of extended cognition, he remains relatively silent on what could turn out to be some of the most pressing neuroethical issues. Namely, the kinds of cases revealed by the framework

³⁰ DeMarco & Ford (2014), however, reject EPP strong and propose the following amendment of EPP weak: “Alterations of external props are ethically on a par with functionally similar alterations of the brain, to the precise extent to which reasons for finding the functional alterations of the brain morally acceptable or unacceptable equally apply to reasons related to the functional alteration of external mental props” (p. 323).

of extended cognition, where agents are so tightly coupled to external resources that such couplings *constitute* the unit of moral concern.

On precisely this point, it is worthwhile to take a short detour to consider a recent paper by J.A. Carter & S.O. Palermos (2016). The paper opens with an invitation to reflect:

How would we feel, and what actions would we take, if someone intentionally broke our phone, stole our smartwatch, or hacked our laptop in a way that significantly undermined our ability to organize our lives? Suppose, for example, that someone compromised the gadgets you rely on daily, such that your diary appointments, your contacts list, photos, system preferences and functionalities, research notes, folders, reminders, push notifications, and so on have all turned into a jumbled, corrupted mess of disorganized data. Would it feel like someone merely damaged your property? Or might the situation perhaps be worse, as if someone had actually assaulted you? (Carter & Palermos, 2016, p. 543).

The intuition here, of course, is that many of us have become so reliant on our smartphones, smartwatches, laptops, etc. for basic day-to-day activities that disrupting our access to them is closer to disrupting our mind rather than disrupting a mere tool. We might also imagine the emotions and feelings that would be elicited if we showed up to teach a class or present at a conference, only to realize we left our laptop at home. If this thought evokes a flash of anxiety, then we are well on the way to appreciating the import of Carter & Palermos's proposal.

Though they do not cite Levy's ethical parity principle, the argumentative strategy is similar: insofar as some external resource is properly integrated with an agent (e.g., the

ongoing feedback loops criterion from Palermos (2014)), then damaging that external resource is akin to damaging the person. In the context of assaults, then, the boundary between inner and outer is not morally significant. Carter & Palermos formalize this with their formulation of the Argument for Extended Assault (AEA):

(P1). Intentional harm to a part of a person which is responsible for her mental and other faculties constitutes personal assault. [Definition]

(P2). Our mental faculties can be partly constituted by external artifacts, so long as these artifacts have been appropriately integrated into our overall cognitive system. [from HEC]

(C). Therefore, having our integrated epistemic artifacts intentionally compromised plausibly qualifies as a case of personal assault [From P1 & P2] (p. 549).

Returning to the context of Levy's ethical parity principle, the worry here is that by backpedaling to the weaker interpretation, it may deliver the wrong verdict on cases of personal assault. If we concede that Otto's notebook merely *embeds* his memories and beliefs, then the willful destruction of the notebook looks like a run of the mill property crime. But if we stay true to the spirit of the thought experiment and grant that the notebook *extends* various parts of Otto's mind, then the willful destruction of the notebook is actually closer to willfully destroying, for example, the hippocampus, and hence, closer to a crime against the person rather than their property.

To sharpen this point, Carter & Palermos also ask readers to consider cases where patients with severe memory defects rely on smartphone reminders, post-it notes, digital picture frames, and the like to remember important activities such as taking medications,

feeding the dog, or brushing one's teeth. The claim is that in these kinds of cases, the props are so crucial to the person's functioning that *property crime* would not adequately characterize the extent and depth of the harm in cases of willful destruction.

The argument for extended assault, however, does not rest only on phenomenological reflection or intuition pumping. Carter & Palermos point to a recent U.S. Supreme Court ruling suggesting that something like the argument for extended assault may have traction in the courts.

According to the legal precedent entitled Search Incited to Arrest (SITA), also known as the Chimel Rule, police officers are allowed to conduct a warrantless search of an arrested person's physical area, specifically the area under their immediate physical control. This is meant to prevent the destruction of evidence and also to remove weapons or other objects that might be used for escape. As Carter & Palermos point out, this does not license a warrantless search of the arrestee's mental states (e.g., a polygraph test), but it does license a warrantless search of physical items in their vicinity. The question, of course, is whether smartphones are more like a part of mental states, or more like any other physical object in the vicinity. According to Chief Justice John Roberts:

modern cell phones... are now such a pervasive and insistent part of daily life that the proverbial visitor from Mars might conclude they were an important feature of human anatomy. Modern cell phones are not just another technological convenience. With all they contain and all they may reveal, they hold for many Americans 'the privacies of life'. The fact that technology now allows an individual to carry such information in his hand does not make the information

any less worthy of the protection for which the Founders fought. (quoted in Carter & Palermos, 2016, p. 552).

And indeed, the Court unanimously held that SITA does not cover the warrantless search of the contents of cell phones. To be sure, this is far from a knockdown argument for extended assault or the ethical parity principle. But it is an important proof of concept. There are legal grounds for claiming that smartphones may be more like *a part of us* and less like *mere gadgets*.

This detour has hopefully served two purposes. First, to illustrate the potential practical import that debates over extended cognition might have. Second, to show how there is a need for something like the strong version of Levy's Ethical Parity Principle. Armed with principled criteria, it does look like there could be important cases where the framework of extending captures features that would be overlooked or underappreciated by the framework of embedding. Insofar as Levy's arguments proceed from EPP Weak then, they risk underselling the force of extended cognition arguments and their application to cases like those of extended assault. That is, Levy may be too quick to grant too much to the embedded mind critics of the extended mind. And in doing so, he may undercut his efforts to emphasize the ethical import of extended cognition.

This second shortcoming is due, at least in part, to the lack of more fine-grained approaches to talking about the various dependence relationships between agents and environments. As I suggested in Chapter II, Sterelny's framework of scaffolding provides a nice middle ground between weaker claims of embedding and the stronger claims of extending. Such a middle-path approach might allow Levy to retain what is distinctive

and normatively important about extended mind-style arguments without entirely alienating proponents of more conservative embedded mind-style arguments.

A final remark about Levy's treatment is that while the introduction promises "an important alteration in the way ethics is understood, and in what we come to see as the bearers of moral values," (2007b, p. xi) this promise is unfulfilled by the end of the book. Levy does provide a number of compelling arguments against the normative significance of the inner/outer distinction. But comparatively much less is said about what this means for the metaphysics of the bearers of moral value. Are skin and skull individuals still the proper targets of the reactive attitudes? Or is there a sense in which assemblies of agents and integrated features of the environment are the targets of virtue and vice attributions? Levy has helped to raise this question and has provided a guide to thinking about it, but as it stands, the question: "what are the bearers of moral values in light of the extended mind?" remains unanswered.

IV.IV Extending autonomy, identity, and the self

I concluded the previous section by noting that while Levy raised the issue of the bearers of moral values, many questions remain unanswered about the underlying metaphysics of such an extended cognition inspired view. In the present section, I will review some of the work that takes steps toward addressing these and related issues. Though I will consider a number of different sources, the theme that unifies this section is the question of whether extended minds entail extended selves, broadly construed.

In a series of papers, Mason Cash (2010, 2013), like Sneddon (2011), casts extended mind-style externalism as the antithesis of individualism about the mind. In this light, Cash identifies parallels with some feminist critiques of individualism and

autonomy (citing e.g., Friedman (1997, 2000); Meyers (1997, 1998); Mackenzie & Stoljar (2000)). Cash argues that feminist theories that highlight the relationality of the self and autonomy are structurally similar to extended mind theorists' arguments for our cognitive reliance and dependence on external scaffolds. Setting to one side questions about whether this analogy is on the right track, I do think Cash begins by asking the right kinds of general questions regarding the ethical implications of the extended mind:

Are the requirements for being a moral agent - an agent that autonomously governs its actions based, in part, on reflection on its goals and values, and on a conception of what it ought to do - affected by this [the extended mind's] hybrid version of cognition and action? Could a decentralized and distributed 'extended' system that controls action itself be a moral agent? Does the influence of these external factors affect the moral agency of the individuals involved in such situations? Furthermore, could taking seriously HEC's [Hypothesis of Extended Cognition] hybrid conception of agency reveal problems with our practice of concentrating moral responsibility on individual persons? Can entities wider than individual persons bear moral responsibility in addition to or instead of the responsibility of the individuals involved?" (Cash 2010, p. 646).

These are wide-ranging questions and the value of Cash's contribution has less to do with any specific answers to them and more to do with a general strategy for approaching them. The general strategy is to identify and call into question the individualism that may still be lurking in the background of many formulations of the extended mind and cognition.

Cash correctly identifies cognitive bloat as one of the strongest objections to the extended mind. One of his contributions to the literature is the recognition that many rebuttals to the cognitive bloat objection seem to double down on a kind of individualism that may ultimately undermine the full thrust of externalist arguments, broadly construed. On Cash's reading, extended cognition theorists tend to respond to the cognitive bloat objection by attempting to formulate more precise criteria which specify just how far the individual extends, but they never call into question the individual as the proper starting point from which various processes may extend. Even the most principled formulations of extended cognition, according to Cash, are still involved in this game of line drawing.

Cash's proposal is that we stop playing that game. Instead, we should reject the premise "that we start with discrete individuals and then define conditions under which their cognitive systems are extended" (2013, p. 66). Instead of policing the boundary lines of cognition, Cash argues we should "*start with socially distributed cognitive processes and then give reasons to focus upon the particular contributions of individuals*" (ibid, emphasis mine).

At first blush, it is difficult to know just how far this line of thought should be taken. It is surely too demanding to suggest that we must simply jettison the idea of discrete individuals all together, once and for all. As Cash rightly notes, there are plenty of pragmatic reasons to not do so, namely our practices of holding individuals responsible for their actions. But on the other hand, fully appreciating Clark & Chalmers's insight that there is nothing sacred about the skin and skull boundaries may require a more thoroughgoing anti-individualism than is usually recognized. To my mind, Cash's most important contribution is the recognition that indeed, extended cognition arguments *do*

require more anti-individualist commitments than are usually supposed, *and* there is a vast body of literature that has already worked out many of these commitments:

This view of cognition as socially distributed, with its critique of overly individualistic conceptions of intelligence, has some important and revealing parallels with many— principally feminist—critiques of individualistic conceptions of self and of moral autonomy...We do not have to reinvent the wheel here. This socially distributed ‘extended’ conception of cognition, by recognizing an alliance with socially distributed ‘relational’ conceptions of self and autonomy, can avail itself of extant arguments in relational theory that address this [cognitive bloot] objection far more thoroughly and effectively (2013, p. 66).

While it is certainly true that jettisoning the view of discrete individuals is a big ask, we are not without assistance in such an endeavor. The question - on which the success of Cash’s arguments hinges - is whether the analogy between feminist critiques of classical liberal individualism and the critiques of internalist accounts of cognition is a good one. Unfortunately, Cash does not so much argue for this connection as assert it. Nonetheless, I think he is on the right track and so it is worthwhile to take a closer look at how this analogy might unfold.

Feminist critiques of individualism constitute a rich tradition in their own right, and I cannot pretend to do this tradition justice in the context of the present section. But insofar as such a tradition can be distilled into a few paragraphs, it is worth noting the similarities on which the argument from might analogy rest. By Cash’s lights, the “relational view of autonomy critiques the ‘liberal’ view of an autonomous moral agent,

conceived as a Cartesian mind or an unencumbered individual who thinks and acts independently from any external influence” (2013, p. 66).

The historical touchstone for this view is Hobbes’s (1658/1991) mushroom metaphor: “Let us return again to the state of nature, and consider men as if but even now sprung out of the earth, and suddenly, like mushrooms, come to full maturity without all kind of engagement to each other” (p. 205). Individuals here are prior to any form of social interaction. To act autonomously is to act freely from these outside influences. A classic feminist critique of this metaphor is offered by Seyla Benhabib: “This vision of men as mushrooms is an ultimate picture of autonomy. The female, the mother of whom every individual is born, is now replaced by the earth. The denial of being born of woman frees the male ego from the most natural and basic bond of dependence” (1986, p. 408).

At the heart of a feminist-inspired conception of relational autonomy, then, is the recognition that as a matter of biological fact, humans do not come into the world as fully formed mushrooms. Instead, as Eva Kittay (1999) famously notes, we are all some mother’s child. We come into the world as vulnerable, dependent beings. The crucial insight for present purposes is that *it is only because of our relationships that we are able to develop the capacities traditionally associated with autonomy in the first place*. Individuality is not prior to social interaction, it is accomplished and achieved as a result of it.

The analogy that Cash wishes to draw here is that in the same way feminists might argue that we should *start* from social interactions, and (biological) dependence in order to derive individuality, so too, should externalists *start* from social interactions and (epistemic) dependence to derive cognition. Returning to our original context of rebuttals

to the cognitive bloat charge, Cash thus writes, “rather than starting with individuals and extending their cognitive systems outward, we start with socially distributed cognitive practices and institutions, and give reasons for focusing on individuals’ contributions to the cognitive milieu” (2013, p. 67).

Again, it is hard to know just how much conceptual revision would be required to implement this proposal. I suspect quite a lot, and quite a lot more than Cash has led on. Nonetheless, Cash does offer a glimpse of a positive view:

My ability to act intelligently is technologically, environmentally and socially enhanced and scaffolded by critical engagement with—and attempts to improve—myself, my cognitive tools, and the cognitive institutions, norms, and practices of my community. An intelligent agent is one who critically engages the cognitive tools around them: selecting, using and reproducing effective cognitive tools; replacing, refining or augmenting less effective cognitive tools; selectively incorporating these social, relational, technological, environmental and bodily resources into their sense of who they are, what they know, what they want, and what they can do. And many such intelligent agents critically engage the tools, practices, and institutions around them by shaping and promoting more effective tools, practices, and institutions for others to use. Such a critical engagement with this social, normative and institutional milieu helps improve the cognitive milieu that influences and supports the cognitive development and cognitive activities of every one with access to such resources” (Cash, 2013, p. 67).

While I think this account, based on a parallel between extended cognition and relational autonomy is promising, I am also struck by the lack of Cash’s engagement with what is

perhaps an even more fundamental concept in feminist philosophy, namely *vulnerability* (e.g., Mackenzie, Rogers, & Dodds, 2014; Straehle, 2016). In order for the arguments for relational autonomy to get off the ground, the essential vulnerability of agents must first be recognized. After all, we must be open to the influence of our social environment (for good or ill) if it is to shape us in the ways relational theories specify.

Perhaps a case can be made that this recognition of vulnerability to environmental influence is similarly required to get extended cognition arguments off the ground. If this line of thought is worth pursuing further, then Cash has done important work of making the first step in connecting these two largely disparate bodies of literature.³¹ In the end, Cash's treatment raises more questions than it answers, but I have tried to show that the analogy between extended mind arguments and relational autonomy arguments is nonetheless worth pursuing further.

IV.V Extending self, identity, and the bearers of psychological states

At the end of "The Extended Mind" Clark & Chalmers mention in passing that it is "far better to take the broader view, and see agents themselves as spread into the

³¹ Though I will not pursue it in these pages, I think feminist accounts of vulnerability are especially well-equipped to wrangle with what might be seen as some of the 'darker' implications of the extended mind. A rhetorical tone throughout much of the situated cognition literature is that the kind of offloading characteristic of embedded/scaffolded/extended cognition is (evolutionarily) beneficial because it reduces internal processing demands, thereby freeing up cognitive resources for more grandiose pursuits. Clark's (1989, p. 64) formulation of the "007 principle" is paradigmatic: "in general, evolved creatures will neither store nor process information in costly ways when they can use the structure of their environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done." Similar sentiments can be found in Rowlands's (2003) "barking dog principle," and Dawkins's (1999) account of extended phenotypes. Less often discussed is the way in which such offloading also renders agents vulnerable to harm, manipulation, or the exacerbation of extant inequalities. I have argued elsewhere, for example, that "on the one hand, scaffolds can reduce the internal processing demands on cognitive agents while increasing their access to information. On the other hand, the use of scaffolding leaves cognitive agents increasingly vulnerable to forming false beliefs or failing to form beliefs at all about particular topics" (Alfano & Skorburg, forthcoming). In a somewhat related vein, Laurin et al. (2016) examine how power dynamics influence the offloading of self-regulatory resources in goal pursuit. For the most part, however, these kinds of considerations are absent from the extended mind and cognition literatures.

world” (1998, p. 18, emphasis mine). This has generated a number of responses in the literature about whether or not the extended mind entails the extended self or extended personhood, and whether such conceptions are desirable or defensible. It is within this context that the project closest to my own has emerged. My central focus in this section will be Robert J. Howell’s (2016) article, “Extended virtues and the boundaries of persons”. As before, I will highlight what I take to be important strengths and shortcomings, and this will both set up an engagement with other treatments of the extended self, and set up my claims in the next two chapters.

I must begin by noting that there is certainly a sense in which Howell has anticipated the arguments of the present project. After all, the titles are the same. And indeed, Howell explicitly raises some of the very questions taken up in this dissertation: Though it is clearly important to analyze what it takes for a person to have a virtue, focusing on that question tends to eclipse even more important questions about what persons must be if they are to have virtues, vices, and characters. What sorts of things can legitimately undergird a person’s character? And what does the answer tell us about the person’s relationship to her body, her environment, and the people who surround her? For the purpose of ascribing virtue, vice, and character—and thus certain forms of praise and blame—what are the boundaries of the person? (p. 147)

Given these similarities, it is important to carefully examine Howell’s arguments so they can be distinguished from my own. To begin with the conclusion, Howell argues that “skindividualism” – the widespread view that persons are contained within the boundaries of their skin – is false. From there, Howell argues that “virtues, vices, and other character traits can be extended, in the sense that the grounds for these traits might

not be within that individual's skin. If this is the case, persons are likewise extended in that persons can have constituents that are not within the skin. Persons are not skindividuals" (pp. 147-48).

I will argue that due to a lack of engagement with existing literature, and lack of attention to important details, Howell fails to make a convincing case for extended virtues. Given this latter claim, of course, we will have to examine the arguments in more detail. But insofar as my argument holds water, my work will be cut out for me in the balance of the dissertation.

Howell begins with a brief review of Clark & Chalmers (1998) argument for the extended mind, but his aim is not to vindicate any particular formulation of it. Instead, the original extended mind argument simply suggests a strategy for arguing that persons are extended beyond the boundaries of the skin. This is important to note, because despite the "extended virtues" in the paper's title, the central argument is actually about extending the boundaries of persons, and it is only secondarily because persons are generally thought to be the bearers of virtues that the formulation of extended virtues arises at all. Given this framing, then, it is important that Howell also marks the disanalogies between arguments for the extended mind on the one hand, and arguments for extended persons on the other.

Howell goes so far as to say that it is actually *better* if the case for extended persons does not depend on the case for the extended mind because then "arguments against extended minds aren't necessarily arguments against extended persons and extended virtues" (p. 151). While this is trivially true (indeed minds are not persons), I will argue below that this separation between the arguments for extended cognition and

extended persons presents serious problems for Howell's argument. But first, we should get the whole argument on the table. On the basis of pulling apart his argument for extended persons from the literature on extended cognition, Howell writes:

since the individuation criteria for persons are apt to be quite different from those for minds, the debates are different. The analogous argument for extended persons will open up a metaphysical possibility that skindividualism is false. I argue that the results of experiments in situationist psychology suggest that if we are to have virtues, they must in fact be extended; situationism speaks against skindividualism (p. 148).

Rather than engaging the literature in extended mind and cognition to mount his argument against skindividualism, Howell, like Sneddon (though without referencing Sneddon), turns to situationist social psychology. But where Sneddon at least acknowledges the history of the person-situation debate as a debate, Howell mostly summarizes other philosophers' summaries of the social psychological research to assert the truth of situationism. The reasons I gave above in Section IV.II for doubting the strength of the situationist premises in Sneddon's argument go doubly here, as Howell's reading is even more superficial. It seems rather straightforward, then, on the basis of the evidence presented in Chapter III, to reject Howell's claim that "if one is looking for a predictor of behavior, one does much better to look at the context of action than at the dispositions of the agent" (p. 152).

This is a potentially devastating blow because by Howell's lights, it is not a formulation of extended cognition which grounds the argument for extended persons. Instead, "the situationist challenge thus offers us an argument for extended virtues and

against skindividualism” (p. 156). Thus, insofar as the argument against skindividualism (and ostensibly also the argument for extended virtues) is premised solely on the truth of situationism, the argument will fail. As I have repeatedly observed, the empirical evidence shows that personality variables and situational variables account for roughly the same amount of variance in behavior. A cursory reading of research in social and personality psychology from the last 30 years reveals that interactionism, not situationism, enjoys the most empirical support and widespread acceptance among psychologists. Philosophers should take note.

Despite this flaw in the argument, we should nonetheless press on to see how, on the basis of situationism, Howell attempts to formulate an account of extended persons. We will have to conditionally assume for the sake of argument that the situationist literature shows what Howell thinks it does, namely, that social psychology calls into question the existence of stable character traits. With this qualification in place, Howell’s move is still quite clever. He notes that:

We are deeply wedded to the claim that people can be evaluated based upon their character traits. Some people are virtuous; others are vicious. What social psychology shows us is that these character traits do not exist in skindividuals. But this shouldn’t lead us to abandon talk of character traits; it should lead us to recognize that people are not skindividuals (p. 158).

It is almost certainly true that most people are committed to the existence of traits, character, personality, virtues, etc. So if, for the sake of argument, we assume Howell’s (almost certainly false) premise that social psychology proves the non-existence of these traits, then extending the boundaries of persons seems like a plausible and creative way to

save these concepts we hold dear. Of course, I have given reasons to doubt that the social psychological evidence shows what Howell thinks it does, and so it looks like we have a solution to a non-problem. This is not to say that the extended persons conclusion is false or unimportant. Rather, it is to say that it is arrived at in the wrong way. In fact, to foreshadow my own argument, I believe that we *do* need an account of something like extended persons (albeit a *much* more careful and detailed formulation than is offered here) to develop an empirically responsible view of virtue.

Returning to Howell's argument, the next question to ask is: if traits, character, personality, virtues, etc. are indeed salvaged by extending the boundaries of persons, what exactly does this look like? Howell writes:

what we are looking for is a set of criteria in virtue of which a ground of a behavioral disposition can be considered part of the person. We can be guided in this rather abstract quest by the more intuitive question of when a behavioral disposition should be called a part of a person's character (p. 159).

That is, once we figure out the answers to the question, "what makes a disposition part of a person's character?" we can then examine whether such components are necessarily skindividualistic. To answer this first question, Howell turns to Christian Miller, whose work I will also consider in detail in the next chapter. According to Miller, character traits are dispositions, and they serve a number of functional roles. Howell highlights three:

One role traits play is explanatory—we can appeal to them to explain why people do what they do. Another is predictive—we can predict with some reliability what

people will do based on their character traits. A third is evaluative—we use character traits to normatively evaluate a person (Howell, p. 159).

In line with widely accepted functionalist approaches in the philosophy of mind, the idea here is that if a given disposition does not play these functional roles, then it is probably not a trait. It does not much matter *where* the dispositions are grounded. What matters is their functional profile. And on the other side, if a disposition does play these functional roles, then it probably is a trait and it does not much matter whether the disposition is grounded inside or outside the skin. To illustrate this, Howell considers the disposition to be on time:

Consider, for example, the ‘virtue’ of punctuality. If I am punctual, this is likely to owe quite a bit to the fact that I have a watch. The watch is located outside of me, but if I am punctual it is in part because of the thing on my wrist. Of course, there are things located inside me that are crucial to my punctuality, such as the desire to be on time. But the virtue of punctuality requires more than that. It requires that I have the disposition to realize that desire. This disposition is partly located in my watch, which is outside of the skin but always with me. (p. 150).

While it is not clear that punctuality is a virtue, at least insofar as virtues have a moral valence (am I a morally bad person because I showed up late?), the suggestion is still provocative and worthy of consideration. In this case it seems correct to say that, from a functionalist perspective, it is only the manifestation of internalist prejudice that would prevent the inclusion of the watch as a ground of the disposition of punctuality. From here, the argument is that we could presumably identify other relevantly similar dispositions. Then, armed with the assumptions that character is a collection of such

dispositions, and that persons are the bearers of character, we can derive the conclusion that in these cases, persons extend beyond the boundaries of the skin. Skindividualism is false and traits are saved.

With the argument now in place, what to make of it? In the first place, we might want to resist the conclusion simply because of its reliance on questionable situationist premises. Fair enough. I have already developed and deployed that argument, however, and so in the balance of this section, I will consider two additional shortcomings of Howell's argument. Both are based on Howell's lack of serious engagement with the extended cognition literature.

The first and most glaring shortcoming is that a number of thinkers have already considered whether arguments for extended cognition entail arguments for extended personhood, self, identity, etc. For example, Eric Olson (2011) argues that the extended self is not entailed by the argument for the extended mind. Lynne Rudder Baker (2009) similarly argues against the idea of extended selves or persons, but on the grounds that the sorts of non-biological couplings usually invoked in service of extended mind arguments are integrated at the sub-personal level. One might expect that in making an argument that persons are extended, Howell would attempt to rebut the extant arguments suggesting persons are not extended. Howell references neither.

Tom Buller (2013) argues that the extended mind is perfectly consistent with the ethical significance of the skin and skull barrier. Rob Wilson and Bartlomeij Lenart (2015) argue that the self is not extended, but drawing on the work of Hilde Lindeman (2010, 2013) and other narrative theorists, they make a case that personal identity is

extended.³² Chris Abel (2014) has written a massive and wide-ranging book entitled *The Extended Self* which takes up extended mind arguments in the context of cultural evolutionary theory. Howell cites none of them. Nor are any of the above treatments by Levy or Sneddon referenced.

The less charitable interpretation is that this constitutes intellectual irresponsibility. More charitably, this could be due to the strategy of pulling apart the arguments for the extended mind from the arguments for extended persons. But as we will see now, this pulling apart leads to even more serious problems with the argument.

By making his case via situationism apart from the extended cognition literature, Howell has not only imperiled his positive argument for extended persons and extended virtues in the ways noted above, he also failed to substantively engage with the most likely objections to his argument. Only a few paragraphs are dedicated to the coupling-constitution fallacy, and here too, situationism shoulders the burden:

We can take the situationist critique as essentially making the point that the skindividualist way of cutting things up does not jibe with personality and trait talk. The extended persons approach does much better on this score. This suggests that in the case of the extensions of persons, we have reason to speak of constitution instead of mere coupling (p. 158).

Even if there were reason to believe the situationist critique could do the work Howell requires, he still has not offered any independent reasons to adopt the language of

³² Their arguments open yet another avenue for engagement with feminist philosophy in the manner suggested above: “The externalist account of personal identity thus reveals a fragile side of personhood that remains hidden in individualistic variants of the neo-Lockean approach. Understanding the sensitivity of narratives to interpretative interventions deepens our understanding of what it means to treat people and their identities authentically” (Wilson & Lenart 2015, p. 436).

constitution instead of coupling. Instead, Howell offers vagaries about persons and virtues *depending* on various contextual features, or being *undergirded* by external resources. Nowhere does Howell approach anything resembling the specificity of the criteria from the extended cognition literature outlined in Chapter II.

To return to the example of watches and punctuality, Howell provides no argument that the watch extends - as in literally constitutes - the disposition of punctuality, rather than merely embedding it. In fact, based on the criteria I developed, the watch looks to be a paradigm case of embedding.³³ This is but one example where more attention to the detail provided in the literature would strengthen the argument and clarify its scope.

Another, and I think more damning example, is that nowhere does Howell take up the issue of cognitive bloat. The analogous charge would be that Howell does not provide any criteria for determining where persons stop. That is, he has offered no principled way to say why watches count, but the entire internet does not. Put yet another way, it does not appear that Howell has any answer to Rupert's challenge: why bother with the counterintuitive view that persons extend beyond the skin when a more intuitive, traditional view will do? "Sure, your watch helps you know what time it is, but the *real* processes still only happen between the ears, even if they do rely on watches and other gadgets in interesting ways." By not considering any of these objections, Howell has left his formulation open to serious counterarguments. And if the extended cognition literature is any guide, many of these counterarguments will be decisive. But by the same token, the mature and principled formulations and distinctions from the extended

³³ Though perhaps a case could be made that smartwatches with access to push notifications, calendars, e-mail, etc. are actually examples of scaffolding or extending. Regardless, this issue is not treated in any detail by Howell.

cognition literature may well vindicate aspects of the arguments for extended persons, and subsequently extended virtues. It is unfortunate that Howell has not availed himself of these resources.

In the end, Howell's article is of a piece with the other work considered in this chapter: promising directions for future research, but unsatisfying as it stands. While I have raised a number of objections to Howell's treatment, this is not in the service of arguing for skindividualism. Instead, I have suggested that insofar as one is interested in developing an account of externalist moral psychology, extended persons, or extended virtues, one should be cautious about building these arguments on situationist foundations. In Howell's case specifically, the argument is imperiled not only by the lack of serious engagement with the psychological literature, but also the lack of engagement with the philosophical literature. So while I agree with Howell's conclusion that skindividualism is false, I will aim to provide stronger premises to support it in what follows.

Conclusion

In this chapter, I have argued that the work of Sneddon, Levy, Cash, Howell, and others have made important strides in recognizing the ethical import of extended cognition-style arguments. More specifically, Sneddon has made a convincing case that individualism need not be the only moral psychological game in town. Levy has helped to raise difficult questions about the bearers of moral virtues and the ethical significance of the inner/outer distinction. Cash has done especially important work in bringing together the largely disparate literatures in feminist philosophy and extended cognition. Howell has broached the central topic I am interested in here, namely, extended virtues.

In all of these cases, however, much work remains. My point is that there is a near total lack of substantive engagement between extended cognition and normative ethical theories such as virtue ethics.

Thus, in the next chapter I will take a step toward addressing this larger issue by showing how the arguments developed in Chapters II and III, and continued in the present chapter, raise serious challenges for widely accepted theories of virtue. Then, in the final chapter I will draw on the resources developed from the extended cognition literature to formulate and defend a novel account of extended virtues which aims to address the shortcomings identified in the present chapter and also respond in a satisfactory and empirically responsible way to the challenge formulated in Chapter V.

CHAPTER V

VIRTUE ETHICS

V.I. Introduction

In the first three chapters, I argued that the views presented therein have significant but heretofore underappreciated implications for virtue ethics. In Chapter IV, I tried to show that while some small steps have been taken to spell out these ethical implications, the efforts are deeply unsatisfactory and cursory. The takeaway was that extended mind and cognition theorists do not substantively engage their work with normative ethical theories such as virtue ethics.

This chapter develops a criticism from the opposite direction: Despite a concern among some virtue ethicists with developing theories of virtue consistent with research from the social and cognitive sciences, most empirically minded virtue theorists have not had much to say about the research programs outlined in Part I. But is this due to benign neglect or a deeper incompatibility between externalism and virtue ethics? I will suggest the latter is the case. That is, I will try to show how much contemporary virtue theorizing is internalist and individualist in precisely the ways that the arguments from Part I identify as inadequate and problematic. The present chapter, then, serves as a brief review of the relevant virtue ethics literature whose shortcomings will motivate the need for an account of extended virtues in the next chapter.

At first pass, here is the sort of thing I have in mind when I worry that virtue ethics is overly internalistic and individualistic:

In virtue ethics, the focus is on the virtuous individual and on those inner traits, dispositions, and motives that qualify her as being virtuous. (Some forms of virtue

ethics do allow for general moral rules or even laws, but these are typically treated as derivative or secondary factors.) Many modern philosophers think of the moral life as a matter of relating properly to moral rules, but in the virtue ethics of the ancient world and in those few instances of virtue ethics one finds in modern or recent philosophy, the understanding of the moral or ethical life primarily requires us to understand what it is to be a virtuous individual and/or what it is to have one or another particular virtue, conceived as an inner trait or disposition of the individual. So the first thing we can say about virtue ethics in an attempt to distinguish it from other approaches is that it is agent-focused (Slote, 2003, p. 4).

Insofar as one can summarize an ethical theory in a single paragraph, I think Slote's is fairly representative. My claim, to be unfolded over the course of this chapter, is that the usual operationalization of the concepts of "individuals," "agents," "inner traits," and "dispositions" in this passage and many others like it render virtue ethics open to the full force of the challenges from the externalist arguments I have developed in Part I.

V.II. What are virtues?

The first step in my argument is to survey the recent virtue ethics literature for answers to the question: "What are virtues?" I will provide abundant evidence to suggest that the dominant view is that virtues are dispositions. I will then take a brief detour into the metaphysics of dispositions to get a better grip on precisely what one is committing to when one says that virtues are dispositions. To begin, consider the following passages from a veritable who's who list of recent virtue theorists:

Bernard Williams: “A virtue, being a disposition of character to choose or reject actions because they are of a certain ethically relevant kind. The word ‘virtue’ has for the most part acquired comic or otherwise undesirable associations, and few now use it except philosophers, but there is no other word that serves as well, and it has to be used in moral philosophy. One might hope that, with its proper meaning reestablished, it will come back into respectable use. In that proper use, meaning an ethically admirable disposition of character, it covers a broad class of characteristics” (1985, pp. 8-9).

Owen Flanagan: “Virtues are psychological dispositions productive of behavior” (1991, p. 282).

Julia Annas: “In some ways Aristotle's word *hexis* answers better to our word ‘disposition’ than to ‘state’: a virtue like courage is a disposition because it is a condition because of which I am so disposed as to act in brave ways; and this is what a *hexis* is...Aristotle's point, that virtue is what we would call a stable disposition, is taken from common conceptions of virtue and regarded as uncontroversial by later schools” (1993, p. 50).

Rosalind Hursthouse: “The virtues (and vices) are all dispositions not only to act, but to feel emotions, as reactions as well as impulses to action” (1999, p. 108)

Christine Swanton: “I begin with a definition of virtue that, I shall claim, is shared not only by forms of virtue ethics that differ from my own (including eudaimonism) but by those virtue theorists who are not virtue ethicists. A virtue is a good quality of character, more specifically a disposition to respond to, or acknowledge, items within its field or fields in an excellent or good enough way. This definition of virtue is

intended to be neutral with respect to a variety of virtue theories and virtue ethics: pluralistic and monistic, eudaimonistic and non-eudaimonistic” (2003, p. 19).

Alasdair MacIntyre: “The virtues therefore are to be understood as those dispositions which will not only sustain practices and enable us to achieve the goods internal to those practices, but which will also sustain us in the relevant kind of quest for the good, by enabling us to overcome the harms, dangers, temptations, and distractions which we encounter, and which will furnish us with increasing self-knowledge and increasing knowledge of the good” (2007, p. 219).

Daniel Russell: “The character virtues are stable dispositions to react emotionally at the appropriate times, about the appropriate things, to an appropriate degree, and so on, and to desire and aim at the appropriate kinds of ends or targets” (2009, pp. 13-4).³⁴

I hasten to add that this brief sampling of the literature is just that: brief and a sample. But it should hopefully be clear from the above that beginning our inquiry from the view of virtues as dispositions is not untoward. To be sure, there are subtle and important differences between many of these thinkers and their conceptions of virtue. There are also different ways of conceptualizing the virtues (for example, as skills, habits or as qualities of a person). And of course, I have not said anything about the role of practical reasoning with respect to these dispositions. Nor do I mean to suggest that virtues are nothing but dispositions. Instead, I simply want to show that across many

³⁴ For an even more extensive list of the virtues as dispositions view, see Doris (1998, p. 509, n. 20; 2002, p. 174, n. 2). For a more extensive list still, see Miller (2013, p. 7, n. 12). It would be needlessly pedantic to reproduce their sources, as both Doris’ and Miller’s views will be treated in detail below. But if my contemporary references here, in addition to the forty or so collected by Doris and Miller, in addition to Annas’ contention that the virtue as disposition view is uncontroversial among various ancient traditions does not establish the virtues as dispositions view as dominant, I do not know what will.

different persuasions, it is highly plausible that “disposition” is the standard answer to the question, “what are virtues?”

Perhaps it could be objected that the above statements should not be taken so literally - as asserting robust metaphysical theses - or that these formulations are somehow taken out of context. I contend, however, that all other things equal, we ought to take authors at their word when they say what virtues *are*. Further, as we will see below, the dispositional view of virtues is present across historical and contemporary sources. So while it is certainly possible to object to the virtues as dispositions view, these objections will have to explain why virtue theorists of different stripes all seem to converge on their formulations of virtues as dispositions. So long as the initial plausibility of the virtues as dispositions view is secured, I will turn my attention to more systematic and nuanced treatments of the structure of virtues.

In her (1993) *The Morality of Happiness*, Julia Annas lays out a general framework for her thinking about ancient theories of virtue. I find this framework particularly useful for two reasons. First it is general enough (by Annas’ own lights, anyway) to cover the varieties of ancient virtue theories (e.g., Aristotelean, Stoic, Cyrenaic, Epicurean, Sceptic). It thus compliments my own more contemporary approach with some historical depth. Second, the framework is also general enough that many of its details can be filled in by the empirical sources I have been engaging with so far.

What, then, is the structure of virtue? According to Annas:

In ancient ethical theory considerable attention was paid to three points: 1. Virtues are dispositional. 2. Virtues have an affective aspect: they involve our feelings, especially our feelings of pleasure and pain, and developing a virtue involves

habituating our feelings in certain ways. 3. Virtues have an intellectual aspect: they involve reasoning about, and grasp of, the right thing to do, and developed virtue implies good practical reasoning or practical intelligence. The development of all three points contributes to our understanding of what a virtue is (pp. 48-9).

Given the wide-ranging historical scope of Annas's project, her claim that, for the ancients, virtues are first and foremost dispositional, tells strongly in favor of my claim for the primacy of this view in the literature. In much of what follows, I will adopt this tripartite structure as a template for thinking about the metaphysics of virtue. The bare claim that virtues are dispositions to act, think and feel, however, merely kicks the can down the road. What are the relevant kinds of dispositions? And what is one committed to when one commits to a dispositional view of virtues? To get some traction on these questions, I turn now to the work of Christian Miller, one of the foremost empirically minded virtue theorists. For my money, Miller (2013, 2014) has provided the most sophisticated analysis of what it means to say that virtues are dispositions.

People have traits such as height, weight, hair color, average cortisol levels, extraversion, conscientiousness, honesty, shyness, wittiness, and compassion. Among these, Miller distinguishes between personality traits (like extraversion and compassion) and non-personality traits (like height and weight). A personality trait is "a disposition to form beliefs and/or desires of a certain sort and (in many cases to act in a certain way, when in conditions relevant to that disposition" (Miller 2014, p. 6). Miller also approvingly cites Brent Roberts' definition: "Personality traits are the relatively enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances" (2009, p. 140).

There is a clear continuity here with Annas's account of the structure of virtue, in that personality traits also involve dispositions to act, think, and feel in characteristic ways. These traits are, among other things, the subject matter of personality psychology, and much of Miller's work in virtue ethics is grounded in and informed by the findings in personality science. While one might quibble with the distinction between personality and non-personality traits (and what falls within the purview of personality psychology), this is not a particularly important distinction for our virtue theoretic purposes.

The more important distinction is within the domain of personality traits (however defined), and that is between character traits and non-character traits. Here, it is worthwhile to slow down a bit and further elucidate what a character trait is, and what it does. Miller notes that character traits serve (at least) five different kinds of functions.

First, character traits help us to *understand* ourselves and others. More specifically, they help us to *explain* and *predict* behavior. We can explain why Sam does not want news of her latest achievement announced far and wide in terms of Sam being a modest and humble person. We could similarly have predicted that if Sam won an award, she would not want it broadcast because she possesses the character traits of modesty and humility - that is, she is a modest and humble person.

Character traits are also the basis for *normative evaluations*. When we say that Dan is cruel, we are not only describing him as thinking and behaving in certain ways, we are also negatively evaluating his character in this regard. Lastly, character traits provide a basis for *imitation*. If I want to become more compassionate, I might well try to model my thinking and behavior after those aspects of the Dalai Lama and his teachings which are particularly compassionate.

According to Miller, “the idea is that character traits are dispositions to form beliefs and/or desires and potentially to act in relevant ways as well when in the appropriate circumstances... a character trait is a disposition of a certain kind, or a ‘trait disposition’ as I will call it” (2013, p. 7). This much, however, is equally true of personality traits. The defining feature of a character trait is that it is the kind of thing people are *held responsible for*. Miller (2014) thus writes:

A character trait is a personality trait for which a person who possesses it is (at least to some degree) normatively responsible for doing so... a person is responsible for possessing a trait, in turn, only if the person is - in this respect - an appropriate candidate for the positive reactive attitudes (praise, admiration, respect, gratitude, congratulations, etc.), the negative ones (blame, condemnation, resentment, indignation, disapproval, etc.), or some combination of the two (p. 11).³⁵

Where it would not make sense to blame someone for having brown eyes or short fingers, it does make sense to blame them for being dishonest or praise them for being courageous. Building up from this definition of character traits as those personality traits for which we are held responsible, character is then understood as the sum of character traits and their interrelations. And indeed, talk of responsibility, praise, and blame brings us in contact with the familiar territory of virtue ethics as distinctively characterological. But we are still left with the question of what makes character traits distinctively moral.

³⁵ See also Zagzebski (1996, p. 116): “A virtue is an acquired excellence and a vice is an acquired defect. It takes time to develop virtues and vices, and this feature is connected with the fact that we hold persons responsible for these traits. Once a virtue or vice develops, it becomes entrenched in a person's character and becomes a kind of second nature. The fact that a trait is among the more permanent of a person's qualities means that she bears a fuller responsibility for it than she does for qualities that are more fleeting. So a person's responsibility for her virtues and vices is connected with the fact that they are gradually acquired and are relatively permanent, and these two properties of permanence and gradual acquisition are not independent. Virtues and vices form part of what makes a person the person that she is.”

This is a difficult question, but Miller suggests that there are indeed character traits such as artistic, lazy, nervous, logical, and witty which are not distinctively *moral* character traits - at least not in the same way character traits like honesty, generosity, courage, gratitude, and compassion are distinctively moral.

Virtues (and vices), then, are those character traits which fall under the heading of morality. Stated formally: “The virtues are all and only those good traits of character which are such that, other things being equal, when they directly lead to action (whether mental or bodily), the action is (typically) a good action and is performed for the appropriate reasons” (Miller, 2013, p. 24). Setting aside issues about whether this is a robust enough conception of morality, or whether people actually have virtues so construed (Miller thinks most of us do not), we can now at least see how a dispositional view of traits, grounded in the empirical study of personality, can be built up to the widespread idea that virtues are dispositions.

It should be clear from the foregoing that Miller’s account of virtues fits squarely into Annas’s tripartite framework. Virtues are character traits, character traits are a species of personality traits, and Miller endorses the view that “personality traits are the relatively enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances” (Roberts, 2009, p. 140; quoted in Miller 2013, p. 6). Taken together, then, we can see how for Annas as well as Miller, virtues are dispositions to think, feel, and act in characteristic ways in certain contexts. Call this view *dispositionalism* about virtues.

I take the foregoing to have established the claim that “disposition” is a widespread and plausible answer to the question, “what are virtues?” The next question

we have to ask is: What is entailed by dispositionalism? I contend that thinking of virtues as dispositions risks commitment to problematic forms of internalism about thinking and feeling, and problematic individualism about the bearers of virtue. My aim in the balance of this chapter, then, is to make clear the connection among dispositionalism, internalism, and individualism.

V.III. Are dispositions internalist and individualist?

The literature in analytic metaphysics of dispositions is vast and this sub-section cannot pretend to adequately summarize it. The literature in virtue ethics which engages with analytic metaphysics of dispositions is much more manageable (and germane) and so this is where my focus will be.

Upton (2009) and Alfano (2014) provide accounts detailing the relationship between virtues and the metaphysics of dispositions. Upton writes: “The similarities between the fundamental structural features of metaphysical dispositions and traits of character justify both classifying and treating traits as dispositions. Unfortunately, however, ethicists have historically failed to examine both metaphysical dispositions and the treatment of traits as dispositions” (2009, p. 29). So far so good. This all looks to be consistent with Miller’s and Annas’s accounts, and also an accurate diagnosis of the virtue ethics literature.

Upton then goes on to summarize four positions in the debates about the role of background conditions and dispositions - that is, the conditions under which some disposition manifests or fails to manifest. First are *obtaining* accounts where for some disposition *D*, for *D* to manifest, background conditions conducive to its manifestations must *obtain*. Upton takes Smith (1977) to be a paradigmatic example here. Second are

ideal accounts, where the background conditions are such that one could reasonably expect *D* to manifest, given *ideal* stimulus conditions. Upton cites Mumford (1998) as representative of this view. Third are *normal* accounts, such as Malzkorn (2000), which restrict the range of relevant background conditions to those that are *normal* in the actual world. Fourth is Upton's preferred *contextualist* account which "places no restriction on the range of background conditions within which an object could causally mediate between initiation and manifestation event pairs for it appropriately to be attributed a disposition" (Upton 2009, pp. 40-1).

In idiom typical of the subject matter, Upton illustrates this account with a thought experiment involving Earth as we know it, and Gravon, a planet with a gravitational pull a thousand times greater than Earth. On the contextualist account, a woodblock dropped on Earth is not disposed to splinter, where the same woodblock dropped on Gravon would be disposed to splinter. So if the woodblock happens to be located on Earth, then it is Gravon-fragile, but not Earth-fragile.

Luckily, I need not defend this account against charges of circularity or triviality. Instead, I only want to highlight that nowhere in the analysis does Upton call into question the equivalence between metaphysical dispositions like the fragility of woodblocks and the dispositional nature of the virtues. In fact, she claims that "the similarities between the fundamental structural features of metaphysical dispositions and traits of character justify both classifying and treating traits as dispositions" (2009, p. 29).

Alfano (2014) concurs with Upton that virtue ethicists ignore the implications of unreflectively assuming the dispositional nature of virtues at their own peril. Alfano's account, however, is more firmly rooted in the mainstream metaphysical debates. He thus

starts where most do, noting that the standard approach to dispositions is through subjunctive conditional analysis. To say that a woodblock is fragile is to say that it is disposed to break, splinter, etc., if it is dropped, struck, etc. The conditional is true if and only if the woodblock would break, splinter, etc. if it were dropped, struck, etc. It is also widely recognized that dispositions are subject to various kinds of disrupters, such as maskers. So while a woodblock might indeed be disposed to splinter if struck, this disposition can be masked by, say, a protective coating. We need not spell out the other kinds of disrupters recognized in the literature. Instead, it should suffice to note that the more sophisticated and plausible accounts of dispositions usually include a stipulation to the effect that a woodblock is disposed to splinter if struck *and no disrupters (like protective coatings) are present*. Taking all of this into account, Alfano then considers how disposition talk might be translated into virtue talk. For example: “at all nearby undisrupted worlds where she is tempted but has sufficient reason to resist temptation, she thinks, feels, and acts appropriately” (p. 81).

Alfano goes on to reject certain aspects (though not all) of the dispositional view of virtues in an argument I will return to in the next chapter. My only purpose in mentioning it here is to provide some of the detail lacking from Upton’s account, and also to drive home the message that even among virtue ethicists, dispositions are still primarily understood with reference to objects like woodblocks, vases, sugar cubes, and wires, and the conditional analyses thereof. The idea seems to be that once the metaphysicians can work out the finer details of dropping vases on various surfaces, then these analyses can be profitably applied to various contexts in which people help people, resist temptations, cheat, lie, and so on.

Perhaps underlying naturalistic commitments could make sense of this tendency: If dispositions figure heavily into scientific explanations, and if one wants to maintain a continuity with the sciences in moral theorizing, then perhaps disposition talk can provide a link. But if the microphysical properties of vases are the standard way of thinking about dispositions, and dispositions are the standard way to think about virtues, what does this say about the virtues?

I worry that all of this disposition talk brings along some unseemly baggage. Insofar as virtues are equated with metaphysical dispositions like fragility, then virtues will be cast as intrinsic, individualist, and internalist - or monadic, in a word. In articulating this worry, I am following Rob Wilson - who has developed wide-ranging critiques of individualism in the life and cognitive sciences - and applying his arguments to the domain of virtue ethics. For example, Wilson writes:

Many properties in the physical sciences are dispositional: They are tendencies that the objects that have them manifest in certain circumstances or under certain conditions. While the manifestation of dispositional properties may require those circumstances or conditions to obtain, what is often called the categorical base of the disposition surely does not: The base is intrinsic to the bearer of the disposition (Wilson 2004, p. 125).

The categorical base of a disposition is often construed as the physical basis which is causally responsible for the manifestation of some disposition (Prior, Pargetter & Jackson, 1982), and this basis is often construed in terms of intrinsic, microstructural properties. One entailment of thinking of dispositions in terms of intrinsic properties, according to Wilson, is an explanatory commitment to the conjunction of *internal*

richness and *external minimalism*. Roughly: when we want to explain some property, ability or capacity, the internal structures are sufficiently rich for the explanatory task. Contextual or external conditions are only minimally relevant, and take a back seat to the intrinsic properties of the bearer of the disposition in question. This conjunction of internal richness and external minimalism is not called into question by any of the discussions of dispositions in the virtue ethics literature.

Perhaps Upton's contextualist account tips the balance away from internal richness, but not without incurring significant theoretical difficulties. And perhaps the basic conditional analyses of dispositions can be scaled up to complex moral behaviors. I have serious reservations about this, and if my arguments go through here, we will have plenty of reason to reject the equivalence between metaphysical dispositions like fragility and the complex clusters of thoughts, actions, and feelings which comprise virtues.

What I hope to have shown here is that the general theme of dispositions as monadic is at the least a respectable, if not predominant, metaphysical view. I have also tried to show how dispositionalism about virtues has been rather uncritically accepted by virtue ethicists who have considered the underlying metaphysics of their views.

Still, I do not think that calling attention to such general themes at a high level of abstraction decisively secures the links among dispositionalism, internalism, and individualism. It is instead meant as a primer to call attention to the less obvious but no less important ways in which disposition talk may entail certain kinds of internalism and individualism about the bearers of those dispositions. Consideration of two further points will hopefully draw a tighter connection.

The first point: Assuming (as Upton, Miller, and others do) the property of shattering when dropped is sufficiently similar to the property of helping when others are in need risks reproducing (perhaps in novel forms) the stimulus-response view of perception and action dismantled by Dewey and his successors (which was detailed in Chapter II). The second point: While there has been a fair bit of analysis about the nature of *dispositions*, there has been relative silence about the nature of the *thinking* and *feeling* to which bearers are so disposed, and what these bearers are in the first place. This lack of attention risks reproducing default internalism about thinking and feeling of precisely the sort dismantled by Clark and his ilk (also in Chapter II). It also risks reproducing default individualism about the bearers of psychological states of precisely the sort called into question in Chapter III.

With respect to the first point, recall the challenge posed by Sterelny-style scaffolding to Rupert-style embedding. The former might charge that the latter holds a view of organisms as passive receptacles of environmental influence, rather than active sculptors of environmental niches. I think the same kind of criticism can be leveraged against dispositionalist view of virtues. If our paradigm example is dropping vases on the floor, then it looks like we have already foreclosed the possibility that the bearer of the disposition might exert some influence over the context and conditions it inhabits. Vases, after all, cannot alter the surfaces on which they are dropped.

In fact, we will probably have foreclosed the much more moderate view that the bearer of the disposition depends on, and is embedded with respect to certain contextual features. The extended view would, of course, be beyond the pale. If this whole range of externalist positions is unavailable to the dispositionalist, then the worry is that

dispositionalism is vulnerable to charges of empirical inadequacy - at least insofar as the externalists enjoy the kind of empirical support which leads Wagman & Chemero (2014) to claim that the debate over the existence of extended cognition is over, owing to the empirical evidence in favor of the view. In the next chapter, I will have more to say about why this charge of empirical inadequacy is so worrisome and problematic.

To be clear, I do not take myself to have shown that dispositionalism logically entails the kind of individualism and internalism critiqued by Dewey, Sterelny, Palermos, and others. Instead, my claim is that in the absence of any qualifications, the default view of dispositions looks much closer to internalism and individualism than the varieties of externalism described in Part I. Thus, insofar as virtue ethicists unreflectively assume a view of virtues as dispositions, they risk cutting themselves off from the externalist insights, which, as I will argue in the next chapter, ought to be at the core of an empirically responsible virtue theory.

The second point is related and, I hope, the most pressing for the empirically minded virtue theorist. By now, there exists decades of scholarship calling for moral philosophers to pay attention to their colleagues' work in the social and cognitive sciences. These calls have been particularly prominent in virtue ethics, all the more after the publication of John Doris's (2002) *Lack of Character*.

In many cases (though certainly not all) the refrain is that virtue theory ought not be incompatible with findings from the social and cognitive sciences. While proponents of these views pride themselves on developing theories in interdisciplinary dialogue, the externalist literature I detailed in Part 1 has been almost entirely absent from these dialogues. Ultimately, my claim here is that the default view of thinking and feeling has

always been, and continues to be such that these processes are bound by the skin and skull. So when, in the absence of any qualifications Miller (2013) says, “as ordinarily understood character traits are personality traits, and are concerned with the *mental life* of a creature, that is, the *mental states and processes that constitute thinking*,” (p. 4, emphasis mine) we are safe to assume that the beliefs he has in mind are not those in Otto’s notebook. *Mutatis mutandis* for claims about the constitution of feeling or affect.

I contend that one cannot make good on a commitment to empirically responsible moral philosophy while making broad claims about mental life which do not engage the vast body of literature which has challenged the internalist and individualist default assumptions about the nature of mental life. Perhaps it will turn out that some brands of virtue ethics will be able to accommodate the externalist insights developed in Part I, or perhaps they will formulate novel arguments to undermine those insights. But insofar as my analysis of the primacy of dispositionalism in contemporary virtue ethics is accurate, and insofar as dispositionalism tends toward internalism (and away from externalism), then unqualified and unreflective endorsement of dispositionalism endorses the internalist and individualist status quo of virtue ethics.

That is, insofar as philosophers like John Doris are convinced that “thinking productively about ethics requires thinking realistically about humanity” (2002, p. 1), then this entails taking seriously the claims that cognitive and affective processing might not be bound by the skin and skull, and that skin and skull individuals might not be the sole bearers of psychological states. Failing to do so is an abdication of empirically responsible moral philosophy.

V.IV. Conclusion

This chapter set out to establish the mainstream answer to the question, what are virtues? I have offered substantial evidence in support of the claim that virtues are generally understood as dispositions. This seems to hold true for both historical accounts of ancient theories of virtue, as well as contemporary accounts informed by the psychological sciences. I then considered what such a view entails. My conclusion was that the virtue theoretic uptake of the metaphysics of dispositions imports the kind of myopic individualism and internalism one would expect from a sub-field whose primary concern seems to be dropping vases on the floor and dissolving sugar cubes.

If my arguments have gone through here, then there are two things to worry about: First, the extent to which empirically minded virtue ethicists have largely ignored the vast body of literature I outlined in Part 1. Second, the extent to which widespread dispositionalism about virtues renders unavailable the externalist insights developed in Part 1. Taken together, these worries call for an account of whether and how empirically minded virtue theorists can incorporate the insights from the extended cognition literature. Providing such an account is the task of the rest of the dissertation.

CHAPTER VI

EXTENDED VIRTUES

VI.I Introduction

The conclusion of the first half of the dissertation was that the views of cognition, emotion, personality, memories, goals, and relationships presented therein have significant but heretofore underappreciated implications for virtue ethics. The second half has aimed to articulate and respond to a resultant two-fold problem. In Chapter IV, I argued that on the whole, extended cognition theorists have not substantively engaged with normative ethical theories such as virtue ethics. In Chapter V, I argued that despite calls for doing empirically informed moral philosophy, virtue ethicists, on the whole, have not substantively engaged with the empirical research surveyed Chapters II and III. In the present chapter, I develop and defend an account of extended virtues in response to these shortcomings.

Here is the plan for this chapter. First, I will make an explicit argument for why an account of extended virtues is needed (Section II). I will then develop my positive account of extended virtues in Sections III through VI. Finally, I conclude by considering possible interpretations of the dissertation's arguments, directions for future work, and a few thoughts about moral agency more generally (Section VII).

VI.II Why bother?

I have argued that a commitment to doing empirically responsible moral philosophy requires a serious engagement with the literature surveyed in Part I: Extended cognition, extended emotion, transactive memory, transactive goal dynamics, and the like. In turn, this will require re-thinking, if not rejecting, some internalist and

individualist assumptions in virtue ethics. I will argue here that my account of extended virtues promises to both uphold the commitment to doing empirically informed moral philosophy, and to avoid the shortcomings previously identified with internalistic and individualistic accounts of virtue.

At the outset, I want to tackle head-on what I take to be the strongest potential objection to my argument. Namely, Rupert's (2009) challenge: *why bother* with complicated, unintuitive, externalist and anti-individualist frameworks when simpler, traditional, intuitive, internalist, and individualist ones will do? There are (at least) three reasons.

First, insofar as our moral theorizing aims to cohere with what humans are actually like, then the research presented in Part I seems to provide empirical evidence for the inadequacy of strictly internalist and individualist accounts of cognition, emotion, perception, memories, relationship well-being, and goal pursuit. In the context of the extended cognition literature, we saw how an internalist perspective could not do justice to the perceptual experience of Tactile Visual Substitution System (TVSS) users. Because the TVSS is so tightly integrated with the user on the basis of ongoing feedback loops, there is good reason to think that the user's perceptual system *extends* to include the TVSS. In turn, this legitimized questions about whether and how other perceptual, cognitive, and affective processes – and especially those relevant to virtue ethics – may be similarly extended.

In the context of the literature on close relationships in social and personality psychology, we saw how an individualist perspective could not do justice to the ways in which romantic partners sculpt each other's ideal selves, and transactively encode and

retrieve information. Similarly, in the literature on transactive goal dynamics, we saw the potential shortcomings of individualist accounts of goal pursuit and self-regulation. The common thread is that on the basis of ongoing feedback loops, there is good reason to think that the bearers of ideals, memories, motivation, and regulation can *extend* to include relationship partners. In turn, this legitimized questions about whether and how the bearers of virtues and vices may be similarly extended.

My claim here is that doing empirically responsible moral theorizing *requires* taking seriously this kind of evidence. This much should be uncontroversial. It may turn out that the evidence does not support the externalist or anti-individualist conclusions I have arrived at. Of course, I think the evidence *does* support such conclusions, but that question must be settled through engagement and argumentation – not ruled out in advance. If my claims are granted, however, the question still remains whether these kinds of cases (e.g., Tactile Visual Substitution System, transactive memories, etc.) are closer to the exception or the rule. I am inclined to think that the phenomena revealed by taking externalist or anti-individualist stances - such as Michelangelo effects, transactive memories, and transactive goals – point to deep insights about the kinds of creatures we are. Recall Clark’s soft selves argument from Chapter I:

Different neural circuits provide different capacities, and contribute in different ways to our sense of self, of where we are, of what we can do, and to decision-making and choice. External, nonbiological elements provide still further capacities, and contribute in additional ways to our sense of who we are, where we are, what we can do, and to decision-making and choice. But no single tool among this complex kit is intrinsically thoughtful, ultimately in full control, or

plausibly identified as the inner ‘seat of the self’. We (we human individuals) just *are* these shifting coalitions of tools. We are ‘soft selves,’ continuously open to change and driven to leak through the confines of skin and skull, annexing more and more nonbiological elements as aspects of the machinery of mind itself (Clark, 2007b, p. 112).

I argued at length in Chapter III that extended cognition theorists overly emphasize agent-artifact couplings at the expense of agent-agent couplings. I think Clark is guilty as charged here. But with the friendly amendment that external, social elements – that is, other agents – can be similarly integrated, the fundamental insight holds: as a rule, humans just *are* these distributed, hybrid coalitions. The traditional internalistic and individualistic agent (think Hobbes’ mushroom) is an *achievement* - perhaps even an exception to this rule.

To be sure, it cannot be definitively settled here what human nature is *really* like. But my contention that we need to bother with an account of extended virtues is grounded in the idea that we need a set of normative concepts which accurately reflect claims about the kinds of creatures that we are. Again, it may turn out that the evidence does not support Clark’s view (though I am inclined to think it does), but again, that will be established through argumentation – not ruled out in advance. And insofar as Clark’s view here *is* independently plausible, then my account of extended virtues will be well-positioned to provide tools for moral theorizing which cohere with this view.

The second answer to the “Why bother?” question is that in addition to accurately reflecting the kinds of creatures we are, adopting a broadly externalist/anti-individualist stance also reveals novel explanatory and predictive resources which may be unavailable

to strict internalists/individualists. This was perhaps clearest in the context of transactive goal dynamics. If one (implicitly or explicitly) assumes that all of the processes relevant for explaining or predicting goal achievement (e.g., losing weight) are intraindividual (e.g., willpower), then it will be easy to overlook interpersonal processes (e.g., spouse temptation) which may turn out to be crucial for understanding how and why people fail to achieve their goals. Moreover, internalistic and individualistic views may foreclose on possible intervention strategies. Indeed, there are diverse sources of evidence suggesting that interpersonal dynamics are uniquely predictive of outcomes in the domain of diabetes and blood glucose monitoring (Anderson et al. 1997; Berg et al., 2013), sleep apnea (Baron et al., 2011), obesity (Best et al., 2016; Franks et al., 2012), arthritis (Hemphill et al. 2016), pain management (Wilson et al., 2017), and aging (Rauers, Riediger, Schmiedek, & Lindenberger, 2011).³⁶

I contend that the same will hold for virtue theory, broadly construed: my account of embedded/scaffolded/extended virtues aims to provide novel explanatory resources and to reveal new and interesting pathways relevant to virtue, character development, and flourishing. That is, in the same way agents may depend (in varying degrees) on their relationship partners to achieve goals like weight loss, so too can agents depend (in the varying degrees specified by the template in II.IV.II, in fact) on their relationship partners to manifest or possess virtues like open-mindedness or compassion. The claim, then, is that there is value-added for adopting an externalist and anti-individualist stance – it reveals pathways and strategies that may otherwise remain invisible or underutilized.

The third answer to the “Why bother?” question is that my account does not require burning down the whole edifice of virtue ethics. In fact, I think virtue ethics is

³⁶ Many thanks to Gráinne Fitzsimmons for calling these sources to my attention.

especially well-positioned relative to other normative ethical theories to incorporate the views of cognition, emotion, relationships, etc. I have been arguing for. More specifically, Annas's (1993, pp. 48-9) plausible and widely accepted account of the tripartite structure of virtue can, with some important tweaks, incorporate the insights from Part I. When Annas says that virtues have a cognitive or intellectual component, the extended virtue theorist concurs, with the friendly amendment that cognition is sometimes embedded in, scaffolded by, or extended to include environmental resources, including other agents. When Annas says that virtues have an affective or emotional aspect, the extended virtue theorist concurs, with the friendly amendment that emotion is sometimes embedded in, scaffolded by, or extended to include environmental resources, including other agents. The sticking point for Annas's account is her claim that virtues are essentially dispositional. Here, the extended virtue theorist will demur. Describing what virtues are, if not dispositions, is the task of Section III.

To sum up, we need an account of extended virtues because such an account will (1) more accurately reflect what humans are like, it will (2) provide novel explanatory resources for virtue theory, and (3) it does not require throwing the baby out with the bathwater. The end result will be an increased awareness of our situatedness, dependence, vulnerability, and responsibility for others (more on which below in Section V)

VI.III Two arguments for extended virtues

I will run two arguments which aim to incorporate the insights from Part I and avoid the shortcomings from Part II. The first is what I will call the *process* argument and the second is what I will call the *bearer* argument. These correspond the conclusions from Chapters II and III, respectively.

The conclusion of Chapter II was that the *processes* relevant to virtue ethics are not solely internal to the agent. The process argument, then, unfolds as follows. Chapter II argued that cognition and emotion are sometimes embedded/scaffolded/extended. Chapter V showed that structure of virtue is comprised of cognitive and emotional components. Therefore, if cognition and emotion are sometimes embedded/scaffolded/extended, and if virtues are essentially cognitive and emotional, then sometimes components of virtues are embedded/scaffolded/extended. In making this claim, I am working from within the empirically established framework of virtue countenanced by Miller (2013, 2014), and arguing that insofar as it claims to be empirically responsible, it must entertain the claim that the cognitive and affective components of virtue do not reside wholly within skindividuals.

The conclusion of Chapter III was that the *bearers* of the states and processes relevant to virtue ethics are not always skindividuals. The bearer argument unfolds as follows. Chapter III argued that some of the interpersonal processes studied by social and personality psychologists can be construed as embedded/scaffolded/extended. For example, the transactive goal dynamics framework describes how the motivational and self-regulatory processes involved in goal pursuit can become so intertwined in close relationships that it only makes sense to posit the coupled system itself as the bearer of those processes. In these cases, motivational and self-regulatory processes may *extend* to include romantic partners. In less intimate relationships, perhaps such processes may be better understood as embedded or scaffolded. Insofar as the cognitive, affective, motivational, and self-regulatory processes relevant to virtue ethics can be similarly intertwined (in varying degrees) with close others, then there is reason to think that

skindividuals are not the sole bearers of the states and processes relevant to virtue ethics, and that the bearers may themselves be embedded/scaffolded/extended. In making this claim, I aim to improve upon Howell's (2016) underdeveloped account of extended persons.

Putting these two arguments together yields a number of possibilities. Cognitive and affective processes can be embedded/scaffolded/extended. The bearers of these cognitive and affective processes can themselves be embedded/scaffolded/extended. Further, it is widely accepted that there are two (often cross-cutting) species of virtues: intellectual and moral. The full scope of what I am arguing, then, is that *the processes relevant to, and the bearers of, moral and intellectual virtues can be embedded/scaffolded/extended*. In what follows, I will explain and describe some of these iterations. Though I certainly will not exhaust the catalog of virtues, I will fill in relevant details of what embedded/scaffolded/extended intellectual and moral virtues might look like, alternating between the most abstract and general features, and specific, concrete examples. If this strategy succeeds, readers will be primed to explore the many ways in which the states and processes relevant to normative ethical theory may not be as internalistic or individualistic as is commonly supposed.

VI.IV Embedded, Scaffolded, and Extended Virtues

This section aims to spell out in more precise detail what it would mean to say that when appropriately integrated, contextual features (where context is broadly construed to include other agents, props, tools, and other environmental resources) can, to varying degrees, constitute virtues.

VI.IV.I Embedded Virtues

Recall some of the general features of the embedded cognition framework: Typical processes depend, in surprising and complex ways, on the organism's use of external resources, but those processes do not literally extend into the environment (Rupert, 2009). The relevant descriptors of this view were dependence (as opposed to constitution) and passivity (as opposed to active construction). The framework of embedding also maintains (or at least does not dissolve) a rigid distinction between the internal and external, organism and environment. All of this is to say that the framework of embedding is not inconsistent with individualism and internalism. In the context of the philosophy of cognitive science, this framework is perhaps best understood as granting claims from the situated cognition literature just up to the point that orthodox internalism and individualism can still be maintained. I contend the same will be true of embedded accounts of virtue. They will grant externalist and/or anti-individualist claims about virtue, just up to the point that the processes relevant to, and bearers of, virtues remain skindividualistic.

In the context of virtue theory, this would mean that embedded accounts of virtue are widespread. Indeed, I think many theorists who have taken the time to consider how context can stymie human flourishing probably endorse something like an embedded account of virtue. Perhaps the most famous instance of this is Aristotle's claim in the *Politics* that one cannot be a perfectly virtuous man in a corrupt polis (1276b28–35). Elsewhere, Aristotle discusses the role of external goods:

Yet evidently, as we said, it [happiness] needs the external goods as well; for it is impossible, or not easy, to do noble acts without the proper equipment. In many actions we use friends and riches and political power as instruments; and there are

some things the lack of which takes the lustre from blessedness, as good birth, satisfactory children, beauty (*Nicomachean Ethics*, 1099a31-1099b2).

The claim here is that these external goods will be necessary (but not sufficient) for human flourishing. That is, well-being for Aristotle *depends on* (among other things) having things like security, noble birth, numerous friends, wealth, good children, fitness for athletic contests, and good luck (*Rhetoric*, 1360b15-22). Flourishing is not *constituted* by such external goods, of course, but there is nonetheless widespread recognition in the virtue ethics tradition that circumstances and context do matter. But for many virtue theorists (Tessman (2005) is an important exception), these external goods (or lack thereof) are often a matter of luck or accident. So while the tradition has recognized the ways in which certain environments can be deleterious and virtue-undermining, the central focus is nonetheless on cultivating a resilient character that can weather the storms: “great and frequent reverses can crush and mar our bliss both by the pain they cause and by the hindrance they offer to many activities. *Yet nevertheless even in adversity nobility shines through*” (*Nicomachean Ethics*, 1100b20, emphasis mine).

Recalling Wilson (2004), this is a paradigmatic instance of internal richness (of character) combined with external minimalism (of circumstances). That is, virtue may turn out to depend on circumstances in complex or surprising ways, but the *really real* content of virtue is inside the individual. Perhaps the distinctive feature of embedded accounts of virtue, then, is the recognition that while various abilities and capacities *depend on* contextual features, this dependence is never so strong as to dislodge the locus of agency from the skindividual.

To see this in the technical terms introduced above, I will consider an example of how a *process* argument for an embedded *moral* virtue might play out. Consider the classic Athenian warrior virtue of courage. The courageous person lies at the mean between the coward who feels too much fear, and judges too many situations as dangerous, and the rash person, who feels too little fear, and judges too few situations as dangerous. There are a number of ways in which the cognitive (judging danger) and affective (feeling fear) processes constituting the virtue of courage might be embedded. In a typical military context, for example, individual judgements about the dangerousness of situations probably depend on (among other things) previously established protocols and orders from commanding officers. Feelings of fear probably depend on (among other things) the expectations and behavioral norms of fellow soldiers. None of this entails that the judgments or feelings extend beyond the individual soldier, but it does give voice to the ways in which such processes are conditioned by and reliant upon contextual features.

More formally, the pathways of influence between the cognitive and affective processes of the individual soldier and various contextual features of the military setting are asymmetrical and imbalanced. The commanding officer, by definition, exerts much more influence on the individual soldier than the soldier does on the officer. This small example is meant to illustrate how cognitive and affective *processes* (judgment and fear) relevant to *moral* virtues (courage) can be *embedded in* social contexts (chain of command). Similar arguments could easily be constructed for the claim that the *bearers* of courage are embedded in this social context, or that the processes relevant to, and bearers of, intellectual virtues are similarly embedded. Because the view of virtue I am describing is widespread, it will perhaps be more instructive to see how the stronger

claims of scaffolding and extending contrast with the examples here.³⁷ But to sum up the present section, the hypothesis of embedded virtue proposes that agents' moral or intellectual capacities depend on various contextual resources, but such dependence is asymmetric and provides little reason to think that the processes relevant to, and bearers of, virtue extend beyond the skin individuals.

VI.IV.II Scaffolded Virtues

As with the philosophy of cognitive science, the hypothesis of scaffolded virtue goes further than the hypothesis of embedded virtue. The former proposes that that agents' moral or intellectual capacities both depend on and have been transformed by contextual resources. Often these resources have been preserved, built, or modified precisely because they enhance agents' moral or intellectual capacities. The relevant features here are construction of contextual features (as opposed to dependence) and active sculpting (as opposed to passive reception). The framework of scaffolding softens the distinction between internal and external, agent and context, and recognizes the ways in which agents engineer their environments to serve various ends.

To illustrate this view, consider an instance of the *process* argument about an *intellectual* virtue. Curiosity is scaffolded by access to the internet via a smartphone. At first pass, curiosity is plausibly cast as the desire to seek out information and knowledge from diverse sources and perhaps to feel excitement in so doing. The claim that the virtue of curiosity is scaffolded by smartphones entails that the cognitive and affective

³⁷ But at first pass, here is how the processes described in the present example might be scaffolded or extended: If, as seems inevitable, police and military become widely equipped with facial recognition software (implemented on Google Glass, or something similar) with heads-up displays presenting information about identified subjects (including name, age, criminal record, and most importantly, potential threat level), then a case could be made that this interface scaffolds, if not extends, the officer's or soldier's judgements about threat, danger, etc.

components comprising curiosity, to some extent, depend on and have been transformed by an ability to quickly, easily, and reliably seek out information and knowledge from diverse sources via smartphones.

To illustrate the active sculpting feature of scaffolded curiosity, consider how smartphone users can configure their interface so as to be a single button press or voice command away from search bars for Google or Wikipedia. These tools afford information and knowledge seeking with unprecedented ease and speed. Moreover (and this bleeds into the closely related intellectual virtue of open-mindedness or the correlative intellectual vice of close-mindedness), users can deliberately sculpt their informational and epistemic environments to be exposed to certain streams of information.

For example, the KIND Foundation has developed a Facebook plug-in called “pop your bubble” which suggests open but differently minded users (on the basis of age, gender, location, political affiliation, etc.) to connect with.³⁸ Users can thus engineer their social media environs to encounter sources of information they otherwise might never have encountered; no small feat when a majority of Americans now get their news from social media (Pew Research Center, Sept 7, 2017)³⁹

The key point here is that smartphones are fast, easy, and reliable tools for seeking information and knowledge, and they afford opportunities to be more curious than users might be without them. It is easy to imagine that once an agent is equipped with this tool, their propensity to seek information and knowledge may increase. In turn, this propensity may lead agents to further engineer their epistemic environment which

³⁸ The plug-in can be found at <https://www.popyourbubble.com/>

³⁹ Of course, this can, and more often does, happen in the opposite direction where users sculpt their epistemic niche to only encounter other users who are likely to share their worldview.

will subsequently entrench the propensity to seek information, and so on. This process of entrenching habits is precisely how many theorists (e.g., Zagzebski, 1998) describe the cultivation of virtue. It would be little more than the manifestation of an internalist bias, then, to deny that smartphones, when properly integrated with the agent, can scaffold an agent's curiosity.

But what precisely is meant here by "properly integrated"? Again, drawing from the template in II.IV.II, we can characterize the smartphone as scaffolding (rather than embedding or extending) curiosity because the pathways of influence between the agent and the smartphone are symmetric but imbalanced. In the first place, unlike with embedding, there is bi-directional or symmetrical information flow. The agent clearly influences the smartphone by initiating search queries or logging on to social media to seek information. And the reliable presence of the smartphone does influence the agent's downstream tendencies and abilities to seek out information. The influence is ultimately imbalanced, however, because the smartphone likely does not influence the agent to the same degree as the agent influences the smartphone (though perhaps parents of the so-called iGen will disagree). That is, for the most part, smartphones do not seek out information and knowledge apart from users' commands. Perhaps in the not so distant future, this influence will become more balanced, at which point the virtue of curiosity could extend to include smartphones or other devices.

V.IV.III Extended Virtues

In Chapter V, we saw how character is a collection of traits, and how those distinctively *moral* traits comprise the domain of virtues and vices. Elsewhere, I have developed a series of arguments for the possibility of embedded and extended character,

and showed how the structure of friendship exemplifies the kind of ongoing feedback loops characteristic of extended coupled systems (Alfano & Skorburg, 2016). These arguments are perhaps best understood as proof of concept. In hindsight, it seems hasty to argue for extended moral character without first specifying the ways in which the components of moral character – virtues and vices – might be similarly extended. In the present section, I will attempt to provide such an account.

Drawing again from the template in II.IV.II, the hypothesis of extended virtue proposes that agents' moral or intellectual capacities can be so tightly integrated with various contextual features on the basis of ongoing feedback loops, that it only makes sense to posit a new, extended moral or intellectual system. The relevant descriptor here is *constitution* (as opposed to dependence or support). That is, contextual features external to the agent literally constitute certain features of the virtue. They do so on the basis of reciprocal interaction (as opposed to comparatively one-way sculpting, as we saw in the case of scaffolding). To illustrate this, I will consider the interplay of affective processes involved in generosity and gratitude, or, using the technical terms, I will consider a *bearer* argument in the context of *moral* virtues.

Robert Roberts (2013), Drawing on a passage from Gareth Matthews (1980), explores the importance of expectation and confirmation of emotional states within relationships. One illustrative example involves a sister who generously gives up a concert ticket so that her brother can go instead. In this scenario, it is plausible to think that when the brother receives the ticket, he may feel the emotion of gratitude for the gift from his sister, which he expresses by some token of thanks. According to Roberts:

She responds to his gratitude by acknowledging his acknowledgment and his desire. She is gratified by his gratitude. His gratitude fulfils her concern that he be gratified by what she has given him and that he acknowledge her as his benefactor. And he may in turn be gratified that she is gratified by his gratitude.

This brother-sister relationship seems to be going on swimmingly (p. 137).

For Roberts, these positive feedback loops are constitutive of friendship and other close relationships (and when oppositely valenced, they are constitutive of enmity). Though I will not defend the claim here, it strikes me that the structure laid out here accurately characterizes many of the most meaningful kinds of human connection, including not only gratitude and friendship, but also love and trust. The important point for present purposes is that this schematic example shows how affective components of a virtue can extend to include other people. That is, essential features of the brother's gratitude are, in a sense, partially *constituted by* his sister's second-order gratitude about his first-order gratitude for her generosity. This ping-ponging between the expectation, expression, and confirmation of affective states illustrates how the bearer of the virtue can extend from the brother to include the sister. To see this, consider how the case fits the template from II.IV.II.

Formally, we said that X extends to include Y when Y is constitutive of some part of X, because the pathways of influence between X and Y are mostly symmetric and balanced. The pathways of influence are symmetric, because the shape of the pathways between brother and sister admit of bi-directional influence between them. She can expect and signal affective states to him in the same way he can expect and signal affective states to her. And in most cases of extending, recall that the degree of influence

is distinctively *balanced* and therefore *reciprocal*. The case of gratitude in response to generosity illustrates this nicely. As Roberts points out, the prepositions used to describe the exchange illustrate the reciprocal structure. For the sister's act to be fully generous, the gift must be *for* her brother's sake (rather than creating a debt he must repay later).⁴⁰ In turn, the brother's gratitude is the proper response *to* his sister's generosity. Her second-order gratitude is her feeling grateful *for* her brother's first-order gratitude *for* her generosity which was *for* his sake in the first place. What starts off as a decidedly one-way interaction (giving for the sake of another) quickly unfolds into a deeply meaningful and edifying reciprocal exchange of affective states.

In Chapter II, we saw how Tactile Visual Substitution Systems are so tightly integrated with its user as to constitute an extended perceptual system. The thrust of this claim was that the user in some sense determines the inputs to the TVSS by orienting to the world in certain way. But this is not quite right, because the TVSS - fluidly and in real-time, informs the wearer's orientation to the world, which in turn changes the inputs to the TVSS, which updates the possibilities for engagement, which in turn updates and drives the wearer's orientation, and so on. This is why Palermos (2014) argued that extended systems, properly understood, cannot be de-composed into simple input and output units. My claim about gratitude is analogous. On the basis of ongoing feedback loops, the brother and sister constitute an extended affective system. Her generous

⁴⁰ Here is how Roberts and Wood (2011) define generosity: "Generosity is a disposition to give valuable things—material goods, time, attention, energy, concessions, credit, the benefit of a doubt, knowledge—to other persons. It is associated with the idea of freedom and is sometimes called liberality (*liberalitas*). It is a disposition to give "freely", gladly, and without calculation of repayment. But this freedom is not mere prodigality; generosity is a disposition to give for the good of the recipient—his well-being, his pleasure, the fulfillment of his purposes; this "altruism" is the kind of motivation characteristic of generosity. The man who sets the contents of his overstuffed garage on the street for anonymous neighborhood scavengers may have a partially generous motivation, but if he does so primarily to clear space in his garage, his act is not a paradigm of generosity" (p. 286)

motivation elicits his gratitude, which in turn creates and invites a new affective possibility for her (feeling grateful for his gratitude), which further influences and indeed *enhances* his affective state (feeling grateful for her gratitude for his gratitude for her generosity). To claim that their affective states merely *depend* on one another is to mischaracterize this rich interplay and coordination, and it is for this reason that the language of bi-directionality, balance, reciprocity - that is *extending* – is necessary.

We have now seen a number of ways in which relationship partners are keenly attuned to their partner’s ideals, goals, memories, motivations, expectations, and emotions. On this basis, it should be easy to uncover other instances of cognitive and affective processes which have important bearing on normative considerations to develop further extended virtue-style arguments. Moreover, the ground covered in this section also suggests another answer to the “Why bother?” question: because an account of extended virtues provides a rich and textured description of the phenomenology of normative relationships.

VI.V A new metaphysics of virtue: What are virtues, if not dispositions? And where are they, if not inside agents?

In the previous section, I sketched but a few permutations of my central claim that the processes relevant to, and the bearers of, moral and intellectual virtues can be embedded/scaffolded/extended. In the present section, I will abstract from some of those particulars in an attempt to address the shortcomings identified in Chapter V, Section IV - namely, that dispositionalism about virtues (the view that virtues are on a metaphysical par with ‘solubility’ or ‘fragility’) reproduces undesirable internalism and individualism. Insofar as virtues have traditionally been construed as dispositions, this view tends to

endorse the theses of internal richness and external minimalism (Wilson, 2004). In turn, the default view holds that virtues primarily inhere within the agent, and the eliciting conditions external to the agent play, at best, a secondary role. This section, then, spells out the metaphysics of extended virtues and attempts to answer the questions, what are virtues, if not dispositions? And where are they, if not inside the agent?

I contend that virtues are less like dispositions, and more like relations. My argument here has three steps. First, I will introduce the claim in Jayawickreme & Chemero (2008) that virtues can be understood as part of moral analogues of affordances. The second step will be to figure out what exactly affordances are, such that virtues are part of the moral analogues of them. Third, I will argue that virtues inhere not so much *in* agents, but rather, at the intersections among agents, other agents, and non-social environmental features, and that this view is well-equipped to deal with the kinds of cases described in the previous section.

Jayawickreme & Chemero (2008) posit that virtues are part of the moral analogues of affordances. They write: “a moral analogue of an affordance (MAA) is an opportunity for moral behavior. As with affordances, MAAs are relations between morally relevant abilities and morally relevant situations...Virtues, in other words, are abilities to behave appropriately in morally relevant situations” (p. 122). Of course, much more needs to be said about abilities and affordances, but with their conclusion on the table, it is worthwhile to briefly note the similarities and differences between their project and my own. In the first place, their paper is framed by debates in positive psychology (e.g., Seligman & Csikszentmihalyi, 2000). One aim of their project, which is largely orthogonal to my own, is to show that positive psychology - concerned as it is with

normative claims about values and well-being – can nonetheless remain thoroughly scientific by availing itself of the resources of ecological psychology (e.g., Gibson, 1979). Thus, many of their claims are aimed at showing how it is possible to rigorously and empirically study human well-being without committing some version of the naturalistic fallacy. While my project is only indirectly related to work in positive psychology, the tradition of ecological psychology from which Jayawickreme & Chemero (2008) draw does have important connections with my larger aims. Consider the following passage:

We recommend that researchers interested in an empirically informed account of moral behavior adopt some of the methods of Gibsonian ecological psychology, a contemporary psychology that is in many ways a continuation of the pragmatist psychology of a century ago (Gibson, 1979; James, 1912/1976). There are several advantages to adopting a broadly Gibsonian perspective to the study of moral behavior. First, it provides a novel theory of the nature of moral virtues. Second, it makes both moral virtues and moral behavior empirically tractable (Jayawickreme & Chemero, 2008, p. 118).

The connection to classical American pragmatism, coupled with the commitment to empirically tractable moral theorizing are, of course, near and dear to the present project. But the most important claim is that a broadly Gibsonian perspective will yield a novel theory of the nature of the virtues, which is precisely what I am after here. We will first need to see, however, what is meant by Gibson's concept of affordances before we can assess its value in the moral domain. This is no small task, because as Tom McClelland has recently pointed out, "the term 'affordance' has become unmanageably polysemous,"

given its proliferation across various disciplines including cognitive psychology, neuroscience, music, anthropology, design theory, artificial intelligence, phenomenology, and philosophy of perception. (McClelland, 2017, pp. 2-3).

Nevertheless, I will start where everyone does, with James Gibson's (1979) magnum opus, *The Ecological Theory of Visual Perception*. Therein, Gibson claims, "the affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill" (p. 127). Thus, for creatures like humans, slow-moving streams just a few inches deep may *afford* walkability, while faster-moving streams a few feet deep may only *afford* wadeability. The same stream *affords* drinkability to deer, or habitability to fish. The chair affords sitability to me, climbability to the toddler, and sleepability and/or scratchability to the cat. This straightforward definition of an affordance is fine as far as it goes, but Gibson muddies the water a few pages later with his more widely cited description:

An affordance is neither an objective property nor a subjective property; or it is both if you like. An affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behavior. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer (Gibson, 1979, p. 129).

As Chemero (2003, p. 182) notes, this makes affordances seem like "impossible, ghostly entities, entities that no respectable scientist (or science-worshipping analytic philosopher) could have as part of their ontology." And indeed, much ink has been spilled by ecological psychologists trying to figure out what to make of this description. I cannot

pretend to do justice to this vast body of literature here, so I will only briefly rehearse Chemero's (2003) argument that affordances are best understood as *relations*, because this formulation paves the way for my own positive account in the domain of virtues.

More specifically, Chemero (2003) claims that *affordances are relations between abilities of an organism and features of the environment*.⁴¹ Thus, the affordance climbability is not merely a one-sided, organism-relevant property of the environment, but is instead a *relation* between organismal abilities (such as leg length, grip strength, flexibility, etc.) and environmental features (such as surface area, surface texture, incline, etc.). Throughout his work on this topic, Chemero emphasizes the importance of this organism-environment mutuality in much the same spirit of Dewey's view outlined in Chapter II, Section II. However, as we will see below, it is questionable whether this commitment to mutuality is upheld in the analogy with the moral domain.

Much more could be said about the affordances-as-relations view and its connection to other formulations in the ecological psychology literature. All that is needed at the moment, however, is a way to see that construing affordances as relations does not necessarily result in ghostly, empirically intractable entities. To accomplish this, Chemero (2003, p. 187) illustrates the "affords" relation with reference to the "taller" relation. If we say that Shaq is taller than Tony, we do not need to posit any other objects besides Shaq and Tony. And the "taller than" is not an intrinsic property inhering in either Shaq or Tony. The relation depends on both of them. "Affords" is like "taller than"

⁴¹ I am glossing over a number of details here about the precise nature of abilities (as opposed to something like body scale), properties, and features. More specifically, Chemero's formulation is developed in contrast to a number of ecological psychologists (e.g. Reed, 1996; Turvey, 1992) who claim that affordances are animal-relative *properties* of the environment. Chemero specifically rejects the species of this view which holds that affordances are dispositional properties of the environment. Although the analogy is not perfect, the argument here does bear some resemblance to my rejection of dispositionalism about virtues in Chapter V. This is taken up in more detail below.

in followings respects: “It is neither of the person, nor of the environment, but rather of their combination. Second, the affordance is not an extra thing in any of the usual senses of “thing.” Yet it exists nonetheless, and, like the fact that Shaquille is taller than Tony, is quite perceivable” (Chemero, 2003, p. 187).⁴²

Although I am again glossing over a number of details, we now have the necessary pieces in place to return to Jayawickreme & Chemero’s application of the notion of affordances to the moral domain. In one sense, the formulation is straightforward: “MAAs are relations between morally relevant abilities and morally relevant situations” (p. 122). In the same way organismic abilities (like flexibility and strength) are related to environmental features (like surface texture and incline) to constitute affordances like walkability, so too are moral abilities (like perceiving suffering and providing comfort) related to features of moral situations (like suffering people in need) to constitute the moral analogues of affordances.

On this analogy, virtues are to abilities as moral situations are to environmental features. Thus, Jayawickreme & Chemero claim “virtues, in other words, are abilities to behave appropriately in morally relevant situations” (p. 122). The part of the moral analogue of affordances that virtues are, then, is *abilities*. With this claim in place, the question before us now is whether the virtues-as-abilities view is able to avoid the shortcomings identified in Chapter V, Section IV with the virtues-as-dispositions view. Put another way, can construing virtues as abilities accommodate the view of virtues as extended?

⁴² Chemero further notes that the claim that the “taller than” relation is perfectly perceivable and perfectly real does rely on the plausibility of something like William James’s (1912/1976) radical empiricist view the relations themselves are just as real as the relata.

On the one hand, construing virtues as abilities is clearly an improvement upon the received view of virtues as dispositions. But on the other hand, construing virtues as abilities may not fully capture the kind of interactionism (or what Chemero would call animal-environment mutuality) characteristic of scaffolded or extended virtues. In the first place, it is worth noting that Chemero's account of abilities – and by extension, virtues - is explicitly anti-dispositional:

The problem with seeing abilities as dispositions is that, when coupled with the right enabling conditions, dispositions are guaranteed to become manifest. The soluble solid sugar will always dissolve in water in suitable conditions. This is not true of abilities. Having the ability to walk does not mean that one will not fall down even in the ideal conditions for walking (Chemero, 2003, p. 189).

My brief against dispositional accounts of virtue is somewhat similar. I argued that construing virtues as dispositions forecloses the possibility of agents (bearers of the disposition) meaningfully modifying their environment (the eliciting conditions). We both agree, then, that dispositionalism is an inadequate metaphysical basis for virtue. The pressing question is whether *abilities* provide the proper basis, or whether ability talk still risks bringing along unseemly individualist/internalist baggage.

In one sense, it is not unreasonable to think of abilities as organism-bound. Chemero (2003) seems to hold that abilities are functional properties of animals, where animals are individualistically construed (see Chemero 2003, pp. 189-90). If this is right, then the moral analogue – virtues – would also be organism-bound. On its face, this result is incompatible with scaffolded and extended accounts of virtue.

But perhaps in another sense, abilities need not be organism-bound. After all, it would be strange to talk about walkability without *any* reference to environmental features. The worry, though, is that even on this more liberal interpretation, walkability may be said to merely *depend on* – that is, it may only be *embedded* in – various features of the environment. To be clear, it is not absurd to think that abilities ultimately inhere inside of organisms. Indeed, something like this is probably the default view. But from the perspective of extended cognition arguments, this looks suspiciously similar to the internalist claim that “sure, various environmental props play an important role in cognition, and perhaps even more so than we first thought. But the *really real* cognitive stuff still only happens in the head.” There is reason to think, then, that all of the problems associated with this view in the philosophy of cognitive science will also obtain for the moral analogue. That is, virtues-as-abilities seems like a paradigmatic instance of embedded virtues, and it is difficult to see how it would be compatible with scaffolded or extended accounts of virtue.

If this assessment is on the right track, it illustrates just how deep skin individualist biases can run. Even in a framework thought to be friendly to extended cognition-style arguments, it is still difficult to eschew organism-bounded thinking. There is no doubt that construing virtues as part of the moral analogue of affordances represents an important improvement over dispositionalism about virtues. But insofar as scaffolded and extended virtues are still on the table, we will need a stronger view than is advanced by Jayawickreme & Chemero (2008). Luckily, the resources for such a view are present, if underdeveloped in the analogy with affordances.

My proposal is that the claim “virtues are *part of* the moral analogue of affordances” be strengthened to “virtues *are* the moral analogue of affordances,” where affordances are construed as relations. Thus, virtues are the relations between various ‘internal’ features of the agent and various ‘external’ features of the environment.

I hasten to add that I am not arguing for a straightforward identification between virtues and affordances. The latter, after all, are entities in the domain of vision science and the former are entities in the domain of normative ethics. Jayawickreme & Chemero make this point by noting: “Affordances are not virtues. Humans have virtues, while affordances are in the environment (or at least in the environment-plus-human system). Thus, affordances and virtues are different kinds of things” (p. 122).

While we agree about the status of the analogy, I contend that in the absence of any qualification, the claim that “humans have virtues,” reproduces undesirable individualism. This is why virtues cannot be merely *part of* the moral analogue of affordances, but are better construed *as* the moral analogue of affordances. In the same way affordances are in the environment-plus-human systems, so too with virtues. If my argument in the previous section is on the right track, then it is at best incomplete, and at worst inaccurate, to think that the brother who received the concert ticket is the sole bearer of the virtue of gratitude. To claim that he just “has” the virtue is to downplay and flatten the rich and intricate emotional feedback loops which characterize his relationship with his sister. Similar examples can be multiplied with respect to romantic partners’ memories, ideals, and goals as these bear on virtue theory. It is not just that one partner “has” these things in the intuitive sense that “humans have virtues.” It is that these

memories, ideals, and goals are constituted and enacted – or, in virtue parlance, possessed and expressed - through mutual interaction.

The reason the affordance analogy is useful is that in the same way it helped perceptual psychologists recognize that vision ain't all (about a rich, detailed, internal representation) in the head, the affordance analogy in virtue ethics helps us to see that virtues ain't all inside the agent. Where Jayawickreme & Chemero seem to hold that the agent side of the relata (e.g. abilities) is sufficient to capture the nature and function of virtue, I contend that a more mutualistic and interactionist approach – one that places virtues more fully within the agent-environment systems - is needed to cash in on the benefits of extended virtues described above in Section IV.

By making this stronger claim, at least two pressing questions arise. First, where are virtues, if not in the agent? Second, if virtues are not solely within agents, do they extend to include everything in the environment? In the balance of this section, I will address each in turn.

I have argued that Jayawickreme & Chemero's account of virtues as abilities does not go far enough in disavowing individualism. But if virtues are not in individuals, where are they? The beginnings of an answer can be found in Alfano's (2014) claim:

given the sorts of creatures we are — embodied, socially embedded, with cognition and motivation extended beyond the boundaries of our own skin...Virtue would inhere, on this view, in the interstices between the person and her world. The object that possesses the virtue in question would be a functionally and physically extended complex comprising the agent, her social setting, and her asocial environment...provided that the social and asocial phenomena outside the

moral agent's skin are appropriately integrated into her functioning, they may count as part of her moral agency and partially constitute her moral virtues (Alfano, 2014, p. 84).

The claim here that virtues are at the intersections among an agent, other agents, and other non-social features of the environment provides a straightforward answer to the question, "where are virtues?" Moreover, this looks to square nicely with my view of virtues-as-relations. So far so good. But, like Howell (2016), Alfano glosses far too quickly over the complex questions about what exactly constitutes "appropriate integration." In turn, this leads to the second question I raised. If virtues are relations, and are located at the intersections among agents, other agents, and non-social features of the environment, which parts are related to which? How do we know where to stop? In other words, does this account commit the coupling-constitution fallacy? Does it suffer from virtue bloat?

Luckily, the template developed in Chapter II, Section IV and applied in Section IV above provides the necessary resources to deflect these objections. To head off these worries, we can say that virtues extend beyond the agent to the extent that the pathways of influence between the agent and context (broadly construed) are symmetric and balanced. For example, in cases where my cognitive and affective processes are continually and reciprocally influencing and being influenced by my partner's, then there is reason to think that insofar as these processes are relevant to the possession or expression of a virtue, the virtue lies not within one of us, but at the intersection between us. The more imbalanced or asymmetric these influences, the more the virtue-relation is weighted toward the individual agent. In just the same way this set of principled criteria

were able to ward off objections of cognitive bloat, the thresholds of symmetry, balance, and reciprocity similarly promise to block objections about virtue bloat. And insofar as Chemero's (2003) account of affordances-as-relations holds, then there is no reason to think that virtues-as-relations is any less real, any less perceivable, or any less empirically tractable than concepts like height, flexibility, or climbability.

To sum up, in this section I sketched some of the metaphysical details of my account of extended virtues. In Chapter V, I identified problems with the dominant view of virtues-as-dispositions. Here, I argued that while Jayawickreme & Chemero's (2008) analogy between ecological psychology and virtue theory – and the subsequent view of virtues-as-abilities - represents an incremental improvement over the dispositional view, it does not go far enough. My positive claim was the virtues are better construed as *relations* which exist at the intersection among agents and context (broadly construed). In the next section, I take up questions about what all of this means for our models and metaphors of moral agency.

VI.VI Toward an ecological virtue ethics

Having now seen some specific components of specific virtues (Section IV) as well as some of the more general metaphysical features of my account of embedded/scaffolded/extended virtues (Section V), I will now consider some of the implications of these views for our thinking about moral agency more generally.

Perhaps the most pressing implication is that all of the foregoing suggests that when we pay careful attention to the many ways agents depend on context, we will see that developing, possessing, and expressing virtues may be difficult, if not impossible in the absence of conducive contextual resources, or in the face of contextual impediments.

The way to think about the claim that virtues *extend* (as opposed to embed or scaffold) to include contextual features is that those contextual features literally constitute essential components of the virtue. Without them, the virtue in question cannot be developed, possessed, or expressed. If my arguments in Section IV.III are on the right track, then a beneficiary cannot fully possess the virtue of gratitude in the absence of reciprocal affective feedback loops from the benefactor. Alfano (2015) makes similar arguments about friendship and trust, namely, that in the absence of a reciprocator, the virtues are unattainable. These are strong claims. To weaken them a bit, we can note that when an agent has developed personalized and customized search features on a smartphone, that agent is less capable of manifesting curiosity in the absence of the smartphone. I have argued elsewhere, drawing from the literature on stereotype threat (e.g., Steele & Aronson, 1995; but see also Flore & Wicherts, 2015), that negative stereotypes in an agent's environment can similarly undermine the development and expression of intellectual virtues (Alfano & Skorburg, 2016; Skorburg & Alfano, forthcoming).

Still, even the weakest of these claims are at odds with much of the virtue tradition. To get a sense of this, consider Johnson's (1993) depiction of the Moral Law folk theory – the set of assumptions, concepts, and metaphors underlying the dominant Western, Judeo-Christian morality which serves as the backdrop for contemporary ethical theorizing:

According to the Moral Law folk theory, then, *morality is a massive, ongoing power struggle* between the forces of reason and the forces of passion. Moral behavior thus requires us to keep our moral reason pure (so that it will give us the right principles of action) and *to keep our will strong* (so that we have the

willpower to do what our reason tells us is right). We thus come to experience our moral lives as ongoing struggles to develop and preserve purity of reason and *strength of will* in the face of constant pressures that arise from our embodiment in the world (p. 17, emphasis mine).

While my arguments have not directly targeted the reason/passion dichotomy described here, they have targeted the direct entailment of this dichotomy. Namely, that *morality is fundamentally a matter of keeping a strong will and cultivating inner strength*. This view is perhaps most prominent in Kant's de-theologized rationalist ethics. Thus, with respect to Kant, Johnson continues, "the Judeo-Christian emphasis on purity of heart and 'inwardness' is translated into an emphasis on correct willing that overcomes an external influence...which may be out of one's control" (p. 25). Insofar as this view is representative of the tradition – and there is good reason to think it is – then, moral virtue is cast as a matter of brute inner strength, and moral cultivation becomes akin to spiritual weightlifting. All of this is of a piece with Aristotle's claim above that "even in adversity nobility shines through." A similar theme is also summed up nicely by Doris (2002):

It's commonly presumed that good character inoculates against shifting fortune, and English has a rich vocabulary for expressing this belief: *stead, dependable, steadfast, unwavering, unflinching*. Conversely, the language generously supplies terms of abuse marking lack of character: *weak, fickle, disloyal, faithless, irresolute*...in this view, it's character, more than circumstance, that decides the moral texture of a life (p. 1, emphasis in original).

And I could go on. The point is that the argument I have developed goes right to the heart of traditional assumptions about the nature of morality. If I have successfully made a case

for embedded/scaffolded/extended virtues, then the moral life cannot be primarily about cultivating inner strength. In light of these arguments, what then, should replace this view?

Consider Howell's (2016) metaphor of a solar-powered traffic light. He notes that there are many ways to build such a light, some better than others. A light that depended so heavily on the sun that it ceased to work on cloudy days would clearly be bad light. A better light would still depend on the sun, but not so heavily that it goes out at night or on overcast days. In both cases, the light does what it does by virtue of its dependence on environmental resources. This provides the basis for the analogy with extended virtues:

To some extent, the degree to which virtues can be extended is itself a function of the environment. On a twin earth where the sun never sets, a solar-powered stoplight could be good without an internal battery. Similarly, if a virtuous disposition's external grounds are stable enough (and nonaccidentally so), there is less need for internal sources of moral power (p. 160).

The idea here is that in some environments, where the 'external' resources are highly conducive, there is less need for the exercise of 'internal' agential resources. Marking a trail offloads the cognitive task of route-finding to a less demanding perceptual task. Informing your friends about your dieting intentions offloads some of the burden of self-control to the social environment. For Howell, moral power can be similarly offloaded to the environment, as it were. This view then casts virtues as threshold concepts, which I take to be a desirable result. That is, virtues can be *achieved* through various combinations of environmental and agential contributions. In amenable environments, the

threshold for agential contributions is lower. In hostile environments, the threshold for agential contributions is higher.

I worry, however, that the analogy is not quite right – at least not completely. First, it risks reproducing the same view we are trying to move past, namely, morality as a source of inner strength or power which is constantly battling the uncertainties and perils of the external world. To be sure, Howell’s view suggests that such internal power can be offloaded and stabilized by external resources. But it is unclear whether this simply kicks the can down the road without addressing the fundamental assumptions that lead to the problem in the first place. If it is the case that moral power has to be offloaded to the environment in the service of freeing up more internal power, then when are right back where we started.

Second, because Howell assumes dispositionalism about virtues, the analogy is not going to do justice to the distinctively symmetrical nature of *scaffolded* or *extended* virtues. The mistake, I think, is the same one made in construing virtues as dispositions, and that is a framing in terms of passive dependence. In the same way vases cannot meaningfully modify their eliciting conditions, Howell’s traffic light is inert with respect to its environment – it passively receives and stores sunlight. In this sense, it is a perhaps a fitting analogy for embedded virtues, but it fails to capture the distinctive bi-directionality of scaffolded and extended virtues. In other words, the traffic light analogy captures the ways in which context can promote (or stymie) the manifestation of certain dispositions, but it fails to capture the ways in which agents can actively modify and be modified by those contextual features. So while Howell’s view takes important steps in

the right direction, it does not go far enough in refuting the traditional view that must also be refuted, insofar either of our accounts of extended virtues are on the right track.

It remains an open question whether there is a better model or metaphor to capture the account of extended virtues on offer. Perhaps this is to be expected, given that what we are up against is, in a sense, the predominant view of individuality that has pervaded ethical and political theorizing in the West. It may very well be that this tradition (from which the present treatment cannot be extricated) does not furnish the requisite conceptual resources to fully capture the insight that developing, possessing, and manifesting virtues *is not always up to us*.

My project is hardly the first to bump up against this possible limitation and in the remainder of this section, I will introduce what I take to be the most promising way forward. Recall that what we are ultimately after here is a model or metaphor which gets away from the individualist, internalist – that is, monadic - connotations which cast the cultivation and expression of virtue as a matter of inner strength. It is in precisely this spirit that ecological metaphors have proliferated across disciplines that are not themselves ecology. Rather than casting agents as monadic singletons, the ecological perspective, in the broadest terms, emphasizes interconnection, dependence, vulnerability, and situatedness.

Indeed, it is this very metaphor which laid the groundwork for the extended cognition arguments at the core of this dissertation. In the philosophy of cognitive science, the ecological metaphor was popularized by Edwin Hutchins (see Hutchins (2010) for a review), though Hutchins traces this to a confluence of theories in the 1970s including the Gibsonian tradition covered above, Gregory Bateson's (1972) cybernetic

approach, and Lev Vygotsky's (1978) cultural-historical activity theory. More recently, Natalia Washington (2015) has invoked the concept of *stewardship* – a hallmark of environmental ethics - to explore the implications of this “ecological turn” for mental health and the philosophy of psychiatry. This is, I think, the most productive way forward. That is, in light of arguments for embedded/scaffolded/extended virtues, moral agents should be construed less like weightlifters and more like stewards. Highlighting the central commitments of such a view here should suffice for setting an agenda for future research.

According to Welchman (2012, p. 299), the word “steward” derives from the Old English “stigweard,” a servant who looks after a hall, manor or landed estate. The concept has taken on a life – or better, lives – to describe various aspects of the relationship between humanity and the rest of the natural world. To my ear, the concept has a distinctively religious ring, and while there is certainly a historical basis for this,⁴³ one need not adopt any theistic commitments to see the value of the model. Of course, I cannot adequately summarize generations of scholarship on the concept of stewardship in the environmental ethics literature here, much less its recent appearance in management theory (Caldwell, Truong, Linh, & Tuan, (2011)), business ethics (e.g., Caldwell, Hayes, & Long, 2010), and healthcare ethics (e.g., Jansen, 2013). Nonetheless, what I have in mind is that a *steward is someone who is charged with taking good care of what is entrusted to them.*

If my arguments in Part II are on the right track, then we are – for better or worse - entrusted with the moral and intellectual character of our close friends and partners. Rather than toiling away in isolation to achieve our ideals, we continually enact these

⁴³ See e.g., Welchman (2012, p. 311, n. 2)

ideals (or fail to) in and through interpersonal relationships. Rather than striving to fortify one's character to be robust against such external influences, the stewardship model prescribes actively managing these external influences and engineering pathways conducive to flourishing. Rather than entrenching reliable dispositions to stimuli, this ecological approach aims to cultivate various *relations* between agents and contexts. Fitting metaphors for extended virtues will evoke less of a battle between opposing forces, and more of a harmonizing or integrating of internal and external influences.

As I noted above in Chapter IV, Section IV, the properties of *vulnerability* and *porosity* seem to be required for getting scaffolding and extending arguments off the ground. After all, if agents were not open to the influence of various contextual features, then how could they be scaffolded by them, or extend to include them? Granting this susceptibility to contextual features opens a can of worms, however. Once granted, we can either ignore the insight, letting the contextual features influence us as they will, or we can attempt to sculpt, engineer, and harness them to promote our values. The latter clearly seems preferable, and the steward as ecological manager seems particularly apt for this role.

I have only begun to scratch the surface of an ecological approach to virtue ethics here, but we can already see how this view implies that there is a sense in which our moral agency is not entirely up to us. In the absence of conducive environmental scaffolds – that is, from within a dysfunctional moral ecology – some virtues may be simply unavailable. It seems unlikely, for example, that any amount of grit or nobility, or any strength of will will be sufficient to overcome the cognitive decrements associated

with lead in the drinking water supply (see e.g., Lanphear et al. (2005)).⁴⁴ Of course, this flies in the face of the Hobbesian mushroom account of autonomy and the inner strength model of virtue. But that is, after all, the whole point.

Perhaps the principal advantage of the steward model is that, in virtue of its ecological connotations, it takes as fundamental (rather than exceptional) the kind of mutuality and reciprocity I have argued is at the heart of extended virtues. Recalling Section I above, I contend that such a view not only provides novel explanatory resources, but also more accurately reflects the complicated, messy, and dynamic nature of the moral life. In much the same way the research on transactive goal dynamics trades off empirically clean experimental design for, as it were, ecological validity, an extended account of virtues eschews dominant assumptions about moral agency in an attempt to more accurately capture the kinds of hyper-social, interdependent, and porous creatures we are.

VI.VII Conclusion

In this dissertation, I first introduced a burgeoning research program in the philosophy of cognitive science which has suggested that cognitive and affective processes are not necessarily organism-bound. I then tried to show how similar insights have been independently developed in various lines of research in social and personality psychology. Taken together, I argued that the insights garnered from these traditions have profound, but largely unrecognized implications for normative ethical theories such as virtue ethics. This was seen most clearly by noting that the cognitive and affective

⁴⁴ Hopefully it goes without saying that this is decidedly *not* an endorsement of crude or cartoonish behaviorism about the cultivation of virtue, such that changing the environment changes the behavior. Instead, as should be clear from all of the foregoing, this only emphasizes the fundamentality of *interactionism* about the cultivation of virtue.

processes posited by plausible and mainstream theories of virtue (e.g. Annas, 1993; Miller, 2013, 2014) *just are* the cognitive and affective processes called into question by, among others, extended cognition and transactive goal theorists. Thus, I argued that strictly internalist and individualist accounts of virtue are empirically inadequate and theoretically undesirable. In this final chapter, I sketched the outlines of a positive theory of extended virtues in response to these shortcomings. I attempted to describe the various ways in which the processes relevant to, and the bearers of, moral and intellectual virtues can be embedded/scaffolded/extended. I then argued that metaphysically speaking, virtues are best understood as relations (rather than dispositions) located at the intersection (rather than within an agent) among agents, other agents, and non-social features of the environment. I concluded by suggesting that ecological metaphors in general, and stewardship in particular, provide the most promising framework for conceptualizing extended virtues.

There are (at least) two interpretations of the evidence and arguments presented in this dissertation. The weaker interpretation holds something like the following. At the least, the burden of proof is now shifted to defenders of the claim that all of the processes and states relevant to virtue ethics are internalist and individualist. That is, if my arguments have gone through, then internalism and individualism are not the only games in town in empirically informed virtue theory. It is no longer tenable to assert without argument that the cognitive and affective processes relevant to virtues ethics are internal to individuals. This weaker interpretation may be willing to grant an *epistemological* thesis about moral agents, but it will likely deny any *ontological* thesis. That is, it might be the case that there is some explanatory benefit to *thinking* about processes relevant to,

and bearers of, moral and intellectual virtues as embedded/scaffolded/extended, but this is not to say, for example, that agents *really, literally, ontologically* are extended to include various parts of the environment.

I hope to have provided abundant evidence to defend this weaker interpretation. I think there can be no question that there are legitimate and plausible *claims* that processes relevant to, and bearers of, moral and intellectual virtues are not solely internalistic or individualistic. It would be irresponsible and inconsistent to ignore these claims while upholding a commitment to empirically responsible moral philosophy.

The strong interpretation on the other hand, claims that *most* virtues are scaffolded or extended in the manner described here in Sections IV.II-III, and thus, generally internalist and/or individualist accounts of virtue are false and should be rejected. And if this right, perhaps virtues like gratitude, friendship, trust, and love, which are essentially social and relational, are better paradigms for virtue theory *writ large* than virtues like courage, which conjure up images of singular inner strength and fortitude. In this chapter, I proposed the *hypotheses* of embedded/scaffolded/extended virtues in the strict sense of that term – as constructs to be tested, verified, and ultimately assessed with respect to their bearing on moral philosophy and psychology. This is my prescription for future research in empirically informed virtue theory, and this is how the adequacy of the strong interpretation will ultimately be settled.

The stronger interpretation may also be friendlier to *ontological* theses about moral agents in light of extended mind and virtue arguments. Namely, that “such agents are genuinely of their worlds, and not simply in them. They are not helpless bystanders

watching the passing show from behind a fixed veil of sensing, acting, and representing, but the active architects of their own bounds and capacities.” (Clark, 2007a, p. 280).

In the end, both the weak and strong interpretations entail that unreflective internalism and individualism in virtue ethics is no longer a viable option. It is both empirically inadequate and theoretically unattractive to think of moral agents as Hobbesian mushrooms or Kantian weightlifters. Clark’s view of agents as not just *in* but *of* the world, and ecological metaphors like stewardship, then, are prime candidates to replace these outdated views. And given our increasing interdependence with various information technologies – not to mention the increasing awareness of our profound interdependence with various microorganisms - I would wager that something like the picture painted by Clark will turn out to be among the most fecund accounts of agency available to us. If the present project has laid the groundwork for constructing a normative ethical theory that does justice to this view, then it will have been a success.

REFERENCES CITED

- Abel, C. (2014). *The Extended Self: Architecture, Memes and Minds*. Manchester University Press.
- Adams, F. & Aizawa, K. (2008). *The Bounds of Cognition*. Blackwell.
- Adams, F. & Aizawa, K. (2009). Why the mind is still in the head. In Murat Aydede & P. Robbins (eds.), *The Cambridge Handbook of Situated Cognition*. Cambridge University Press, 78-95.
- Adams, F. & Aizawa, K. (2010). Defending the bounds of cognition. In R. Menary (ed.), *The Extended Mind*. MIT Press, 67-80.
- Alexander, T. (1987). *John Dewey's Theory of Art, Experience, and Nature: The Horizons of Feeling*. Albany: State University of New York Press.
- Alexander, T. & Hickman, L. (1998). Introduction. In *The Essential Dewey: Ethics, Logic, Psychology, Vol 2*. Bloomington: Indiana University Press.
- Alfano, M. (2012). *Character as Moral Fiction*. Cambridge University Press.
- Alfano, M. (2014). What are the bearers of virtues? In H. Sarkissian and J. C. Wright (Eds.), *Advances in experimental moral psychology*, 73–90. New York: Continuum.
- Alfano, M. (2015). Friendship and the structure of trust. In J. Webber and A. Masala (Eds.), *From personality to virtue: Essays in the psychology and ethics of character*, 186–206. Oxford University Press.
- Alfano, M. & Skorburg, J.A. (2016). The embedded and extended character hypotheses. In Julian Kiverstein (ed.), *Philosophy of the Social Mind*. Routledge. pp. 465-778.
- Alfano, M. & Skorburg, J.A. (forthcoming). Extended knowledge, the recognition heuristic, and epistemic injustice. In D. Pritchard et al. (eds.), *Extended Knowledge*. Oxford University Press.
- Allport, G. W. (1937). *Personality: A Psychological Interpretation*. New York: Holt Rinehart & Winston.
- Anderson, B., Ho, J., Brackett, J., Finkelstein, D., & Laffel, L. (1997). Parental involvement in diabetes management tasks: relationships to blood glucose monitoring adherence and metabolic control in young adolescents with insulin-dependent diabetes mellitus. *The Journal of Pediatrics*, 130(2), 257-265.
- Annas, J. (1993). *The morality of happiness*. Oxford University Press.
- Anscombe, G. E. M. (1958). Modern Moral Philosophy. *Philosophy*, 33(124), 1-19.
- Bach-y-Rita, P., & Kercel, S. W. (2003). Sensory substitution and the human–machine interface. *Trends in cognitive sciences*, 7(12), 541-546.

- Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, 723–767.
- Barnes, J. (Ed.). (2014). *The complete works of Aristotle: The revised Oxford translation, one-volume digital edition*. Princeton University Press.
- Baron, K. G., Smith, T. W., Berg, C. A., Czajkowski, L. A., Gunn, H., & Jones, C. R. (2011). Spousal involvement in CPAP adherence among patients with obstructive sleep apnea. *Sleep and Breathing*, 15(3), 525-534.
- Bateson, G. (1972). *Steps to an ecology of mind*. New York: Balentine Books.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74, 1252-1265.
- Benhabib, S. (1986). The generalized and the concrete other: The Kohlberg-Gilligan controversy and feminist theory. *Praxis international*, 5(4), 402-424.
- Berg, C. A., Butner, J. E., Butler, J. M., King, P. S., Hughes, A. E., & Wiebe, D. J. (2013). Parental persuasive strategies in the face of daily problems in adolescent type 1 diabetes management. *Health Psychology*, 32(7), 719.
- Bernecker, S. (2014). How to understand the extended mind. *Philosophical Issues*, 24(1), 1-23.
- Best, J. R., Goldschmidt, A. B., Mockus-Valenzuela, D. S., Stein, R. I., Epstein, L. H., & Wilfley, D. E. (2016). Shared weight and dietary changes in parent–child dyads following family-based obesity treatment. *Health Psychology*, 35(1), 92.
- Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139-159.
- Buller, T. (2006). Brains, lies, and psychological explanations. In J. Illes (ed.), *Neuroethics: Defining the issues in theory, practice and policy*. Oxford University Press, 51-60.
- Buller, T. (2013). Neurotechnology, Invasiveness and the Extended Mind. *Neuroethics* 6(3), 593-605.
- Caldwell, C., & Hayes, L. A. (2010). Leadership, trustworthiness, and ethical stewardship. *Journal of Business Ethics*, 96(4), 497-512.
- Caldwell, C., Truong, D. X., Linh, P. T., & Tuan, A. (2011). Strategic human resource management as ethical stewardship. *Journal of business ethics*, 98(1), 171-182.
- Cantor, N. & Kihlstrom, J. (1987). *Personality and Social Intelligence*. New Jersey: Prentice Hall.

- Carter, J. A., Gordon, E. C., & Palermos, S. O. (2016). Extended emotion. *Philosophical psychology*, 29(2), 198-217.
- Carter, J. A., & Palermos, S. O. (2016). Is having your computer compromised a personal assault? The ethics of extended cognition. *Journal of the American Philosophical Association*, 2(4), 542-560.
- Cantor, N. & Kihlstrom, J. (1987). *Personality and Social Intelligence*. New Jersey: Prentice Hall.
- Cash, M. (2010). Extended cognition, personal responsibility, and relational autonomy. *Phenomenology and the Cognitive Sciences*, 9(4), 645–671.
- Cash, M. (2013). Cognition without borders: ‘third wave socially distributed cognition and relational autonomy. *Cognitive Systems Research*, 25-26, 61-71.
- Chemero, A. (2003). An outline of a theory of affordances. *Ecological psychology*, 15(2), 181-195.
- Chemero, A. (2011). *Radical Embodied Cognitive Science* (Reprint edition). Cambridge, Mass.: A Bradford Book.
- Colombetti, G., & Roberts, T. (2015). Extending the extended mind: the case for extended affectivity. *Philosophical Studies*, 172(5), 1243-1263.
- Colombetti, G., & Krueger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology*, 28(8), 1157-1176.
- Cook, W. & Kenny, D. (2005). The actor-interdependence model: a model of bidirectional effects in developmental studies. *International Journal of Behavioral Development*, 29(2), 101–109.
- Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT Press.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis* 58(1), 7-19.
- Clark, A. (2007a). Re-inventing ourselves: The plasticity of embodiment, sensing, and mind. *Journal of Medicine and Philosophy*, 32(3), 263-282.
- Clark, A. (2007b). Soft selves and ecological control. In D. Ross, D. Spurrett, H. Kincaid, & G.L. Stephens (eds.) *Distributed cognition and the will: Individual volition and social context*, 101-122. MIT Press.
- Clark, A. (2008). *Supersizing the Mind: Action, embodiment, and cognitive extension*. Oxford University Press.
- Clark, A. (2009). Spreading the joy? Why the machinery of consciousness is (probably) still in the head. *Mind*, 118(472), 963-993.
- Darley, J. & Fazio, R. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, 35(10), 867-881.

- Darley, J., & Latané, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
- Darley, J. & Batson, C.D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27, 100-108.
- Dawkins, R. (1999). *The extended phenotype: The long reach of the gene*. Revised edition. Oxford University Press.
- DeMarco, J. & Ford, P. (2014). Neuroethics and the Ethical Parity Principle. *Neuroethics* 7(3), 317-325.
- Dewey, J. (1896/2008). The reflex arc concept in psychology. In *The Early Works, 1882–1898. The Collected Works of John Dewey, 1882–1953*. Ed. Jo Ann Boydston. Carbondale: Southern Illinois University Press.
- Dewey, J. (1898/1976). *Lectures on Psychological and Political Ethics: 1898*. Ed. Donald F. Koch. New York: Hafner Press.
- Dewey, J. (1901/1991). *Lectures on Ethics: 1900–1901*. Ed. Donald F. Koch. Carbondale: Southern Illinois University Press.
- Dewey, J. (1916/2004). *Essays in Experimental Logic*. Dover.
- DiPaolo, E. (2009). Extended Life. *Topoi*, 28(1), 9–21.
- Doris, J. M. (1998). Persons, situations, and virtue ethics. *Nous*, 32(4), 504-530.
- Doris, J. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge University Press.
- Doris, J. (2015). *Talking to Ourselves: Reflection, Ignorance, and Agency*. Oxford University Press.
- Drigotas, S. M., Rusbult, C. E., Wieselquist, J., & Whitton, S. W. (1999). Close partner as sculptor of the ideal self: behavioral affirmation and the Michelangelo phenomenon. *Journal of personality and social psychology*, 77(2), 293-323.
- Finkel, E.J., Fitzsimons, G.M., & van Dellen, M.R. (forthcoming). Self-Regulation as a transactive process. In K. Vohs & R. Baumeister (eds.) *Handbook of Self-Regulation: Research, Theory, and Applications*. New York: Guilford Press.
- Fitzsimons, G. M., Finkel, E. J., & van Dellen, M. R. (2015). Transactive goal dynamics. *Psychological Review*, 122(4), 648-673.
- Flanagan, O. (2016). *The geography of morals: Varieties of moral possibility*. Oxford University Press.
- Flanagan, O. (1991). *Varieties of moral personality: Ethics and psychological realism*. Harvard University Press.

- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of school psychology, 53*(1), 25-44.
- Franks, M. M., Shields, C. G., Lim, E., Sands, L. P., Mobley, S., & Boushey, C. J. (2012). I will if you will: Similarity in married partners' readiness to change health risk behaviors. *Health Education & Behavior, 39*(3), 324-331.
- Friedman, M. (1997). Autonomy and social relationships: Rethinking the feminist critique. In D. T. Meyers (Ed.), *Feminists rethink the self*. Boulder, CO: Westview Press.
- Friedman, M. (2000). Feminism in ethics: Conceptions of autonomy. In M. Fricker & J. Hornsby (Eds.), *The Cambridge companion to feminism in philosophy*. Cambridge University Press.
- Froese, T., & Fuchs, T. (2012). The extended body: A case study in the neurophenomenology of social interaction. *Phenomenology and the Cognitive Sciences, 11*(2), 205–235.
- Funder, D. (2006). Towards a resolution of the personality triad: persons, situations, and behaviors. *Journal of Research in Personality, 40*, 21-34.
- Funder, D. & Ozer, D. (1983). Behavior as a function of the situation. *Journal of Personality and Social Psychology, 44*(1), 107-112.
- Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research, 25-26*, 4–12.
- Gallagher, S. (2014). Pragmatic Interventions into Enactive and Extended Conceptions of Cognition. *Philosophical Issues, 24*, 110-126.
- Gallagher, S., & Crisafi, A. (2009). Mental institutions. *Topoi, 28*(1), 45–51.
- Gazzaley, A., & Rosen, L. D. (2016). *The Distracted Mind: Ancient Brains in a High-Tech World*. MIT Press.
- Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
- Griffiths, P. E., & Scarantino, A. (2005). Emotions in the wild: The situated perspective on emotion. In P. Robbins & M. Aydede (Eds.) *The Cambridge Handbook of Situated Cognition*, pp. 437-454. Cambridge University Press.
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.
- Harris, C, Keil, P., Sutton, J., Barnier, A., & McIlwain, D. (2011). We remember, we forget: collaborative remembering in older couples. *Discourse Processes, 48*(4), 267-303.

- Heersmink, R. (2015). Dimensions of Integration in Embedded and Extended Cognitive Systems. *Phenomenology and the Cognitive Sciences* 13(3), 577-598.
- Heersmink, R. (forthcoming). Extended mind and cognitive enhancement: moral aspects of cognitive artifacts. *Phenomenology and the Cognitive Sciences*.
- Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Mahwah, NJ: Erlbaum.
- Hemphill, R. C., Martire, L. M., Polenick, C. A., & Stephens, M. A. P. (2016). Spouse confidence and physical function among adults with osteoarthritis: the mediating role of spouse responses to pain. *Health Psychology*, 35(10), 1059-1068.
- Hobbes, T. (1658/1991). *Man and Citizen*. Indianapolis: Hackett
- Holt, E.B. (1931). *Animal drive and the learning process: An essay toward radical empiricism*. New York: Henry Holt and Company.
- Howell, R.J. (2016). Extended virtues and the boundaries of persons. *Journal of the American Philosophical Association* 2(1), 146-163.
- Huebner, B. (2016). Transactive Memory Reconstructed: Rethinking Wegner's Research Program. *Southern Journal of Philosophy*, 54(1), 48-69.
- Huebner, B. (2013). Socially embedded cognition. *Cognitive Systems Research*, 25-26, 13-18.
- Hurley, S. (2010). Varieties of externalism. In Richard Menary (ed.), *The Extended Mind*, 101-153. MIT Press.
- Hursthouse, R. (1999). *On virtue ethics*. Oxford University Press.
- Hutchins, E. (2010). Cognitive ecology. *Topics in cognitive science*, 2(4), 705-715.
- Ickes, W., Snyder, M., & Garcia, S. (1997). Personality influence on the choice of situations. In R. Hogan, J. Johnson, & S. Briggs (eds.), *Handbook of personality psychology*. New York: Academic Press, 165-195.
- James, W. (1981/1890). *The Principles of Psychology* (2 vols). Cambridge, MA:Harvard University Press.
- James, W. (1976/1912). *Essays in radical empiricism*. Cambridge, MA: Harvard University Press.
- Jansen, L. A. (2013). Between beneficence and justice: the ethics of stewardship in medicine. *Journal of Medicine and Philosophy*, 38(1), 50-63.
- Jayawickreme, E., & Chemero, A. (2008). Ecological moral realism: An alternative theoretical framework for studying moral psychology. *Review of General Psychology*, 12(2), 118-126.
- Johnson, M. (1994). *Moral imagination: Implications of cognitive science for ethics*. University of Chicago Press.

- Jones, E.E. (1986). Interpreting interpersonal behavior: the effects of expectancies. *Science*, 234(4772), 41-46.
- Kant, I. (1994). *Ethical philosophy: the complete texts of Grounding for the metaphysics of morals, and Metaphysical principles of virtue, part II of The metaphysics of morals, with On a supposed right to lie because of philanthropic concerns*. Indianapolis: Hackett Publishing.
- Kass, L. (2003). *Beyond therapy: Biotechnology and the pursuit of happiness*. New York: HarperCollins.
- Kashy, D. A., & Kenny, D. A. (1999). The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology*. New York: Cambridge University Press.
- Kenny, D. A. (1996). Models of nonindependence in dyadic research. *Journal of Social and Personal Relationships*, 13, 279–294.
- Kihlstrom, J. (2013). The person-situation interaction. In D. Carlston (ed.) *The Oxford Handbook of Social Cognition*. Oxford University Press, 786-806.
- Kirchhoff, M. (2015). Composition and transactive memory systems. *Philosophical Explorations*. <http://dx.doi.org/10.1080/13869795.2016.1085593>
- Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science* 18(4), 513-549.
- Kittay, E. F. (1999). *Love's labor: Essays on women, equality and dependency*. Routledge.
- Krueger, J. & Szanto, T. (forthcoming). Extended emotions. *Philosophy Compass*.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Lanphear, B., Hornung, R., Khoury, K., Baghurst, P., Bellinger, D., Canfield, R., Dietrich, K., Bornschein, R., Greene, T., Rothenberg, S., Needleman, H., Schnaas, L., Wasserman, G., Graziano, J., & Roberts, R. (2005). Low-level environmental lead exposure and children's intellectual function: An international pooled analysis. *Environmental Health Perspectives*, 113(7), 894-9.
- Laurin, K., Fitzsimons, G. M., Finkel, E. J., Carswell, K. L., vanDellen, M. R., Hofmann, W., & Brown, P. C. (2016). Power and the pursuit of a partner's goals. *Journal of personality and social psychology*, 110(6), 840.
- Levy, N. (2007a). Rethinking Neuroethics in the Light of the Extended Mind Thesis. *American Journal of Bioethics* 7(9), 3-11.
- Levy, N. (2007b). *Neuroethics: challenges for the 21st century*. Oxford University Press.
- Lewin, K. (1936/2015). *Principles of Topological Psychology*. (F. Heider, trans.). Martino Fine Books.

- Lewin, K. (1939/1951). Field theory and experiment in social psychology: concepts and methods. In K. Lewin (ed.), *A Dynamic Theory of Personality*. New York: Harper & Row.
- Lewin, K. (1951). *Field Theory in Social Science*. New York: Harper.
- Lewis, K. & Herndon, B. (2011). Transactive Memory Systems: Current Issues and Future Research Directions. *Organization Science*, 22(5), 1254-1265.
- Lindemann, H. (2013). *Holding and letting go: The social practice of personal identities*. Oxford University Press.
- Lindemann, H. (2010). In a Time of Dementia. In Eva Feder Kittay & Licia Carlson (eds.), *Cognitive Disability and its Challenge to Moral Philosophy*. Wiley-Blackwell.
- Locke, J., (1690/1975). *Essay Concerning Human Understanding*, P.H. Nidditch (ed.), Oxford: Oxford University Press.
- Loughlin, V. (2013). Sketch this: Extended mind and consciousness extension. *Phenomenology and the Cognitive Sciences*, 12(1), 41-50.
- MacIntyre, A. (2007). *After virtue: A study in moral theory*, 3rd ed. South Bend: University of Notre Dame Press.
- Mackenzie, C., & Stoljar, N. (2000). Introduction, in C. Mackenzie & N. Stoljar (eds.), *Relational autonomy: Feminist perspectives on autonomy, agency, and the social self*. Oxford University Press.
- Mackenzie, C., W. Rogers, & Dodds, S. (eds.) (2014). *Vulnerability: New essays in ethics and feminist philosophy*. Oxford University Press.
- Madzia, R. (2013). Chicago Pragmatism and the Extended Mind Theory: Mead and Dewey on the Nature of Cognition. *European Journal of Pragmatism and American Philosophy*, 5(1), 279-297.
- Matthews, G. (1980). Ritual and religious feelings. In A. Rorty (ed.) *Explaining Emotions*, 339-53. Berkeley: University of California Press.
- Malzkorn, W. (2000). Realism and the conditional analysis of dispositions. *The Philosophical Quarterly* 50, 152–69
- Martz, J. M., Verette, J., Arriaga, X. B., Slovik, L. F., Cox, C. L., & Rusbult, C. E. (1998). Positive illusion in close relationships. *Personal Relationships*, 5, 159–181.
- McClelland, T. (2017). The mental affordance hypothesis. Presented at the Minds Online Conference, September 9, 2017.
<http://mindsonline.philosophyofbrains.com/2017/2017-session-1/the-mental-affordance-hypothesis/>

- Mead, G.H. (1936/1967). *Mind, Self, and Society from the Standpoint of a Social Behaviorist*. University of Chicago Press.
- Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.
- Menary, R. (2010a). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, 9, 459–463.
- Menary, R. (2010b). Introduction. In Richard Menary (ed.) *The extended mind*. MIT Press.
- Meyers, D. T. (1998). Agency. In A. M. Jaggar & I. M. Young (Eds.), *A companion to feminist philosophy*. Malden, MA: Blackwell.
- Meyers, D. T. (1997). Introduction, in D.T. Meyers (ed.) *Feminists rethink the self*. Boulder, CO: Westview Press.
- Miller, C. (2013). *Moral character: An empirical theory*. Oxford University Press.
- Miller, C. (2014). *Character and moral psychology*. Oxford University Press.
- Mischel, W. (1965). Predicting the success of Peace Corps volunteers in Nigeria. *Journal of Personality and Social Psychology*, 1(5), 510-517.
- Mischel, W. (1968). *Personality and Assessment*. John Wiley & Sons.
- Mischel, W. (1974). Processes in delay of gratification. In L. Berkowitz (ed.), *Advances in experimental social psychology*. New York: Academic Press, pp.249-292.
- Mischel, W. (2009). From Personality and Assessment (1968) to Personality Science, 2009. *Journal of Research in Personality*, 43, 282-290.
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. Harper & Row.
- Mumford, S. (1998). *Dispositions*. New York: Oxford University Press
- Murray, S. L., & Holmes, J. G. (1997). A leap of faith? Positive illusions in romantic relationships. *Personality and Social Psychology Bulletin*, 23, 586–604.
- Murray, S. L., Holmes, J.G., & Griffin, D. W. (1996a). The self-fulfilling nature of positive illusions in romantic relationships: Love is not blind, but prescient. *Journal of Personality and Social Psychology*, 71(6), 1155-1180.
- Murray, S. L., Holmes, J.G., & Griffin, D. W. (1996b). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality and Social Psychology*, 70, 79–98.
- Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford University Press.
- Nietzsche, F. (1874/1997). *Untimely Meditations*. Trans. R. J. Hollingdale. Cambridge University Press.

- Nussbaum, M. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.
- Noë, A. (2009). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness*. New York: Macmillan.
- Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: the neglected process in evolution*. Princeton University Press.
- Olson, E. (2011). The Extended Self. *Minds and Machines*, 21(4), 481-495.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), 1-8.
- Palermos, S. O. (2014). Loops, Constitution and Cognitive Extension. *Cognitive Systems Research*, 27, 25-41.
- Palermos, S. O., & Pritchard, D. (2013). Extended knowledge and social epistemology. *Social Epistemology Review and Reply Collective*, 2(8), 105-120.
- Parens, E. (2006). Creativity, gratitude, and the enhancement debate. In J. Illes (ed.), *Neuroethics: Defining the issues in theory, practice and policy*. Oxford University Press, 75-86.
- Peirce, C.S. (1903/1998). Three normative sciences. *The Essential Peirce: Selected Philosophical Writings vol. 2 (1893-1913)*. Indiana University Press, 196-207.
- Pew Research Center. (January 12, 2017). "Mobile fact sheet." <http://www.pewinternet.org/fact-sheet/mobile/>
- Pew Research Center. (September 7, 2017). "News Use Across Social Media Platforms 2017." <http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/>
- Pew Research Center. (June 28, 2017). "10 facts about smartphones as the iPhone turns 10." <http://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/>
- Pöyhönen, S. (2014). Explanatory power of extended cognition. *Philosophical Psychology* 27(5), 735-759.
- Prinz, J. (2007). *The Emotional Construction of Morals*. Oxford University Press.
- Prior, E. W., Pargetter, R., & Jackson, F. (1982). Three theses about dispositions. *American Philosophical Quarterly*, 19(3), 251-257.
- Pritchard, D., Kallestrup, J., Palermos, O., & Carter, J. A. (Eds.) (forthcoming). *Extended Knowledge*. Oxford University Press.
- Putnam, H. (1975). The meaning of meaning. In *Mind, language, and reality: Philosophical papers, vol. 2*, pp 215-271. Cambridge University Press.

- Rauers, A., Riediger, M., Schmiedek, F., & Lindenberger, U. (2011). With a little help from my spouse: Does spousal collaboration compensate for the effects of cognitive aging?. *Gerontology*, *57*(2), 161-166.
- Reed, E. S. (1996). *Encountering the world*. New York: Oxford University Press.
- Ren, Y. & Argote, L. (2011). Transactive Memory Systems 1985–2010: An Integrative Framework of Key Dimensions, Antecedents, and Consequences. *The Academy of Management Annals* *5*(1), 189–229.
- Richard, F.D., Bond, C.F., & Stokes-Zoota, J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology* *7*(4), 331-363.
- Roberts, B. (2009). Back to the future: *Personality and Assessment* and personality and development. *Journal of research in personality* *43*, 266.
- Roberts, R. C. (2013). *Emotions in the moral life*. Cambridge University Press.
- Roberts, R. C., & Wood, W. J. (2007). *Intellectual virtues: An essay in regulative epistemology*. Oxford University Press.
- Rockwell, W.T. (2005). *Neither Brain nor Ghost: A Non-Dualist Alternative to the Mind-Brain Identity Theory*. Cambridge: MIT Press.
- Ross, D & Ladyman, J. (2010). The Alleged Coupling-Constitution Fallacy and the Mature Sciences. In Richard Menary (ed.), *The Extended Mind*. MIT Press.
- Rowlands, M. (2003). *Externalism: Putting Mind and World Back Together Again*. McGill-Queens University Press.
- Rowlands, M. (2013). *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*, MA: Massachusetts Institute of Technology Press.
- Rudder-Baker, L. (2009). Persons and the extended mind thesis. *Zygon*, *44*(3), 642-658.
- Rupert, R. (2004), "Challenges to the Hypothesis of Extended Cognition," *Journal of Philosophy*, *101*, pp. 389-428.
- Rupert, R. (2009). *Cognitive systems and the extended mind*. Oxford University Press.
- Rusbult, C. E., Finkel, E. J., & Kumashiro, M. (2009). The Michelangelo phenomenon. *Current Directions in Psychological Science*, *18*(6), 305-309.
- Russell, D. C. (2009). *Practical intelligence and the virtues*. Oxford University Press.
- Savulescu, J. (2005). New breeds of humans: the moral obligation to enhance. *Reproductive BioMedicine Online*, *10*, 36-39.
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*, 1–20.
- Skorburg, J.A. (2013). Beyond embodiment: John Dewey and the integrated mind. *The Pluralist*, *8*(3), 66-78.

- Skorburg, J.A. (2017). Lessons and new directions for extended cognition from social and personality psychology. *Philosophical Psychology*, 1-20.
<http://dx.doi.org/10.1080/09515089.2017.1282606>
- Skorburg, J.A. & Alfano, M. (forthcoming). Implications for virtue epistemology from psychological science: intelligence as an interactionist virtue. In H. Battaly (ed.), *Handbook of Virtue Epistemology*. New York: Routledge.
- Slote, M. (2003). *Morals from motives*. Oxford University Press.
- Smith, A. (1977). Dispositional properties, *Mind* 86, 439–45.
- Sneddon, A. (2011). *Like-minded: Externalism and moral psychology*. MIT Press.
- Sneddon, A. (2008). The depths and shallows of psychological externalism. *Philosophical Studies*, 138(3), 393–408.
- Snow, N. E. (2010). *Virtue as social intelligence: An empirically grounded theory*. New York: Routledge.
- Srivastava, S., McGonigal, K., Richards, J., Butler, E., & Gross, J.J. (2006). Optimism in close relationships: how seeing things in a positive light make them so. *Journal of Personality and Social Psychology*, 91(1), 143-153.
- Steele, C. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Stephan, A., Walter, S., & Wilutzky, W. (2014). Emotions beyond brain and body. *Philosophical Psychology*, 27(1), 65-81.
- Sterelny, K. (2006). Cognitive load and human decision, or, three ways of rolling the rock uphill". In P. Carruthers, S. Laurence, and S. Stich (Eds.) *The innate mind: Culture and cognition*, pp. 218-233. Oxford University Press.
- Sterelny, K. (2010). Minds: Extended or scaffolded?. *Phenomenology and the Cognitive Sciences*, 9(4), pp. 465-481.
- Sutton, J. (2010). Exograms and interdisciplinarity: History, the extended mind and the civilizing process. In R. Menary (ed.) *The Extended Mind*. MIT Press, 189-225.
- Sutton, J., Harris, C. B., Keil, P., & Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences*, 9,521–560.
- Straehle, C., (ed). (2016). *Vulnerability, autonomy, and applied ethics*. New York: Routledge.
- Swann, W. & Seyle, C. (2005). Personality psychology's comeback and its emerging symbiosis with social psychology. *Personality and Social Psychology Bulletin* 31(2), 155-165.
- Swanton, C. (2003). *Virtue ethics: A pluralistic view*. Oxford University Press.

- Taylor, E. (2011). William James's radical empiricism: Did E.B. Holt get it right? In E.P. Charles (Ed.), *A new look at New Realism: The psychology and philosophy of E.B. Holt*. pp. 105-126. New Brunswick: Transactions Publishers.
- Tessman, L. (2005). *Burdened virtues: Virtue ethics for liberatory struggles*. Oxford University Press.
- Theiner, G. (2013) Transactive Memory Systems: A Mechanistic Analysis of Emergent Group Memory. *Review of Philosophy and Psychology* 4(1), 65-89.
- Tollefson, D., Dale, R., & Paxton, A. (2013). Alignment, transactive memory, and collective cognitive systems. *Review of Philosophy and Psychology*, 4(1), 49–64.
- Turvey, M. (1992). Affordances and prospective control: An outline of the ontology. *Ecological Psychology*, 4, 173–187.
- Twenge, J. M. (2017). *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy--and Completely Unprepared for Adulthood--and What That Means for the Rest of Us*. New York: Simon and Schuster.
- Upton, C. L. (2009). *Situational traits of character: Dispositional foundations and implications for moral psychology and friendship*. Lexington Books.
- Vaesen, K. (2014). Dewey on extended cognition and epistemology. *Philosophical Issues*, 24, 426-438.
- Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Washington, N. (2015). *Mental health and human minds: Some theoretical criteria for clinical psychiatry* (doctoral dissertation, Purdue University). Retrieved from: <https://search.proquest.com/docview/1777617026?pq-origsite=gscholar>
- Wagman, J. B., & Chemero, A. (2014). The end of the debate over extended cognition. In *Neuroscience, Neurophilosophy and Pragmatism* (pp. 105-124). Palgrave Macmillan UK.
- Wegner, D. (1995). A computer network model of human transactive memory. *Social Cognition* 13(3), 319–339.
- Wegner, D. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen, G. R. Goethals, (eds.) *Theories of Group Behavior*. New York: Springer-Verlag, 185–208.
- Wegner, D., Erber, R, & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, 61(6), 923–929.
- Wegner, D., Giuliano, T, Hertel, P. (1985). Cognitive interdependence in close relationships. In W. Ickes, (ed.) *Compatible and Incompatible Relationships*. New York: Springer-Verlag, 253–276.
- Welchman, J. (2012). A defence of environmental stewardship. *Environmental Values*, 21(3), 297-316.

- Wheeler, M. (2015). Extended consciousness: an interim report. *The Southern Journal of Philosophy*, 53(S1), 155-175.
- Williams, B. (1985). *Ethics and the limits of philosophy*. London: Fontana Press.
- Wilson, S. J., Martire, L. M., & Sliwinski, M. J. (2017). Daily Spousal Responsiveness Predicts Longer-Term Trajectories of Patients' Physical Function. *Psychological Science*, 28(6), 786-797.
- Wilson, R. A. (2004). *Boundaries of the mind: The individual in the fragile sciences: Cognition*. Cambridge University Press.
- Wilson, R. A., & Clark, A. (2009). How to Situate Cognition: Letting Nature Take its Course. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition*. Cambridge University Press, 55–77.
- Wilson, R.A. & Lenart, B. (2014). Extended mind and identity. In Jens Clausen & Neil Levy (eds.), *Handbook of Neuroethics*, 423-439. Springer.
- Wagman, J. & Chemero, A. (2014). The end of the debate over extended cognition. In T. Solymosi and J. Shook (eds.) *Neuroscience, Neurophilosophy, and Pragmatism: Brains at work with the world*. Palgrave.
- Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.
- Zawidzki, T. (2012). Trans-human cognitive enhancement, phenomenal consciousness, and the extended mind. *International Journal of Machine Consciousness*, 4(1), 215-227.