

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Virtue and Empirical Psychology

Permalink

<https://escholarship.org/uc/item/43n6r8w2>

Author

Robertson, Kyle Stuart

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

VIRTUE AND EMPIRICAL PSYCHOLOGY

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

PHILOSOPHY

by

Kyle S. Robertson

June 2015

The Dissertation of Kyle Robertson is
approved:

Professor Daniel Guevara, chair

Professor Jonathan Ellis

Professor John Bowin

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Kyle Robertson

2015

Table of Contents

1. Introduction.....	1
1.1 Key Concepts for Virtue Ethics.....	5
1.2 Virtue Ethics and Naturalism.....	14
1.3 Virtue Ethics and Empirical Psychology.....	17
2. Empirical Investigations of Virtue: Situationism, Positive Psychology, and Mindfulness.....	23
2.1 Lack of Character: Doris' Situationist Challenge.....	26
2.2 Replies to Doris in the Current Literature	34
2.2.1 Sabini & Silver - The Problem of Vicious Moral Psychology	35
2.2.2 Sreenivasan - Operationalizing Virtue.....	39
2.2.3 Annas	43
2.3 First Reply to Doris: Concepts of Virtue	46
2.3.1 The Competence-Performance Distinction.....	46
2.3.2 The Inconsistency of Vice	55
2.4 Second Reply to Doris: Positive Psychology	62
2.5 Conclusion	70
3. A Virtue Ethics Take on Situationism: Mindfulness as a Virtue.....	76
3.1 Doris: Automatic and Controlled Thinking.....	77
3.2 Mindfulness as a Solution: Implicit Bias.....	85
3.3 Mindfulness as a Corrective to Implicit Bias	90
3.4 The Benefits of Mindfulness	108
3.5 Conclusion	118
4. Humility as a Virtue.....	122
4.1 The Concept of Humility	123
4.1.1 The Literature on the Puzzle.....	125
4.1.2 A Solution: The Skill Analogy Argument	132
4.2 Humility v. Modesty.....	141
4.3 Empirical Humility: Motivated Reasoning and Cognitive Dissonance.....	145
4.3.1 Humility Benefits its Possessor	147
4.3.2 Humility Benefits Others.....	152
4.3.3 Humility Corrects for a Characteristic Human Weakness.....	158
4.4 Conclusion	162
5. Conclusion	167

List of Figures:

1. Milgram Experiment Diagram.....	82
2. Numeracy v. Correct Interpretation of Data (Kahan et. al. 2013)	160

Abstract for Virtue and Empirical Psychology
by Kyle Robertson

This dissertation analyzes the relationship between empirical social psychology and the philosophical tradition of virtue ethics. It argues that social psychology is conceptually important for virtue ethics, that their relationship has been misunderstood by some prominent philosophers, and that a wide variety of evidence from social psychology is useful and inspirational for virtue ethics. First, it offers several novel criticisms of a well-known critique of virtue ethics. This critique claims that the body of evidence in situationist psychology shows that humans are incapable of possessing general dispositions of characters, of which virtues are a subset, and that therefore virtue ethics is empirically inadequate. This critique is mistaken because it fails to distinguish between moral competence and moral performance, it fails to appreciate the difference between the moral psychology of virtue and that of vice, and it fails to acknowledge the wide breadth of psychological evidence that paints a more nuanced picture of human capacity to possess a general character disposition. Second, this dissertation takes some of the evidence grounding the situationist critique and analyzes it within the virtue ethics framework. This evidence, along with historical traditions valuing mindfulness in Stoicism and non-Western philosophy, suggests that virtue ethics ought to entertain mindfulness as a potential virtue. Rather than grounding an argument against the possibility of virtue, this evidence can be used to help illuminate a key aspect of practical wisdom. Third, this dissertation takes further topics in social psychology, unrelated to the situationist critique, as evidence for the importance of the disposition of humility.

Dedication:

To my love, Olivia, who has taught me more about human flourishing than I can
grapple with in a lifetime of philosophy.

1. Introduction

In her essay “Modern Moral Philosophy”, Elizabeth Anscombe famously claimed that “it is not profitable for us at present to do moral philosophy; that should be laid aside at any rate until we have an adequate philosophy of psychology” (Anscombe 1958, 26). She seems to have been speaking about a philosophical account of important psychological concepts, such as “‘action,’ ‘intention,’ ‘pleasure,’ ‘wanting.’” (Anscombe 1958, 40). This is a problem she worked on herself, most importantly in her book *Intention*. A main thesis of this dissertation is: we should follow Anscombe’s advice and at least place moral psychology on a par with moral philosophy, and we will be unable to do so unless and until we grapple with the wealth of empirical psychological research. While Anscombe herself showed the power of the so-called ‘armchair’ approach to a philosophy of psychology, I am skeptical that we can rise to the challenge she set based on introspection and reasoning alone. An adequate philosophy of psychology must engage with, or at least be consistent with, the results of empirical research.

Rather than discuss moral philosophy in general, this dissertation focuses on virtue ethics, one of the three major schools of ethical thought in contemporary moral philosophy. Put simply, virtue ethics is a theoretical stance that prioritizes human flourishing and happiness as the goal of human life. It claims that a necessary means, both instrumental and constitutive, for achieving flourishing and happiness is cultivating the virtues, which are enduring dispositions of character. They move their possessor to act in response to certain circumstances and reasons, which are different

for and define each virtue, in the right way, for the right reason, and with the right perception and judgment. Courage, for example, is the virtue in regard to fear. The virtues are also general dispositions—to have courage, one must consistently respond rightly to fear in all circumstances. This focus on human character makes virtue ethics better situated, in comparison to the other major theories, to engage with, learn from, and critique scientific work on the nature of moral decision-making and moral action. Virtue ethics involves claims about human natural capacities as part of its theory of goodness, whereas such natural capacities do not figure largely into the theories of deontological or utilitarian ethics.

I am not the first to claim such a potential affinity, but virtue ethics and modern empirical psychology have had a relatively fraught relationship in contemporary philosophy. On the one hand, research in situationist psychology grounds an influential critique of virtue ethics (Doris 2002). This critique is the focus of Chapter 2, but briefly, the situationist experiments are purported to show that human action is determined much more, or even exclusively, by situational pressures rather than personal character traits. For example, we will shock a stranger under orders from an authority figure, ignore someone in need if we are in a hurry, and be more likely to help a stranger if we find a dime in a payphone coin return slot. Some philosophers have taken this research as evidence that humans are incapable of possessing virtue, and that virtue ethics is therefore in error. These arguments have generated a contrary skepticism on the part of some virtue ethicists about the relevance of empirical psychology to ethics. For example, Julia Annas is skeptical

that any of this empirical psychology is settled enough to be useful at all to philosophers (see, e.g. Annas 2005). On the other hand, some psychologists have embraced virtue ethics as a model for positive character development (Peterson and Seligman 2004; Haidt 2012, 441 n. 68), and some philosophers have engaged with this work with cautious optimism (see, e.g., Tiberius and Plakias 2010).

This dissertation considers this literature and attempts to stake out a position reconciling virtue ethics with empirical psychology. In particular, it defends the view that the empirical psychology is critically important for, rather than fatal to, to virtue ethics. Empirical psychology can set limits on the appropriate scope for virtue ethics, though those limits are broader and more permissive than the situationist critique suggests. More importantly, empirical psychology can serve as an inspiration for virtue ethicists examining issues about what character traits are virtues, how those traits are expressed, how those traits are thwarted, and many other questions in virtue ethics. This dissertation particularly engages with the first question, about what character traits are virtues, by arguing that there is strong empirical evidence in favor of considering humility and mindfulness to be important virtues.

The first chapter of this dissertation offers a novel criticism of the situationist psychology argument against virtue ethics. This chapter has two lines of argument. First, the arguments inspired by situationist psychology misconstrue virtue ethics in a number of important ways which weaken the force of Doris's, and other situationist's, claims. Second, situationist arguments ignore relevant, and contrary, empirical evidence about the plausible existence of virtue traits from other areas of psychology,

particularly those reflected in the positive psychology movement. The combination of these two arguments substantially weakens the arguments of philosophers such as John Doris in A Lack of Character. Most importantly, the situationist claims that survive this critique are exactly the sorts of claims that virtue ethicists themselves can endorse, claims about the difficulty and frailty of virtue.

In the next two chapters, this dissertation examines other relevant results from modern experimental psychology and argues that they can enliven debates in virtue ethics about what character traits constitute virtues. Chapter 3 argues that the situationist literature, and other literature in psychology on implicit bias and mindless action, imply that mindfulness should be taken seriously as a virtue. Chapter 4 argues that the extensive literature on forms of motivated reasoning indicate that humility deserves a place on our list of cardinal, or most-important virtues. Both of these chapters integrate discussions about humility and mindfulness as virtues with evidence from empirical psychology in order to argue that the psychology buttresses, and helps clarify, already existing notions of virtue. And both show, I hope, that virtue ethicists and those working with ethics and empirical psychology need not be at odds.

The remainder of this introduction sets the stage for the dissertation by defining some key virtue ethics terms, discussing the role of empirical evidence in virtue theory, and discussing some key issues of statistical inference in the social science research.

1.1 Key Concepts for Virtue Ethics

This dissertation does not try to present a creative or contentious interpretation of classical Aristotelian virtue ethics. The goal is rather to bring mainstream ideas of virtue ethics into an analysis with contemporary social psychology. The idea of virtue presented here, therefore, is unambiguously situated in the recent Aristotelian tradition of virtue ethics, particularly as articulated by scholars such as Julia Annas, John McDowell, Sarah Broadie, Philippa Foot, Robert Adams, and Rosalind Hursthouse. These authors do not present a univocal account of virtue ethics. But the arguments in this dissertation are intended to fall within the area of overlap among all of these authors. I do not intend to take a stance on the issues that define the disagreements among these authors. As a basis for the rest of this dissertation, this section lays out an understanding of a few key terms, and their relationship with each other.

I generally understand *eudaimonia* as human flourishing. It also has the quality of happiness, the common classical translation of *eudaimonia*, but the key feature of *eudaimonia* is that it describes the notion of the best sort of life, of the quality that makes a life choice-worthy, a kind of life we all hope to live. Just as there are many different choice-worthy lives, there are many different ways that *eudaimonia* can play out in different circumstances and contexts. The life of someone in Santa Cruz, CA, today, for example, looks very different from the life of someone born in ancient Athens. The concept of *eudaimonia* does not demand that persons

born in these two disparate communities appear to have similar lives in order to both be *eudaimon*.

Human flourishing is defined by what it is to be human as a matter of logic, for human flourishing is a matter of the flourishing of humans, not of plants or animals. But what it is to be a human is complex and contentious. I think there is general agreement that human flourishing involves some version of a list like: doing things in life we value; enjoyment in doing these things; having rich friendships and family relationships; general experiences and sentiments of well-being; etc. This broad notion of human flourishing includes a wide variety of lives and lifestyles, but according to virtue ethics, all must include possession of the virtues. Broadly understood, virtues are excellences of character, and they exist as consistent, enduring dispositions to do the right thing, along with all of the emotional and rational features that make an action right. Virtue ethicists see these dispositions as necessary for *eudaimonia* for at least two reasons. First, these are the dispositions one needs in order to secure the qualities of a *eudaimon* life; they make one a good friend, a good family member, a good citizen, and the like. Second, these qualities are themselves intrinsically rewarding and valuable. Having a virtue is itself a constituent part of living a life of *eudaimonia*.

Virtues are dispositions of character which include general, non-situation-specific behavioral consistencies. For example, a person who embodies the virtue of courage consistently acts rightly, and for the good, in the face of fear. Virtues also are ways of addressing characteristic human flaws or challenges. Courage is a virtue

because fear is a universal impediment to human life projects. The virtues also exist as an intermediate state between two vices, or extremes. Courage is the mean between cowardice and brashness, for Aristotle, and similar descriptions work for every virtue.¹

Being virtuous is an activity, as Aristotle famously concluded in his function argument, where he says that “the human good proves to be activity of the soul in accord with virtue” (NE 1098a16-18). He compares this to athletic competitions, where “Olympic prizes are not for the finest and strongest, but for the contestants—since it is only these who win—the same is true in life; among the fine and good people, only those who act correctly win the prize” (NE 1099a3-6; Cf. NE 1096a1-3). The activity of virtue is superior to the mere possession of a virtuous state. In particular, if our project is practical, as Aristotle’s is (NE 1103b28-30), we ought to focus on actions over theory. On the other hand, it is also clear that Aristotle considered virtue a part of human nature as a potentiality, not an actuality. In his words,

“And so the virtues arise in us neither by nature nor against nature. Rather, we are by nature able to acquire them, and we are completed through habit. Further, if something arises in us by nature, we first have the capacity for it, and later perform the activity.” (NE 1103a24-26).²

¹ The most difficult classic virtue on this picture is justice, or the disposition to be just, for it is reasonable to construe injustice as involving more than one disposition (Williams 1980, 198). In such a case, justice would not form a single mean between two vices but perhaps an intermediate point between many pairs of vice. Alternately, we could think that the word “justice” captures more than a single virtue. Either option is better, I think, than abandoning the doctrine of the mean.

² See also NE 1106a8-12: “For these reasons the virtues are not capacities either; for we are neither called good nor called bad, nor are we praised or blamed, insofar as we are simply capable of feelings. Further, while we have capacities by nature, we do not become good or bad by nature; we have discussed this before. (Irwin references the block quote in the main text in relation to the “we have

Virtues are not innate dispositions, but we are by nature innately capable of acquiring them. And virtue is within the innate capacity of humans regardless of whether we actualize it in any particular life.

This relates to a number of arguments about virtue theory. For example, for a virtue theory to be empirically adequate, as discussed extensively in Chapter 2, it needs to describe things that are within the power of humans to do, not things that are done by an actual living human. The statement “humans can lift 1000 pounds” is empirically adequate in just the same way that “humans can be courageous” potentially is; there are at most a few people alive who can lift 1000 pounds off the ground, and very recently there were no people alive who had ever lifted 1000 pounds. We should judge the empirical adequacy of virtue claims based on what is in the power of humans to do, not what humans have actually done, though of course in many situations the latter is evidence relevant to the former.

The relationship between virtue and *eudaimonia* justifies some of this dissertation’s investigation into empirical psychology. Evidence that some disposition or habit is systematically associated with human flourishing is evidence that this disposition or habit is like a virtue, or has an important quality of a virtue. Not all things which lead to human flourishing are virtues, of course. As Aristotle observed, human flourishing is influenced by lucky circumstances, and luck is no Aristotelian virtue. But all virtues lead to or are constitutive of human flourishing. And so virtue

discussed this before” passage) Virtues are not *simply* capacities, they are actualized capacities, but they are a part of human nature as a potential even if they are not actualized in any particular agent.

theory suggests that when asking whether disposition X is a virtue, we might try a Popperian *modus tollens*: if X is a virtue, then it is a means to human flourishing; if agents with disposition X systematically fail to flourish, then X is probably not a virtue; and if there is strong evidence that agents with disposition X do systematically flourish, then we have failed to falsify the claim that X is a virtue. This failure to falsify is a confirmation of the claim that X is a virtue. If we repeatedly fail to falsify the claim that X is a virtue, then we ought to seriously consider that claim.

This model of virtue ethics is a sort of ideal theory, where the practical imperative is to act like the ideally virtuous person. This grounds a common criticism of the theory, specifically that it does not provide sufficient guidance for right action. The criticism amounts to something like: the rule “do what the virtuous person would do in any given situation” does not tell us what to do, unless we can find some virtuous person to ask for advice. And even if we do have access to a virtuous person, virtuous people would not get themselves into the everyday pickles in which we find ourselves, and so it is unclear how the rule could guide us even in these cases. This criticism gets virtue theory wrong, as Rosalind Hursthouse has explained clearly (Hursthouse 1999). But the way in which virtue ethics is an ideal theory is complex, and sorting some of its complexity out here helps illustrate the notion of virtue ethics in this dissertation.

Julia Annas’ latest work can provide clarity. There are a number of ways in which the practical imperative of virtue ethics, to be virtuous, can be cashed out. Here are a few for this discussion:

(a): Agent A is virtuous in S when there is nothing A could have done in S that could be considered better than what A actually did. A acted as a virtuous agent would act in S.

(b): Agent A is virtuous when A always acts according to (a) in all situations.

(c): Agent A is virtuous when A has a perfect, complete character that guarantees that A will satisfy (b).

This list is not exhaustive, but it gives a sense of the different ways we might interpret the notion of virtuous action and character. The virtuous agent, in order to be a virtuous agent, acts according to the standard in (a). Annas' key insight is that such an agent aspires to (b) but does not aspire to (c). Or rather, they might aspire to (c) if that were a coherent possibility, but it is not. In fact, (c) is inconsistent with the view of virtue ethics in this dissertation because aspiration is a key quality of virtue, and virtue necessarily involves the will to improve, to get better at practical reasoning (See generally Annas 2011). This drive is inconsistent with the existence of a state of perfection along the lines of (c). In short, being virtuous involves the desire for constant improvement, and rejects the notion of a static, perfect disposition embodied in (c).

The prominent role of aspiration in this description of virtue may strike the reader as problematic. If aspiration is an intrinsic part of virtue, and the aspirer never achieves what they aspire to (or they would stop aspiring), then is not the pursuit of virtue futile? Annas says something to this point in regards to skills, as part of her general argumentative strategy to analogize virtue to practical skills:

“Does the need to improve eventually recede, or is it always present? If it is, this seems to imply that nobody can master a skill completely. The need to improve never in fact entirely disappears, but the implication is simply that

mastery of a skill is incompatible with its being mere routine; experts in a skill need to maintain it as a skill and not mere routine.” (Annas 2011, 18)

In a similar manner, the virtuous person strives toward right action in the (a) and (b) sense, but not toward a state of character like (c), because such a state would render virtuous action routine. Doing the right thing in uncertain and unknowable future circumstances involves vigilance, effort, and striving, not resting on the laurels of having achieved some particular disposition. In Annas’ words again:

“[the ideal and aspirational side of the virtues] is not to be confused with any form of perfectionism, in which we know beforehand what the ideal achievements are that we are to attain. The ideal aspect of virtue leads us to aspire continually, not to get the prize and then retire.” (Annas 2012, 116 n.23)

This view does not violate the axiom of futility, which holds that we should not strive for things we cannot achieve, because it does not claim that we cannot achieve a lifetime of right actions, only that we cannot achieve a state of character that guarantees a lifetime of right action. While Annas uses the term “aspiration” to capture this feature, perhaps “striving”, or “active effort”, could serve the same purpose while not raising the specter of a violation of the axiom of futility. When the term “aspiration” is used in this dissertation to refer to a quality of the virtuous person, it is always meant in this sense.

While this view is motivated by Annas’ work, the part important for this dissertation overlaps, for example, with Robert Adams’ religious viewpoint. Adams believes that virtue is something we strive for in the image of God and that human virtues are necessarily fragmented versions of the complete and perfect virtue of divinity. For Adams, God would have perfect virtue in the sense of (c) above (Adams

2006, 173). This is virtue as an unrealizable ideal, for while we can aspire to do as God would want us to do, we cannot aspire to be God. Adams' human virtue is therefore necessarily aspirational in a manner similar to Annas' virtue.³ In both cases, although for different reasons, the state of human virtue is aspirational. Annas makes an interesting comment about this feature of virtue:

“Imagine a pianist whose goal is to play like Alfred Brendel, but mistakenly thinks that she will achieve this by copying all his mannerisms and niceties of style, playing only the pieces he plays and playing them just as he plays them. The result would be an *impersonation* of Brendel, not the achievement of his skill” (italics added, Annas 2011, 17).

This observation about piano playing captures a key aspect of skill and of virtue. Just as with the Brendel impersonator, someone who tries to be virtuous by copying the actions of the virtuous agent misses the point, and misses virtue. In piano playing, the artistic expressive aspect of performing makes the impersonator's attempt at playing like Brendel obviously misguided. Even if the person could literally perform like Brendel, that performance is derivative from the original creative work of the artist Brendel himself. It is not clear, however, that this feature is analogous with virtue. We do not need a million piano players copying Brendel, but is it really a bad thing to have a million people copying Mother Theresa?

³ Adams' notion of virtue involves aspiration to a state that is more strictly impossible than Annas'. It seems to violate the axiom of futility. I am inclined to be charitable towards Adams' view and say that, while he may ask us to aspire to an impossible ideal, he may believe that this aspiration is itself the best means we have to improvement, and that aspiring to something lesser, something human, would lead us to achieve less. One might notice the break here between epistemic and pragmatic reasons to act. It could be the case that while I have no good evidence that I can be like God, I might have good reason to try to be like God if that attempt is what leads me to be the best person I can be. And someone who has a Christian set of ontological beliefs could, I think reasonably, hold that humans made in the image of God are made to aspire to goodness, with God as the paradigm case.

There are a few reasons why the impersonation of a virtuous person does not itself constitute virtue. First, Annas distinguishes the circumstances of a life from the living of a life (Annas 2011, 92-95). Virtue is about how we live our lives, but we always live our lives in various circumstances. And because no two people have the same circumstances, we cannot trust imitation as a reliable indicator of how to act in our own circumstances. We cannot simply copy the virtuous person because we have not actually seen the virtuous person act in our own circumstances. We would always necessarily be extrapolating from what we have seen the virtuous person do to some notion of how they would act in our circumstances. In comparison with music, it is as if we were all presented with different instruments that demanded different sorts of playing styles for greatness. More fundamentally however, this sort of impersonation is not virtue because impersonating the virtuous person is an improper motive for moral action. The virtuous person does what they do because they perceive it to be the good and right thing to do in these circumstances. They do it because they are trying to do what is right, not because they are trying to have some particular virtuous character trait. Impersonating the virtuous person is therefore doomed to fail. To the extent that you act as you act because you are impersonating, you are doing it for the wrong reasons. To the extent that you begin to take on the right reasons for action, you stop impersonating the virtuous agent and you begin to try to exercise virtue in a “self-directed” way (Annas 2011, 18; Cf. NE 1106b36-1107a2).

1.2 Virtue Ethics and Naturalism

Much of this dissertation argues the potentially contentious claim that empirical psychology is relevant to ethics, particularly to virtue ethics. Any purported link between normative and empirical claims is contentious, and there are many aspects of this claim that I will not be able to consider here. I will briefly grapple with some of the most serious concerns in this section of the introduction, and leave a fuller discussion for another author, or another time.

Some believe that attempts by virtue ethicists to engage with the psychological or natural sciences is misguided (see Hursthouse 2012, 16 for a general discussion). Robert Adams, for example, argues that with regard to questions about the human good, “I think we cannot expect naturalistic (for instance biological) investigations of ‘human nature’ to answer these questions convincingly” (Adams 2006, 51).⁴ Ethics is normative, and normative statements both cannot and should not be derived from natural statements. In particular, Adams thinks that ethical value must come from a transcendent good, and we cannot learn about transcendent goods from natural facts. “Biological and other natural facts” about our lives establish the context within which we can seek the good, but they cannot tell us about the good itself (Adams 2006, 52).

Adams’ position is not the only one critiquing this aspect of naturalism in virtue ethics. One might argue, for instance, that virtue ethics commits itself to a form

⁴ Adams, and others, cite William Fitzpatrick’s book *Teleology and the Norms of Nature* as an authoritative example of this sort of critique. I have unfortunately not had the chance to review that work.

of outdated naturalistic teleology by claiming that the fact that humans naturally are some way implies that this way is good, or that humans ought to be this way.

Rosalind Hursthouse presents a basic version of the modern virtue ethics response:

“*Eudaimonia* in virtue ethics, is indeed a moralized concept, but it is not only that. Claims about what constitutes flourishing for human beings no more float free of scientific facts about what human beings are like than ethological claims about what constitutes flourishing for elephants. In both cases, the truth of the claims depends *in part* on what kind of animal they are and what capacities, desires and interests the humans or elephants have” (Hursthouse 2012, 16; Cf. Foot 2001).

The idea is that while we cannot read ethical norms straight off of human biology or psychology, we can also not talk about human flourishing, or *eudaimonia*, without talking about empirical facts about human nature, including human “capacities, desires and interests”.

This view is not in error because it says nothing about whether particular human capacities, desires, and interests are themselves good in reference to some standpoint external to human life. Virtue ethics claims that *eudaimonia* is the state of a flourishing, excellent human. It does not claim that the human is excellent because she possesses these capacities, desires, and interests. It rather claims that it is good to develop (or restrict) the capacities humans happen to have because we want flourishing, excellent lives. We could make a parallel statement about desires and interests, and in each the “because” does not function as it does in a naturalistic teleology.

In my view, this position ought to be less controversial than it appears to be. Basic facts about human biology and psychology should, and I think do, matter on

any theory of ethics. For Utilitarianism, this is perhaps more obvious, for facts about what brings us pleasure and pain, what is healthy and bad for us, and what sorts of activities we most enjoy, these are all based on contingent facts about humanity discoverable by natural science. In particular, much of the work in social psychology discussed in this dissertation should matter very much to Utilitarians, because it provides varied evidence about what sorts of interests humans have, and what sorts of activities promote human welfare.

For Kantian deontology, the view is perhaps a bit more controversial. But it seems clear that facts about human nature influence our moral obligations to other people. To borrow an example from H.L.A. Hart, if humans were born “like giant land crabs with an impenetrable carapace” (Hart 1958, 623), the question of what sorts of maxim we could will universally, and of what constituted using another as a mere means, would be drastically different. We would be reasonable in assuming, for example, that other crab-people would not care about physical violence, since they are invulnerable to harm, and many violent actions that are impermissible for us would be permissible for crab-people. There is a sense in which the formal structure of deontology could remain intact, but much about how we are obligated to relate to each other physically would be different.

We ought to embrace the basic relevance of the natural sciences, to the extent that they help us understand human capacities and interests, to virtue ethics.

1.3 Virtue Ethics and Empirical Psychology

Since natural facts about humans are important for ethics, it is critical to consider how we should use empirical data in moral theorizing. In particular, this dissertation uses extensive empirical results from social and personality psychology. While there are a few discussions of clinical psychology here, the malleability of therapeutic practices makes clinical psychology an unstable ground for philosophical argument. Empirical psychological practices at least agree on a fundamental scientific method of investigation. Social psychology is widely considered relevant to ethics, because most pressing ethical questions involve our conduct with and toward other people. Most troubling cases of immoral action, for example, involve social, or rather anti-social, behavior. But virtue ethics should also consider personality psychology, and so much of the research and argument in chapters 3 and 4 concern this sort of evidence. Virtue ethics, to a greater degree than the other major schools of ethics, considers actions that only affect ourselves to be morally significant. This provides a special incentive for virtue ethics, as opposed to ethicists in general, to take broader results in psychology seriously.

Given the extensive discussion of psychology in this dissertation, it is important to be clear about the appropriate use of psychological studies in philosophical argument. This is particularly true given the substantial, reasonable concern about the truth of many psychological findings (See e.g. Meehl 1978; Ioannidis 2005; Nuzzo 2014). As one psychologist has said,

“There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims...

However, this should not be surprising. It can be proven that most claimed research findings are false.” (Ioannidis 2005, 696)

This literature is based on the existence of particular statistical problems in many areas of modern experimental science, where the focus is placed on obtaining a p-value to show “statistical significance”. It is the case that single studies, even those with statistically significant p-values, can tell us very little about the likelihood of the experimental hypothesis in question. Accordingly, as a first step, in this dissertation I have not relied in any serious manner on single studies. The one notable exception is the discussion of famous studies which would be unethical to replicate, such as Milgram’s obedience studies and the Stanford Prison experiment. These studies have had such a large influence on the literature in this area that they demand attention, and so I follow John Doris, among others, in discussing their import. But it is important to note that my dialectic with Doris assumes some psychological facts about humanity that could turn out to have no empirical support. Given the high visibility of the work of Doris and other situationists, I think it is worth addressing their arguments on their own terms, but there is also substantial reason to think that it might turn out that some of the more fantastic empirical results are not replicable. Aside from these sorts of studies, this dissertation focuses on psychological results that have been widely and variously confirmed, so as to try to minimize the problems of statistical significance testing. This is why I focus, for example, on books written by well-known psychologists that are summarizing hundreds of studies they undertook to examine a relevant effect (see, e.g., Kahneman 2011; Langer 1989; Langer 2009; Tavris 2008). I

may relate particular studies in order to explain the point being made, but they are all a part of large bodies of experimental work.

A second step taken in this dissertation to mitigate these worries is to focus on the psychological data, not the psychological theory. When the popular media covers psychological studies, the story often concerns what psychologists think their studies mean rather than the data that the studies generate. Philosophers ought not to follow suit and defer to what psychologists think their studies mean. The evidence psychologists have collected about human behavior is important, their theories explaining that data less so. It is not that psychologists are necessarily any worse at explaining the data than other people; surely they are not. It is rather that a philosophical analysis ought not to appeal to psychologists as authorities on explaining the data. The expertise of psychologists that this thesis defers to is their experimental expertise, not their theoretical expertise. Wherever I have considered any particular psychological study in some detail, therefore, I have tried to relate the experimental procedure and results directly rather than relying on a summary, and I have tried to make conclusions independent of the psychologists' own theorizing on the importance of their data.

Despite these steps, it is still quite likely that some of the research referred to in this dissertation will turn out to be false. For example, there has been a recent debate about the replicability of some of John Bargh's studies on priming, studies which Kahneman discusses at length (Kahneman 2011). Recently, a number of research groups have tried to replicate some of Bargh's findings and failed (Harris et.

al. 2013).⁵ This dissertation takes two approaches to the danger that this study, or indeed any of the studies presented in this dissertation, turn out to be false. First, in regard to chapter 2 and my responses to the situationist social psychology, this danger only helps my claims. If there is reason to be more cautious about the famous situationist social psychology studies, then virtue ethics is on even stronger footing. In regard to Chapters 3 and 4, and the parts of Chapter 2 on positive psychology, where I take psychological evidence to be in line with virtue ethics, I have tried to speak in the language of inspiration. For example, the psychological evidence strengthens the claim that humility is a virtue, given the strong philosophical reasons already present to think that humility is a virtue. Likewise with mindfulness. Psychological evidence can be inspiration for considering certain traits to be virtues; it is at most *prima facie* rather than sufficient evidence in favor of considering certain traits to be virtues.

One interesting implication of these problems with empirical psychology is that we should not take empirical evidence in favor of unlikely virtues very seriously. This is because when the prior probability of some trait being a virtue is low, even a very statistically significant study is unlikely to be measuring a true effect. So if, for example, some psychologist generated a study in favor of the hypothesis that cruel behavior had some hallmarks of a virtue, it is unlikely that this study is a good reason to consider cruel behavior to be virtuous. Since we think it is so unlikely that cruelty

⁵ It is important to note that Kahneman has been at the forefront of pushing his colleagues, particularly priming effects researchers, to confront these issues and develop strategies and practices to mitigate these problems. See generally his open email to colleagues, available at [http://www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman Letter.pdf](http://www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf)

is a virtue, even a very statistically significant study should raise our belief in that hypothesis very little (Nuzzo 2014). Bayes' famous theorem is a quick way to see why this is the case. The traditional formulation of Bayes' theorem is:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}.$$

In this notation, A and B are events related by Bayes' Theorem. It says that the probability of A given that B, or on the condition that B, is equivalent to the probability of B given that A times the probability of A divided by the probability of B. In a scientific context, A is the hypothesis in question and B is the experimental evidence, so Bayes Theorem in this context expresses the probability of the hypothesis given the experimental evidence. The term "P (B/A)" expresses a number that the p-value tells us about (though the p-value itself is really an expression of P(B/~A)). Bayes' Theorem shows that the link between p-values and the likelihood of the relevant hypothesis must take into account P(A), the prior probability that the hypothesis is true. If P(A) is very low, then even a very low p-value, which would indicate a very high "P(B/A)", does not make P(A/B) particularly large. As an example, if we are testing a long-shot hypothesis that we think is only 5% likely, and we generate a statistically significant study in favor of the hypothesis (p-value = 0.05), the post-study likelihood of the hypothesis is only 11% (Nuzzo 2014). It is still much more likely than not to be false. Even a very statistically significant study (p-value = 0.01) only raises the likelihood of the hypothesis to 30% (Nuzzo 2014). On the other hand, if we independently generated many studies showing this result, or we robustly repeated the statistically significant study and obtained further statistically

significant results, then the likelihood of the hypothesis may increase to the point where we have good reason to believe in it.

One could object that it is too self-serving to say that psychological evidence is only important if it confirms our traditional beliefs about virtue. It might indeed be the case that I am sensitive to this argument because I am motivated to believe its conclusion. But that does not necessarily make the argument bad. Widespread practices of statistical significance testing are flawed because they fail to account for the prior probability of a hypothesis. This is simply a true statement about p-values, as their creator Ronald Fisher, understood (see generally Meehl 1978).

More importantly, however, this problem suggests that psychologists probably need the help of philosophers about as much as philosophers need the help of psychologists. That philosophy should look to psychology for inspiration is a main focus of this dissertation. This brief mathematical detour is good support for the converse claim, because psychologists should be concerned about the plausibility of different traits as virtues before they study these traits. There is no better resource for thinking about the prior probability of virtue traits than the philosophical tradition in virtue ethics.

2. Empirical Investigations of Virtue: Situationism and Positive Psychology

For Aristotle, the existence of virtue is a fundamentally empirical claim about human nature. In his famous “function argument”, for example, he grounds the concept of virtue in an argument about human nature and the reasoning function of humanity: the highest human good, *eudaimonia*, is “the activity of the soul in accord with virtue” (NE 1098a5-10). Therefore, empirical psychology, to the extent that it studies these capacities of the human soul, is an important set of empirical data for Aristotelian virtue theory. Aristotle’s understanding of human nature deeply informed his ethics, and I think ethicists today following in Aristotle’s footsteps should be equally interested in human nature. We should therefore care very much about the arguments of John Doris, and his book A Lack of Character, who claims that the psychological structures necessary to support a robust virtue ethics are not present in humans. Doris grounds his argument on the situationist program of research in empirical psychology, which purports to show that human action is more influenced by situational factors than general, non-situation specific character traits. While Doris casts his arguments as a challenge to virtue theory, I think that an Aristotelian ought to see his arguments as continuous with the proper concern of virtue theory with human nature. It is the main project of this chapter to show that Doris’ arguments fail, but it is important to understand at the outset that Doris, and others bringing empirical psychology into virtue ethics, are working within the tradition established by Aristotle, not against it. It is always awkward to invoke notions of what any philosopher ‘would say’ in particular circumstances, but I think that if we are to take

Aristotle's empirical commitments seriously, he would be first in line to renounce his normative virtue theory if it turned out to be empirically inadequate, as Doris claims it is. This chapter is therefore a critique of Doris' reasoning, not a critique of his move to bring empirical psychology to bear on virtue ethics' normative claims.

I offer two rebuttals in this chapter to Doris, and other philosophers who argue similar claims on the basis of situationist research. First, I argue that they fail to understand virtue theory's theoretical commitments. They attack virtue theory as empirically inadequate without properly grasping what empirical claims virtue theory needs to make. In particular, it seems to me that they wrongly infer conclusions about moral competencies from evidence about moral performances. In addition to my new argument about the competence-performance distinction, I show that situationists have failed to adequately consider some important past criticism leveled against situationism about the inconsistency of vice. In general, it is important that those philosophers using empirical psychology to criticize virtue ethics actually have a realistic virtue ethics as their target. Clearing these conceptual difficulties is a prerequisite to discussing the impact of empirical psychology on virtue ethics.

Situationists' failure to appreciate what sorts of descriptive empirical claims virtue theory does make also leads them to overlook some psychological literature that is relevant to the plausibility of virtue ethics. In particular, in my second criticism I argue that situationists fail to take account of psychological evidence that confirms the moral psychology of virtue theory. Situationism is a powerful area of psychological research, and I believe it does have important implications for our

moral life, but there are other areas of psychology that are also important for our moral life and provide evidence in favor of a virtue moral psychology.⁶ I will examine the recent movement of positive psychology in particular, and look at the research surrounding the virtue of gratitude as a paradigm. I claim that this research provides empirical evidence that aspiring to virtue is causally linked with increased feelings of subjective well-being. This empirical link is exactly what virtue theory predicts, and while falling short of confirming a virtue moral psychology, at the very least it helps render plausible the claim that attempting to live a virtuous life leads to happiness.

The focus of this chapter is on critiquing the reasoning of Doris and the situationists, and I hope I can lay bare some of their mistakes. My hope is that by understanding some of their missteps, we can more fruitfully bring empirical psychology to bear on ethics. Indeed, in the rest of my thesis, I will try to show how a virtue theorist could use empirical research to ground arguments in favor of particular virtues. I will argue in chapter 3 that the situationist research itself grounds an argument for the virtue of mindfulness. I will argue in chapter 4 that another area of psychological research that is receiving a lot of recent attention, research on motivated reasoning and cognitive dissonance, grounds a secular argument for the historically religious virtue of humility. Through all of these arguments, I hope to show that how we take modern empirical psychology for virtue ethics is more a

⁶ I will use the term “virtue moral psychology” throughout to refer to the basic moral psychology commitments of an Aristotelian virtue theory. I believe that the ones I have chosen to focus on in this chapter are obvious and basic enough not to generate serious objections to my arguments.

matter of the author’s pessimism or optimism about virtue prior to encountering the research than something that is implicit in the research itself. This thesis is an attempt to engage with the very same empirical data with which Doris engages, but as a virtue theorist trying to find the best virtue theory.

2.1 Lack of Character: Doris’ Situationist Challenge

This is how Doris summarizes the challenge he brings:

“The experimental record suggests that situational factors are often better predictors of behavior than personal factors, and this impression is reinforced by careful examination of behavior outside the confines of the laboratory. In very many situations it looks as though personality is less than robustly determinative of character. To put things crudely, people typically lack character” (Doris 2002, 2).

Doris believes that the situationist psychology research program has shown that our everyday notions of character are mistaken, and that what character traits we do have are insufficiently robust to support a reasonably powerful Aristotelian virtue theory.

Doris takes aim at a particular notion of character, which he labels globalism (Doris 2002, 22). Globalism maintains three theses, which Doris contrasts with analogous situationist theses, that he says “amount[] to a qualified rejection of globalism” (Doris 2002, 24). These theses are:

GLOBALISM (Doris 2002, 22)	SITUATIONISM (Doris 2002, 24-25)
“ Consistency. Character and personality traits are reliably manifested in trait-relevant behavior across a diversity of trait-relevant eliciting conditions that may vary widely in their conduciveness to the manifestation of the trait in question.”	“Behavioral variations across a population owes more to situational differences than dispositional differences among persons. Individual dispositional differences are not so behaviorally individuating as might have been supposed; to a surprising extent it is safest to predict, for a

	particular situation, that a person will behave in a fashion similar to the population norm.”
“ Stability. Character and personality traits are reliably manifested in trait-relevant behaviors over iterated trials of similar trait-relevant eliciting conditions.”	“Systematic observation problematizes the attribution of robust traits. People will quite typically behave inconsistently with respect to the attributive standards associated with a trait, and whatever behavioral consistency is displayed may be readily disrupted by situational variation. This is not to deny the existence of stability; the situationist acknowledges that individuals may exhibit behavioral regularity over iterated trials of substantially similar situations.”
“ Evaluative integration. In a given character or personality the occurrence of a trait with a particular evaluative valence is probabilistically related to the occurrence of other traits with similar evaluative valences.”	“Personality is not often evaluatively integrated. For a given person, the dispositions operative in one situation may have an evaluative status very different from those manifested in another situation; evaluatively inconsistent dispositions may “cohabitate” in a single personality.”

Situationist psychology is rich with compelling and memorable experiments.⁷ I will relate three favorites to illustrate these three disputed areas in turn, though it is important to note that each experimental result has been recreated many times.

A. Consistency

Stanley Milgram’s obedience experiment is the most famous example of a situationist experiment, so I will begin with the Milgram experiment as an example of inconsistency (Doris 2002, 39-51; Milgram 1963). Milgram’s experiment involves a single test subject, an experimenter, and another apparent subject who is actually a

⁷ Indeed, I often wonder if part of the explanation for what I think is the outsized impact of situationism in ethics is the vividness of some of its experimental results. The Milgram (Milgram 1963) and Stanford Prison (Zimbardo 1973) experiments spring to mind.

confederate (Doris 2002, 39-40). There is a rigged drawing wherein the test subject is assigned the role of ‘teacher’ and the confederate the ‘learner’. The test subject is told, as the teacher, that they are to administer a word association test to the learner, and for every wrong answer they must administer an increasing shock to the learner. The teacher and learner are in a separate room, no shocks are actually administered, and the learner’s response at each point is really a pre-recorded audio track. If the test subject refuses to continue administering shocks at any point, the experimenter will exhort the subject to continue, saying things such as “The experiment requires that you continue” and “You have no other choice, you *must* go on” (Doris 2002, 40). The voltage level that the test subject believes they are administering goes from 15 to 450, with appropriately dire warnings associated with the higher voltages – the labels end with “Danger: Severe shock” and then finally simply “XXX”. At approximately 150 volts, the pre-recorded voice rescinds its consent to participate in the experiment. The voice gets increasingly agitated, and increasingly angry, as the shock level rises. By 270 volts, the recorded voice lets out an “agonized scream” with every shock and begs to be let out, sometimes complaining of heart trouble. At 330 volts the voice lets out an extended scream, complains of heart pain, and begs to be let out. For the remaining eight shocks, there is no sound from the apparent victim. An appalling sixty-five percent of Milgram’s test subjects completed the entire experiment, administering those eight silent shocks (Doris 2002, 42).

Regardless of how this experiment strikes the reader, in this chapter I accept that the Milgram experiment is powerful and disturbing evidence of the power of

situations to overcome any disposition to compassion or justice most of us have. I agree with Doris' use of this experiment to question the general prevalence of the virtue of compassion in our society. He claims, "my argument requires only that the effects of situational stimuli often seem quite disproportionate to their intuitive magnitude, and such disproportion clearly obtained between the experimenter's instructions and the shocking behavior that they produced" (Doris 2002, 49). I also agree with Doris' substantive claim here, that the Milgram experiment shows some situational stimuli often, or even systematically, have effects disproportionate to their magnitude, but I disagree with Doris' claim about what his argument requires. In particular, as I will argue in more detail in section 2.3.1, while the Milgram experiment provides disturbing evidence about failed moral performance, it is not particularly dire with regards to moral competence. Some of Milgram's test subjects refused to administer any shocks beyond the point when the subject apparently withdraws consent. That seems to me clear evidence that humans are generally competent to refuse this situational pressure, unless Doris wants to claim that all the test subjects who failed to resist the situational pressure were somehow morally impaired and thereby unable to resist the situational pressure. But such an explanation would be its own form of appeal to personological characteristics, in this case an appeal to a vicious characteristic rather than a virtuous one. Milgram shows that virtue may be harder than we thought, and particularly harder in surprising ways, but not that it does not exist.

B. Stability

Situationism is historically traced, according to Doris, to experiments done on schoolchildren in the 1920s (Doris 2002, 24, 62-63). In these early experiments, researchers tested the schoolchildren's behavior for traits of honesty and deception. Hartshorne and May looked at behaviors in many different situations that might be relevant to honesty, including "22 opportunities to cheat in classroom work, 4 opportunities in athletic contests, 2 in party games, and 1 in school work done at home" (Hartshorne 1928, 407-408). They also looked at 46 total questions given to students with the opportunity to lie and three opportunities for students to steal (Hartshorne 1928, 408). The experiment showed that honest behavior in different situations showed minimal cross-situational consistency: "The results of these studies show that neither deceit nor its opposite, 'honesty,' are unified character traits, but rather specific functions of life situations. Most children will deceive in certain situations and not in others" (Hartshorne 1928, 411). Their studies indicated that "honesty is not an 'inner entity' but is instead 'a function of the situation'" (Doris 2002, 24).

This is a classic case of the situationist argument against stability: in different opportunities to manifest a particular virtue trait, almost everyone is inconsistent, doing the virtuous thing in at best only some of the situations. And as with the Milgram study, I think that this experiment is critically important to a nuanced virtue theory, particularly given the importance of moral education to the cultivation of virtue. But that is precisely what the experiment shows: a challenge for moral

education. Nothing in the experiments suggest that it is impossible to be honest across situations. And again, the experiment itself shows that such honesty is possible for humans, for it found honest humans in each and every situation. I also think that the objection to the use of this experiment given by Sabini and Silver, which I relate in 2.3 below, is fatal and continues to be under appreciated by situationists.

C. Evaluative Integration

One of the most compellingly simple situationist studies involved some students at Princeton Theological Seminary (Doris 2002, 33; Darley 1973). They were given the task of filling out a questionnaire in one building and then going to another building to give a short presentation, on either The Good Samaritan parable or a normatively less loaded passage. Students were given different instructions about timing, that either they were running late to the presentation, right on time, or a little early. On the way to the presentation, the students passed an experimental confederate slumped in a doorway, apparently in need of help. The experiment examined which students stopped to help the apparently injured confederate, and found that it varied widely with the timing instructions, where only ten percent of the students told they were running late helped, forty-five percent who were on time helped, and sixty-three percent who were running early helped. None of the other conditions, including the personality tests of the subjects or the assigned presentation content, showed any significant correlation with helping behavior. That is, in each situation, the timing

condition and only the timing condition heavily influenced the likelihood of whether a student would help the incapacitated confederate.⁸

Doris says that the situationist literature supports the following claim about evaluative integration: “the dispositions operative in one situation may have an evaluative status very different from those manifested in another situation; evaluatively inconsistent dispositions may ‘cohabit’ in a single personality” (Doris 2002, 25). The Good Samaritan experiment is good evidence of this, for the students are presumably headed to talk about the goodness of acting to help the injured party in the Good Samaritan parable. The students presumably had an intellectual belief in the values communicated by the parable, to the extent that they were going to lecture on it, that helping in a situation like the experimental one is laudable. But very few of them, under the instruction that they were running late, acted on such a belief. Hurriedness overwhelmed any personological, and also any other situational, variables.⁹

⁸ While I accept Doris’ basic characterization of this experiment for purposes of this chapter, I feel obliged to point out that Doris’ articulation is in fact overly skewed by his own view. In particular, Doris piles on criticism of the test subjects saying “Nor was there special reason to think, in the green fields of 1970s Princeton, New Jersey, that the victim posed some threat, as might be supposed in more threatening urban climes. Similarly, the placid suburban environment should have worked to reduce situational ambiguity” (Doris 2002, 34). But this ignores the experimental reports, where Darley and Batson said “The parable of the Good Samaritan also suggested how we would measure people’s helping behavior—their response to a stranger slumped by the side of one’s path. The victim should appear somewhat ambiguous—ill-dressed, possibly in need of help, but also possibly drunk or even potentially dangerous” (Darley 1973, 102). I do not know how well the experimenters constructed their experiment, but this description makes it hardly clear what response we might expect from the virtuous person, and gives the whole experiment less emotional impact as an example of human callousness. This issue with the Darley and Batson experiment has been noted by other critics of Doris (Sabini and Silver 2005, 558). Doris acknowledges this complexity in the Darley and Batson experiment in later writings (Merritt 2012, 369).

⁹ It is worth noting that there were facets of the experiment that seemed specifically intended to distract the test subjects as they walked. In the Good Samaritan condition, subjects were told in a written prompt “Because we are interested in how you think on your feet, you will not be allowed to use notes

I have focused on John Doris' Lack of Character here because it is the best extended defense of the situationist position I know of, but it is worth noting that Doris has recently published work affirming the key arguments of that book (Merritt 2012).¹⁰ In this text, he reaffirms the basic claim of situationism that "Behavior is *not* typically ordered by robust traits" (Merritt 2012, 358), where he defines robust traits as a trait, the holder of which "can confidently be expected to display trait-relevant behavior across a wide variety of trait- relevant situations, even where some or all of these situations are not optimally conducive to such behavior" (Merritt 2012, 356). This is essentially the globalism thesis related in Lack of Character. Therefore, while he does soften some of the positions taken in Lack of Character in this latest article, his central premise is unchanged.

D. Normative Importance

Doris does not just think that the situationist literature is important for virtue ethics to consider, and that virtue ethics can accommodate these empirical results (which is the position I will generally defend in this thesis). He believes that this research is fundamentally fatal to virtue ethics, and fatal to any reliance on notions of ethical character. He believes that we can, and ought, to revise our moral theories to exclude notions of ethical character from consideration. In his words, "I'm claiming

in giving the talk" (Darley 1973, 103). In such a situation, I know I, at least, would be mentally practicing my talk as I walked to the room to give it and be even less attentive to details of my environment than otherwise. I do not think that a limited capacity to focus on things in one's perceptual range is evidence of viciousness.

¹⁰ While Doris wrote this paper with two other authors, Maria Merritt and Gilbert Harman, given that it is the most recent restatement of situationist arguments under his authorship, I will take this paper as representing Doris' current view throughout my work in order to support my discussion of his *Lack of Character*.

that ethical reflection can safely dispense with notions of character” (Doris 2002, 108). He thinks we would be better off if we focused on actions and not character attribution (Doris 2002, 114-117), if we appreciated more the role that luck plays in ethical action (Doris 2002, 117-119), and if we dropped character traits from therapeutic practices aimed at making us better, happier people (Doris 2002, 119-120). In this thesis, I agree with Doris’ basic approach to the situationist research: “that reflection on situationist moral psychology can help us to judge and act better in ethically trying circumstances, despite the presence of continued theoretical dispute” (Doris 109). There are also parts of Doris’ suggested reformation of ethical reflection on which I concur.¹¹ But I will argue directly against the claim that ethical reflection would be better off absent concepts of character. In this section, I will argue that the empirical evidence does not demand that we drop notions of ethical character from our discourse, that Doris’ situationist argument, so to speak, is wrong. In later chapters, I will develop several arguments that in fact, the language of character is a good way to understand and discuss many of the most interesting and important results in modern empirical psychology.

2.2 Replies to Doris in the Current Literature

There have been quite a few criticisms of Doris, and the other situationists, in the literature. In this section, I will summarize and comment on some representative critiques.

¹¹ See Section 2.4.2, below.

2.2.1 Sabini & Silver - The Problem of Vicious Moral Psychology

Sabini and Silver (2005) critique the situationists on their use of the psychological evidence. They divide the data into four categories: “data on trans-situational correlations, data on social influences, data on mood effects on behavior, and, indirectly, data on the fundamental attribution error” (Sabini & Silver, 539). They mainly discuss the first two categories. They do not discuss the fundamental attribution error, rightly saying that data on that question is only relevant if the globalism thesis is actually false. Attributing character traits is only fundamentally an error if in fact humans have no traits.¹² It might be an error to attribute character traits in an unwarranted way regardless, but Harman’s argument relies on the claim that there are no real character traits to attribute. If the globalism thesis is in fact true, Sabini and Silver are arguing that the attribution might not be in error at all. With regards to the mood effects data, they reject the notion that the mood effects measured, such as the likelihood a subject will pick up dropped papers based on finding or not finding a dime in a payphone coin return slot, are relevant to moral virtue. In the experiment they are referencing, Isen and Levin observed whether people using a telephone booth would help an experimental confederate who dropped papers on the ground nearby (Isen 1972) or mail a letter left in plain sight near the

¹² The fundamental attribution error refers to an early situationist argument in philosophy first pushed, to my knowledge, by Gilbert Harman (Harman 1999). Harman uses it to describe the commonplace error in which “in trying to characterize and explain a distinctive action, ordinary thinking tends to hypothesize a corresponding distinctive characteristic of the agent and tends to overlook the relevant details of the agent’s perceived situation. Because of this tendency, folk social psychology and more specifically folk morality are subject to what Ross calls ‘the fundamental attribution error’” (Harman 1999, 316). Harman claims “that there is no empirical basis for the existence of character traits” (Harman 1999, 316).

telephone booth (Levin 1975). In both experiments, they found that whether the subject found a dime, and were therefore in a good mood, had a substantial effect on helping behavior. As Sabini and Silver say, however, “we just do not believe that picking up or not picking up your papers is a very important manifestation of a moral trait” (Sabini & Silver, 539-540).¹³

With regard to the data on trans-situational correlation, Sabini and Silver first argue that the low correlations found among, for instance, honesty in different situations in the Hartshorne and May study, is actually high enough to support a virtue theory (Sabini & Silver, 540-544). They point out, for instance, that a high correlation in cross-situational honesty “would require not only that some people are consistently and virtuously honest but also that other people are consistently, and perhaps perversely, dishonest” (Sabini & Silver, 543). This is because a high correlation requires that the general population measured be consistent across test cases. A high consistency means that the stratification of moral dispositional traits across the entire population tested is consistent, at the virtuous end, the vicious end, and everywhere in between. High consistency would therefore imply that there is a consistently virtuous subset of the population, but also that there is a consistently vicious subset. A vicious person being sometimes virtuous will lower the correlation just as much as a supposed virtuous person being sometimes vicious. Low trans-

¹³ I do not think Sabini and Silver are being fair here. It does seem true that this particular helping behavior is not a significant measure of virtue. But the point of the use of this experiment, I think, is to emphasize the mismatch between the act, finding a dime, and the effect on mood. It is not unreasonable to think that this mood effect could influence more morally significant actions. At the very least, I would think that Sabini and Silver would want to show that the mood effect is lessened for more morally significant actions before dismissing this argument out of hand.

situational correlations do not show a lack of consistency among the virtuous, only a lack of consistency among the population as a whole, and it is perfectly possible that a consistent virtuous subset of people could exist within a generally inconsistent population.

Sabini and Silver also point out that honesty in classroom situations depends on more than one's moral character. For instance, whether one cheats on an exam has to do with both how honest one is and how smart one is, for the dishonest smart person does not need to cheat (Sabini & Silver, 543). Their study therefore fails to present the same moral question to all its subjects, because smart students were not necessarily induced in any way to cheat, undermining claims that it is a useful statistical sample for purposes of measuring virtue.

For the situationist arguments on social influence, Sabini and Silver focus on the Milgram experiment. Their basic line is: the circumstances of the Milgram experiment upset the subjects' sense of the world in a manner that is more powerful, and less threatening to virtue, than the situationists think. They compare the Milgram experiments, for instance, to Asch's studies of people's unwillingness to depart from a unanimous group consensus on something as trivial as stating which of a series of lines is longer than the others (Sabini & Silver 554; Asch 1955). In a typical experiment, "about 75 percent of the subjects gave the wrong answer—the answer the confederates gave—on at least one trial" (Sabini & Silver 554). They use these, and other, studies to claim that situationism has not discovered some diffuse propensity for people to be overwhelmed by situational factors to the detriment of character, but

rather a more specific thesis that we are “weak, morally weak... when confronted with a resolute authority or a unanimous group of other seemingly normal people who seem to see the social, moral, and even physical world differently from the way we do” (Sabini & Silver 560). This is not a particular challenge to virtue, they argue, but rather to any human perception at all. They therefore claim that these results do not call into question theses about general character traits, but rather provide very specific moral advice to those of us trying to have good characters: “that people should stay away from slippery slopes, that they should be very wary when their moral perceptions seem to clash with others’, and that they should understand both that it is hard and that it is possible to confront other people who are doing wrong” (Sabini & Silver 562).¹⁴

Sabini and Silver use this evidence to claim that “[p]eople are, we suggest, oblivious in their day-to-day lives of the degree to which their cognitive world could be thrown into chaos by the unanimous rejection by others of their view of it” (Sabini & Silver 554-55). They describe how these effects exist in experiments with moral significance, as in Milgram, but also in Asch and other studies with no moral significance, even to the point of ones that test the subjects’ own selfish self-preservation instinct (Sabini & Silver 557). The effect persists across all categories, suggesting that there is something more fundamental than a moral failing going on in these types of situations.

¹⁴ I will return to this sort of argument below in Chapters 3 and 4. My argument in those chapters, that we ought to consider mindfulness and humility to be virtues, is inspired partly by these sorts of concerns about situations in which we characteristically make mistakes.

I think that Sabini and Silver's objections are generally excellent. In particular, I think that their argument about the inconsistency of vice is critical and, sadly, still unappreciated by Doris. I try to articulate why this objection is important, and still underappreciated, in Section 2.4.2 below.

2.2.2 Sreenivasan - Operationalizing Virtue

Gopal Sreenivasan has criticized the situationist literature on the basis of its interpretation of the psychological evidence (Sreenivasan 2002). He first distinguishes the fundamental attribution error from a claim about the actual existence of character traits (Sreenivasan 2002, 53-54). He is right to do so, for the justification of attributing character traits based on any set of observations is a distinct question from whether such traits exist (see Vranas 2005, 2009). As Sreenivasan says, "the predictive inefficacy of everyday trait attributions does not itself tell us anything either about the predictive efficacy of warranted trait attributions or about the prospects of acquiring the relevant warrants" (Sreenivasan 2002, 54). This criticism is similar to the one levied by Sabini and Silver, discussed above, but it is also not the focus of situationists such as Doris, who argue that in fact character trait attributions are never warranted.

Sreenivasan levels two more serious charges at situationists such as Doris. He claims that two assumptions are required for the situationist argument to work, neither of which are justified. First, the low cross-situational correlations among behaviors in different situations, such as in Hartshorne and May's studies, corresponds to the cross-situational correlations among the same person's behavior in

different situations. The consistency coefficients in Hartshorne and May are aggregates, over all children, and “this figure does not exclude there being some individuals for whom the correlation between stealing and lying situations was much higher” (Sreenivasan 2002, 56). That is, the study does not tell us much about whether there exist honest individuals, only that they are not the majority. This is a similar critique to Sabini and Silver’s point about what they call “trans-situational” consistency, discussed in more detail above.¹⁵

Sreenivasan’s second charge is more important and novel: situationism assumes that studies like the Hartshorne and May’s properly operationalize the virtuous character traits (Sreenivasan 2002, 55). Sreenivasan challenges this assumption on three points. First, the studies assume without argument that an objective, external viewpoint should be privileged over the agent’s subjective view of the act (Sreenivasan 2002, 58). Second, the studies do not distinguish the relative importance of the studied actions to the trait in question (Sreenivasan 2002, 58-59). For example, taking change from a table in an empty classroom may not be as important to honesty as cheating on a test (Sreenivasan 2002, 49). Finally, different cases of the same act are not equally normatively sensitive: lying is certainly a relevant case for testing the trait honesty, but what about lying for some other good, such as saving a fellow student from trouble (Sreenivasan 2002, 59-60)? Lying in

¹⁵ To my knowledge, Doris and other situationists do not directly address this concern. On the one hand, this is not entirely unreasonable: the situationist is in the enviable position of having a mountain of evidence for her empirical claims, and attacking particular studies in this manner is not likely to weaken the overall weight of evidence. On the other hand, a cavalier attitude to the actual statistical methods used in landmark situationism studies is troubling.

such a case does not necessarily show a lack of virtue or of honesty. Depending on one's substantive theory of the virtues, it could be consistent with virtue to lie in some cases where you would prevent a great harm from happening. Sreenivasan reinforces this point in relation to Darley and Batson's Good Samaritan experiment: just because there is reason to help someone in distress does not imply that this reason is infeasible. "[T]he fact that one is in a hurry can defeat the reason to help someone in distress" (Sreenivasan 2002, 60). This is not just a theoretical point, as a subsequent experiment showed that the reason one is in a hurry can substantially influence whether people stop to help. In particular, Batson showed that people told that the task they were going to complete was "not essential" were more likely to help than if they were told the task was "of vital importance" (Sreenivasan 2002, 60 n.12).

These factors together lead Sreenivasan to conclude that studies like Hartshorne and May's do not properly operationalize the character trait honesty, and he suggests three requirements that a study must satisfy to properly operationalize any character trait. It must measure a "central or *paradigm case*" of the virtue in question, it must not have any feature that would defeat the reason supporting the virtuous response, and the subject and observer must agree on these characterizations (Sreenivasan 2002, 61-62).¹⁶ Sreenivasan says that a theory of virtue presupposes that "there is some non-trivial number of people whose [virtue] is cross-situationally

¹⁶ Sreenivasan rightly points out that his test does not limit *the virtues* to traits that subject and observer agree upon. His requirements are requirements for empirically measuring character traits. That is, these are suggested as requirements for empirical adequacy of character trait measurements, not their moral worth (Sreenivasan 2002, 62). I think Milgram is a paradigm of what Sreenivasan has in mind here, where the experimental subjects themselves often agreed that the actions the subjects did were unethical (Milgram 1963, 376).

consistent across a range of behavioural measures that satisfy all three generic requirements” (Sreenivasan 2002, 63). On this definition, he claims that no theory of virtue has been falsified by the situationist data.

Doris recently characterized Sreenivasan’s criticism as an “attempt to show that ostensibly disturbing behavior is actually *compatible* with virtue, instead of merely rationalizable” (Merritt 2012, 369). While he agrees that some studies, such as the Darley and Batson Good Samaritan study, fit this argument, he also argues that “such stories will not be equally apt to all cases”. I do not think this is a fair criticism of Sreenivasan. As I outlined above, Sreenivasan’s main critique of situationism is about the empirical adequacy of the psychological tests to establish the non-existence of any character trait at all, regardless of whether we label it virtue. Sreenivasan is accusing the situationist studies of being beside the point, so to speak, in relation to virtue. He is not trying to argue that the behavior captured in the situationist studies is actually virtuous, he is arguing that the studies are not properly measuring the things we need to measure to answer the question of whether they are virtuous. To answer this charge, Doris ought to say how the situationist studies he cites actually properly operationalize any character trait at all. That being said, Sreenivasan’s argument does have more intuitive plausibility in relation to studies like the Hartshorne and May as opposed to ones like Milgram and Zimbardo. Milgram in particular seems to satisfy the three requirements, given that Milgram claims, for example, that “it is clear from the remarks and outward behavior of many participants that in punishing the victim they are often acting against their own values. Subjects often expressed deep

disapproval of shocking a man in the face of his objections, and others denounced it as stupid and senseless” (Milgram 1963, 376). Sreenivasan might have mentioned this study and focused on the fact that substantial numbers of subjects acted rightly in Milgram, and that the experiment therefore fails to falsify his virtue theory.

I think that Sreenivasan’s criticisms are largely good, but he does not go far enough. In particular, as I will outline below, his standard description of virtue theory grants too much to Doris and the situationists. There is no reason, I think, that virtue theory is committed to the proposition that “some non-trivial number of people” must have a cross-situationally consistent trait in order for that trait to be a virtue. As I will argue below, it is possible that a virtue trait exists that is not currently expressed by anyone. In essence, I think that Sreenivasan makes the same mistake about empirical adequacy that the situationists make.

2.2.3 Annas

In a commentary on Doris’ text, Julia Annas argues that he misconceives the ancient virtue tradition he is attacking (Annas 2005). Her criticisms are the closest to mine that I have found in the literature, because she points out some of the ways in which Doris misconstrues his target. However, I do not think all of her charges are fair.

Annas first accuses Doris of selectively quoting from Aristotle in order to paint a picture of virtue as “an uncritical and rigid habit” (Annas 2005, 637). Annas says rather that virtue is “a disposition to *act on reasons*...The firmness of the virtuous person’s character is a firmness in intelligent deliberation, not, as Doris

thinks, the firmness of fixed response” (Annas 2005, 638). This is an excellent point, and gets at some of the same issues Sreenivasan raised. In particular, the focus on outcomes in social psychology experiments does not capture the actual deliberations of the agent, which on Annas’ view is all that matters to an assignment of virtue. Of course, this is also why Doris and other situationists come to rely more and more, I think, on experiments like Milgram’s, where it seems clear that no good deliberation could lead one to deliver the final shocks.¹⁷

Annas uses this misunderstanding of virtue theory to show how the situationist position does not make sense as a description of how we interact morally. Doris accepts that local traits exist, that for example some people can be relied upon to be courageous in particular types of situations. Annas points out that if we encounter someone, Mary, only in those situations where Mary acts virtuously, then we might consider her virtuous (Annas 2005, 640). What happens when we encounter Mary in a different situation where she does not act as virtue demands? Annas thinks Doris would say this is not very significant, for we have just discovered that Mary is a more complex individual: her situational character traits are more complex than we presumed from our prior interactions with her, but she does not lack some sort of universal or general trait, because such things do not exist. And if Doris did say that, then I think that Annas is right to point out that this is not what we would say at all, that rather the new situation would lead us to question our judgment of Mary’s virtue

¹⁷ Perhaps we are overlooking good reasons to deliver the shocks. But it does not seem so to me, and since Doris and I are in agreement here, I will not pursue this suggestion.

in the first place. Perhaps Mary only acted well in the earlier situations because she wanted to impress us, or because someone bribed her to do so.¹⁸ The fact that she clearly lacks virtue in other situations leads us to think that she lacked virtue in the original situation too, and therefore leads us to seek some other, non-personological, explanation for her apparently virtuous conduct in the first instance.

While I am sympathetic to many of Annas' points, she ultimately misses her mark by committing the same basic sin she accuses Doris of: misconstruing her target. Doris is not committed to many of the claims "the situationist" is alleged to support in her arguments. Doris might believe, for instance, that it is a bad thing that we are not more morally integrated. His position is consistent with holding that morally integrated, general character traits would be a normatively good thing, and that even if they are empirically impossible, we still ought to aspire to embody them as much as possible. That is, he could accept something like the Stoic sage as a practically impossible normative ideal. Doris' situationist argument is concerned with the explanatory adequacy of character traits to explain behavior, not with the normative value of character traits. He is free to agree that virtuous character traits would be wonderful to have, he is just concerned that such things do not exist. Thus, it seems to me that Annas' comments are ultimately tangential to Doris' arguments.

¹⁸ I actually think, however (and I do not know if Annas agrees), that Mary could have acted for the right reasons in her first action. Our later skepticism should be focused on her character trait, not necessarily on the moral worth of the initial action.

Many of the objections to Doris in the literature summarized here are, I think, trenchant. But as I will argue below, I think there are important criticisms of the situationist position that have not yet been made.

2.3. First Reply to Doris: Concepts of Virtue

2.3.1 The Competence-Performance Distinction

“And so the virtues arise in us neither by nature nor against nature. Rather, we are by nature able to acquire them, and we are completed through habit.” (NE 1103a23-25)¹⁹

My argument in this section was inspired by an intriguing new book, where John Mikhail defends Rawls’ linguistic analogy as setting forth a legitimate framework for empirical investigation into human moral behavior (Mikhail 2011). Rawls suggested that we might fruitfully analogize our capacity to make moral judgments to our capacity to make grammatical judgments. He suggested that the fact that our judgments of grammaticality extend beyond our explicit principles of grammar might indicate that our judgments of morality also extend beyond our explicit principles of morality (Rawls 1971, 40-42). Mikhail follows this line of reasoning to argue that we ought to design empirical investigations into the possibility of a human cognitive structure of moral decision-making that corresponds to the cognitive structures of human linguistic decision-making. I would like to focus on one point in his argument here, where Mikhail deploys the competence-performance distinction concept from Chomsky in analyzing moral psychology. “Linguistic competence, in Chomsky’s framework, denotes a person’s *knowledge* of her

¹⁹ All *Nicomachean Ethics* English translations are from the Irwin 1999 translation.

language; linguistic performance refers to how, in actual situations, her knowledge of language gets put to *use*” (Mikhail 2011, 18).

Mikhail argues that a similar competence-performance distinction is useful in moral psychology. I agree, though I will use the concept somewhat differently from Mikhail. Mikhail says, “I will use the term *moral competence* to refer to an individual’s moral knowledge and the term *moral performance* to refer to how that knowledge is put to use” (Mikhail 2011, 18). I do not know why Mikhail limits moral psychology to knowledge and its use, though he seems to construe knowledge very broadly. Suffice to say, however, that my notion of virtue is Aristotelian in inspiration, and so my notion of moral competence includes moral knowledge, but also extends to everything included in the will (cf. McDowell 1998, 50-73). For my argument, I take moral competence to be synonymous with moral potential, not in relation to any specifically posited Chomskian structures in the brain, but in relation to an Aristotelian understanding of moral potential.

My argument in this section is simple in form. The philosophical situationists are making a mistake by arguing that moral performance, as measured in the situationist psychology research, puts strict limits on moral competence. While there is nothing necessarily illicit about arguing from moral performance to moral competence, it requires that the research support very strong claims about moral performance in order for it to imply much of anything about moral competence. The situationist literature is not so strong. Accordingly, the philosophical situationists essentially infer from the fact that most of us have failures of moral performance that

we therefore lack moral competence of the sort posited by virtue theory. In other words, they take the fact that no one is perfect to imply that we do not have the potential to be perfect. This inference is faulty.

Each of the representative experiments I described fails to provide the sort of evidence that would challenge virtue theory's moral competency claims. In the Milgram experiment, seven out of forty experimental subjects refused to administer any shocks beyond the point where the confederate rescinds his consent to participate (Doris 2002, 41). This does not imply that those people are virtuous, but it does not disqualify them from virtue either. Indeed, 17.5% would be a surprisingly high base rate of virtue in our society. Likewise with the Good Samaritan experiment, where 10% of the test subjects helped even in the hurry condition (Doris 2002, 34). Both of these experiments provide clear evidence that people can resist the situational pressures and do the right thing. It is no objection to virtue theory to show that most people do not have virtue, it is only a challenge for teachers of virtue to try to overcome.²⁰

The honesty experiments are harder to quantify in any particular way. On this topic, however, Sabini and Silver's critique about correlations is appropriate (Sabini & Silver, 540-544). The fact that the population in general had a low cross-situational correlation does not imply anything about whether there existed honest individuals with high correlations. Furthermore, it is important to note that the Hartshorne and

²⁰ Doris briefly attempts to address this sort of argument at the very end of his book. I believe his response is entirely unsuccessful, as I argue below in section 2.6.

May experiments were performed upon children. For an Aristotelian, children do not have full virtues. At best, these children would be learning honesty, and we ought not to expect perfect performance from those learning virtue.

Let me try to put my argument another way, in terms of the moral buzz phrase “ought implies can”. This is a principle that is often given as an axiom with which any moral system must comply. I think the philosophical situationists are using situationism to limit “can” in an illicit way, and then use this sort of axiom to attack virtue theory itself. In particular, the situationists use “can” in a performance sense, while for virtue theorists the “can” in this phrase is a “can” of moral competence. To put this another way, the situationists are using a valid *modus tollens* of the form: P1 - If we ought to be virtuous then we can be virtuous; P2 - It is not the case that we can be virtuous; C - It is not the case that we ought to be virtuous. But they are equivocating on “can”, and the “can” in P1 is different from the “can” in P2. No amount of evidence about failures of virtuous performance necessarily shows that “it is not the case that we can be virtuous” from the point of view of moral competence. To take an extreme example, consider Lawrence Becker’s writings on Stoicism (Becker 1998). Becker has an extremely strong version of what counts as virtue, but even for his theory it is no problem that we cannot find a Stoic sage who embodies virtue, even though Becker’s Axiom of Futility essentially formalizes “ought implies can” into his logic. Becker agrees that we ought not to take on impossible projects, but he also thinks that we ought not to assume projects to be impossible simply because they have not yet been done: “when we know that a given undertaking is

impossible, we are prohibited from undertaking *it*. This does not imply that we ought to refrain from other efforts to make it possible” (Becker 1998, 44-46, 99).

Situationism does not even show that the virtue traits it challenges do not exist, only that they are rare if they do exist. This is hardly an argument against the human competency to virtue, any more than arguments about human performance in general allow us to infer statements about human competence. And even if situationism did show that virtue traits do not exist yet, that is no argument against taking actions to try to make those traits possible.²¹

My argument here is, I think, widely accepted in other areas, and it should be similarly acceptable here. Compare the case of human physical competencies. Usain Bolt’s sprinting ability is legendary. In the one hundred meter dash, the world’s fastest runners first broke ten seconds in 1968. It took 25 years for someone to go under 9.9 seconds, Carl Lewis in 1991. In 1999, Maurice Greene was the first to break 9.8, and only four men have ever done that in competition. Usain Bolt, in 2009, went under 9.6, and he has run nearly two tenths of a second faster than any other human in competition.²²

Prior to Bolt’s amazing performances, various researchers had made predictions about the possible lower bounds of human performance in the one hundred meters. As one mathematician wrote, “To measure the effect Bolt has had so far and may have in [the] future we calculated the lower bound excluding his three

²¹ See my arguments below in 2.5 about the value of seeking to be grateful, even if one fails to achieve a full-fledged trait of gratitude, for an empirical example of how this might work.

²² These numbers are all available at http://en.wikipedia.org/wiki/Men%27s_100_metres_world_record_progression.

records. The result is 9.62... so Bolt has already beaten the estimated ultimate record based on other runner's times" (Noubary, 1) (Bolt's fastest time in competition to date is 9.58; some think he may go as low as 9.4). Bolt's sprinting performances have shown that predictions of human running competencies, on the basis of human running performances prior to Bolt, were simply wrong. One might claim that the theoretical models used to predict such competencies were flawed, and it is clear they were flawed because Bolt exists. But I think that this story shows two more important things: 1) predicting human competencies on the basis of human performances is risky business, and we should always err on the side of unexpected competencies outside the range of known human performance, and 2) sometimes we need a genius, physical or otherwise, to show us what competencies we actually possess. The second claim is particularly important, as there is evidence that other sprinters have run faster themselves since Bolt came on the scene. According to Professor Steve Haake, "It is a little jump in performance when [Bolt] appeared in that year [2008]. If we look at the top 25 sprinters and take Usain Bolt out of that list, so that you just analyze the other 24, you still get this step change" (Connor 2012). Bolt is not just outside the previously predicted bounds of human performance himself, he stretches the bounds of human performance for others by showing that our competencies are greater than we thought.

Philosophical situationists are making a mistake similar to the one the mathematicians made in the case of sprinting: they are being unduly pessimistic about human competencies on the basis of known human performances. In fact, they are

making a worse mistake, because they are ignoring the fact that in all of the situationist experiments, some of the participants in fact did the right thing. It is as if someone were to watch a race with Bolt in it and then declare that his times were outside the realm of human competency because the other runners were too slow.

As a general matter, I think that virtue theorists, myself included, are making claims about moral competencies and not performances. It is true that the classical Aristotelian notion of virtue involves action, and that a virtue is not a virtue unless it expresses itself in action. But virtue, on any robust account, is extremely hard. And virtue is, even for Aristotle, the actualization of a human potential (see, e.g., NE 1103a19-27), and the failure to actualize such a potential does not warrant the inference that the potential was not there in the first place.²³ Particular failures to be virtuous do not warrant giving up on trying to act virtuously any more than failing a particular lift is reason to give up on weightlifting. There is value in trying to be as

²³ Becker, for example, does not believe that he knows anyone who has achieved the level of a Stoic sage, a perfectly virtuous actor, and even doubts the possibility of such achievement for us (Becker 1998, 120). Yet he does not think that Stoic sagehood is by nature impossible, for then seeking it would violate the Stoic axiom of futility (Becker 1998, 120). These beliefs may seem contradictory, but I think they are better interpreted as making it clear that the virtue Becker is talking about is an issue of human moral competence, not performance. I'll briefly sketch out why this is the case. First, it is implausible to think that it is possible for a person to have perfect virtue in the sense that they are 100% certain to do the right thing any situation, no matter the circumstances. It may be the case that this is the Stoic view, but it seems implausible in light of basic facts about our physical and psychological frailty; such a person seems more divine than human. But an extremely virtuous person is still possible, and this makes the Stoic Sage still possible as a matter of competence, if grossly unlikely as a matter of performance. Imagine a very great Stoic who is 99.9% accurate in determining the virtuous action and doing it in all situations. They are, by any account, an unusually virtuous person. And to make extremely generous assumptions, say they are confronted by only *one* test of virtue a day, and that we will only start measuring their virtue when they are twenty-one. Such an exemplary person has an approximately two and a half in one hundred thousand chance to make it to their fiftieth birthday without making a mistake of virtue. Even the best of us are much less accurate in finding the virtuous action, and much more overwhelmed by decisions and actions that demand virtuous action. But even so, in this example the Stoic Sage is possible as a matter of moral competence, but so unlikely as to never have happened as a matter of moral performance.

strong as I can be, just like there is value in perfecting my agency as far as I can perfect it.

Doris seems aware of the sort of argument I am making, and he discusses it in relation to perfect virtue. In relation to the inseparability thesis (Doris' term for a unity or reciprocity of the virtues thesis), he says,

“Defenders may claim that inseparability holds only for perfect virtue; they can thereby allow the abundant appearances of separability and simply insist that these cases involve something less than the full realization of virtue... since we can expect perfect virtue to be extremely rare, neither a paucity of cases suggesting inseparability nor a plethora of cases suggesting separability need give defenders of inseparability pause” (Doris 2002, 22).

I am sure Doris sees how this perfect virtue thesis could ground a more general response to his situationism critique; if all he has shown is that perfect virtue is rare, that is hardly a novel claim. But Doris does not appreciate, I think, how the perfect virtue dodge actually undermines his argument. For there is an implicit assumption in his style of argument, that a partial or incomplete or imperfect virtue looks just like perfect virtue, just with less of it, so to speak. He does not explicitly say this in his book, but it is an implicit assumption in his articulation of the globalism thesis, and he makes it explicit in recent work when he says that the empirical conclusion “behavior is *not* typically ordered by robust traits” somehow falsifies virtue theory (Merritt 2012, 357-358). For to argue that perfect virtue satisfies the requirements of globalism, as I think virtue theory does say, is not to claim that imperfect virtue satisfies some, or any, of the requirements of globalism. Perfect virtue should be typically ordered by robust character traits, but perfect virtue is not typical, and imperfect virtue is not necessarily typically ordered by robust traits. The virtue

theorist is asserting globalism as a statement of human competence, and at best only extremely rarely as a statement of human performance.

While I have exploited the notion of a competence-performance distinction in this section, my argument has, I think, a broad intuitive appeal. For if virtue theorists were only talking about human performances, as Doris seems to think, and not also human potentials, then virtue theory would be transparently wrong. It would, for instance, be historically contingent in an unacceptable way. If there were a time during which all or mostly all humans were engaged in some vicious activity, this would falsify a virtue theory. Any student of history can generate a long list of such cases; the case of slavery is sufficient to make my point. I do not think anyone should take the widespread acceptance of slavery in the ancient world as an argument against the theoretical claims of virtue theory, but that is what would be implied by Doris' argumentative strategy. This widespread acceptance of slavery does constitute good evidence against the ascription of full or perfect virtue to any individual in the ancient world, but does not call into question the human potential for perfect virtue.²⁴ If anything, the historical story of slavery, of a general overcoming of such widespread, socially-acceptable viciousness, is a hopeful story for human potential, not an argument against it. My hope is that our understanding of situationist psychology begins an analogous story: that we may someday learn to overcome the situational

²⁴ Annas has an insightful discussion of what forms of virtue would be available in the ancient world in light of its acceptance of slavery (Annas 2011, 58-65).

factors that pervert or distort our moral judgments, and that we may thereby become more virtuous.

2.3.2 The Inconsistency of Vice

In addition to the competence-performance distinction, Sabini and Silver's criticism that Doris claims vice is governed by the same sort of robust character traits as virtue deserves greater discussion. This criticism is worthy of a serious response from the situationist because it is deeper than Sabini and Silver understood. It infects Doris' understanding of virtue ethics throughout his book, and undermines many different argumentative strategies he uses. If Doris has good arguments in favor of his view that there is consistency to vice, he ought to share them. But without a response to this argument, it seems that virtue ethicists and the situationists may simply be speaking past each other.²⁵

Doris considered this point in his original work, saying:

"The thought that the virtues have this sort of evaluative dimension is respectably Aristotelian; Aristotle (1984: 1105a30-b1) maintains that genuinely virtuous activity is undertaken knowingly and for its own sake. It is less clear that this thought neatly applies to vices and other negatively valenced traits of character; is cowardly behavior necessarily the expression of the actor's values?²⁶ This is a reasonable concern, but I'll insist that cowardice and other negatively valenced character traits do involve the relevant sort of evaluative dimension; perhaps the coward values safety more than honor, loyalty, and dignity. To get tolerably clear on this would take rather more discussion of evaluation, and the evaluations associated with each trait of character, than I'm going to provide. For my interest is not so much what distinguishes character and personality traits as what they have in common: behavioral consistency as a primary criterion of attribution" (Doris, 20).

²⁵ This sort of misunderstanding is probably, I think, what led scholars like Julia Annas to be so dismissive of Doris' work (See, e.g., Annas 2005, 637 n.2).

In his footnote 26, he notes “Gilbert Harman has pressed this objection against my account. I’ve reluctantly, and no doubt imprudently, declined to heed his warning” (Doris, 177). Unfortunately, Doris does not offer any reason to think that vice has an evaluative consistency akin to virtue. Aristotle clearly rejects this thesis. Aristotle’s moral psychology famously includes at least two sorts of people who act wrongly, the vicious and the incontinent (*akratic*), who do not have robust, integrated character traits at all (See generally NE Book VII). As Aristotle says, “virtue preserves the principle, whereas vice corrupts it; and in actions the end we act for is the principle” (NE 1151a15-17). The virtuous are expected to act consistently in accord with a principle, but the vicious are not expected to act consistently in accord with some form of anti-principle, the vicious rather reject or corrupt the principle of the virtuous. The incontinent on the other hand, while they recognize the principle, continually fail to act on the principle they rationally endorse as their end (NE 1145b11-13). The incontinent person “is like a city that votes for all the right decrees and has excellent laws, but does not apply them” (NE 1152a20-21). It is characteristic of the *akratic* individual that they act on orectic impulse, not rational principle. It would have helped Doris get a hearing among virtue theorists if he had focused on situationist experiments that did not rely on the consistency of evaluative factors that virtue relies upon. It is particularly unclear why Doris continues to include the evaluative consistency of vice in his arguments today. In recent work, for example, Doris articulates what he calls “the skeptical argument” pressed by situationists against virtue theory in the following *modus tollens* form:

“(1) If behavior is typically ordered by robust traits, systematic observation will reveal pervasive behavioral consistency.
(2) Systematic observation does *not* reveal pervasive behavioral consistency.
(3) Behavior is *not* typically ordered by robust traits” (Merritt 2012, 357-358).

Doris thinks this argument is sound (Merritt 2012, 358), but even if it is, the conclusion is not something that is a challenge to virtue theory, for virtue theory does not endorse the negation of proposition (3). Aristotelian virtue theory generally does assert a related specific proposition, something like “virtuous behavior is typically ordered by robust virtuous traits”, but it clearly does not make the general claim that all behavior is typically ordered by robust traits. In an Aristotelian virtue theory, many people will not typically behave according to robust character traits, even if they recognize and rationally endorse the proper virtues, and it is a mystery why Doris does not acknowledge this in his argument.

This misunderstanding crops up in other parts of Doris’ argumentation. For example, Doris claims that his suggested moral psychology, one that does not have notions of general character traits, would improve our moral discourse. As an example, he discusses what we should do in relation to personal character evaluations:

“If we take situationism to heart, we will eschew some familiar ways of thinking about other people. We will be reluctant to evaluate persons in terms of robust traits or evaluatively integrated personality structures, because we will think it highly unlikely that actual persons instantiate such psychological features. Accordingly, we will be unwilling to speak in terms of general evaluative categories as ‘good person’ and ‘bad person.’... Robust traits and evaluatively integrated personality structures are constructs that underwrite substantial stretches of evaluative discourse, but these stretches too often enable unfair condemnations, on the one hand, and unwarranted approbation, on the other” (Doris 2002, 115-116).

As a practical matter, Doris has described “substantial stretches of evaluative discourse” to which Aristotelian virtue ethicists could object with as much vigor as Doris. Aristotelian virtue ethics does not imply that one bad act shows that the actor adheres to some sort of evil or vicious principle. I would think that a virtue ethicist, and certainly anyone trying to embody justice, would limit condemnation to morally blameworthy actions and acknowledge the difference between bad character and not good character. One bad act may disqualify someone from being virtuous, but it does not necessarily label them as vicious.

Once we see that this reformation of ordinary ethical discourse is also in line with virtue ethics, what about praise for virtuous action? Does virtue ethics lead to “unwarranted approbation” of a morally serious kind? I ask “morally serious” because it is not clear to me what problems are really caused by unduly ascribing virtuous traits to people. There is certainly the potential for pragmatic problems, for instance when you rely on the character judgment of honesty and loan your friend some money which he never returns, but I am not clear on what the moral problem is with over-attributing virtue traits on the basis of virtuous acts. This is a problem intrinsic to the concept of virtue independent of the situationist evidence.

This being said, it is nevertheless fair of Doris to be critical of virtue theory here. I think that it is possible to do some things right, as the virtuous agent would do them, without having a virtuous character trait, but not all virtue theorists agree. To illustrate, Doris tells the story of Max Redlicht, a gangster in Nazi-occupied Cracow (Doris 2002, 130-131). He was a part of a group of men that was forced by the

Germans to “spit on the Torah scroll in the city’s oldest synagogue.” Redlicht apparently said “I’ve done a lot. But I won’t do that.” He was killed, along with all the others in the synagogue. Doris argues that his reading of this story, “Redlicht was an ignoble person who behaved nobly—is a perfectly natural one.” Doris goes on to use this story to criticize character-based ethics:

“Furthermore, should it in such cases be determined that an admirable deed was not the function of an enduring and admirable character trait, I suspect that many people would still be inclined to praise the person who did it. Many seem to hold dear the possibility of redeeming actions, the hope that there is a flicker of goodness even in the worst sorts; I think those that do such things are justly credited, however their character should finally be judged. But if responsibility is linked to character assessment, such reactions are apparently prohibited” (Doris 2002, 131).

I agree with Doris, and I think virtue ethicists should leave room for redemptive action, conversion experiences, and even cases like Redlicht’s where an agent suddenly, and likely temporarily, comes to see things in the right way.

I admit, however, that there are virtue ethicists who say the sorts of things Doris is critiquing here. For example, he might have someone like Rosalind Hursthouse in mind, who discusses a similar case. Hursthouse asks us to consider an agent

“whose other intentional actions, reactions, and talk license the descriptions ‘selfish’, ‘unjust’, and ‘timorous’... [who] suddenly surprises us. She gives with an open hand, spurns the offer of an unfair advantage, she speaks out boldly in defence of an unpopular colleague—and gives the appropriate... reasons. It turns out that she is in love, or on top of the world because of recent success; love or success has momentarily transformed her. And shortly afterwards, she lapses back into behaving as she did before” (Hursthouse 1999, 133).

Hursthouse thinks there is something lacking in this agent's actions. She says, "The agent who surprises us by her virtuous actions when momentarily transformed by love or success seems to recognize the value of the [virtuous] actions only when it is, as it were, lit up for her by her love or success" (Hursthouse 1999, 135).

But I do not think there is any requirement that a virtue theory be so demanding about what right action requires. In particular, as I will argue below in Chapter 3, I think a more reasonable virtue theory would consider us to be very much in this woman's position nearly all of the time. In any lifetime, the emotions and reasoning that generate virtuous actions begin, if we are lucky, at some point. They can come and go. Part of the work of virtue is to cultivate and maintain these dispositions. Only a consistent disposition counts as a virtue, but it does not seem like a good idea to claim that right actions are only those that come from a settled disposition. I believe that virtue theory can be subtle enough to distinguish between the goal of a virtuous character and the marks along the way to developing that character, which would involve right, or at least increasingly morally better, actions. Indeed, for those of us not already virtuous, we can only hope that we might find such love or success that will illuminate the world in such a way as to make rightness and virtue apparent! Someone who intermittently sees virtuous reason for action still sees them, and is capable of acting from them. It is not full virtue, which is worth noting, but I do not see a strong reason to undermine the value of a right action simply because it is not the part of a complete virtuous life.

I think Doris is right to criticize the view espoused here by Hursthouse as unreasonably demanding, and to suggest that our moral discourse would be improved if we discarded it. This reform of virtue ethics would be for the better. Given what I have argued about how Doris' moral psychology fits virtue better than vice, it is perhaps not surprising that on this point, his arguments have some force against a theory of virtue but miss the point against a theory of vice. Doris should distinguish virtue and vice in order to engage more directly in dialogue with virtue ethicists.

I am interested in the situationist response to these questions in particular because I think that virtue ethicists can learn a lot from empirical psychology. If the situationist challenge and the virtue ethics tradition can be brought into alignment, to be talking about the same issues, then there is more room for each side to learn from the other. Much of the rest of my thesis, in fact, will be arguing for some of the things that I think virtue theory can learn from empirical psychology. These are positive arguments; I believe the empirical evidence is generally encouraging for virtue ethics. But I also think that some of Doris' criticisms of virtue ethics have force, and virtue ethicists ought to take them seriously. But for criticism either way to stick, Doris and the situationists have to be talking about the same thing that virtue ethicists are talking about. In particular, Doris and the situationists must take seriously the fact that virtue ethics does not demand that vice operate with general, cross-situational consistency.

2.4 Second Reply to Doris: Positive Psychology

The situationist experiments were clever, surprising, and despite my criticism of Doris' arguments based on them, I believe very important for moral psychology and moral philosophy. But they were and are the tip of the iceberg, so to speak, when it comes to understanding human moral reasoning. This is important because Doris might accept my argument in section 2.3, but reply with this argument: it may be true that virtue could exist in rare cases, but if that is the case, virtue ethics is still bankrupt because it tries to base a general ethics on the behavior of extraordinary people. We need an ethics for everyone, and situationism shows that this is not virtue ethics. As he says, he is "concerned with what kind of moral-psychological reflection can help people do right... and help people live more ethically desirable lives" (Doris 2002, 109). But his focus on one area of empirical psychology warps the lens through which he pursues this goal to the extent that he unfairly dismisses virtue moral psychology as empirically inadequate without taking account of its empirical support. In this section, I argue that Doris' own normative claims about practical ethics do not follow from his basic premise. Even assuming for the sake of argument that general character traits do not exist, or are so rare as to form a poor ground for a general ethical theory, there is empirical evidence that simply aspiring to a general virtue trait can itself lead to well-being.

Positive psychology, a relatively recent movement within psychology, addresses this issue by examining human strengths. Positive psychology is "as focused on strength as on weakness, as interested in building the best things in life as

in repairing the worst, and as concerned with fulfilling the lives of normal people as with healing the wounds of the distressed” (Peterson and Seligman 2004, 4). This movement involves the application of standard psychological research methods to topics like positive experiences, positive traits, and the effects of institutions on these positive experiences and traits (Peterson & Seligman, 5). Among other projects, this movement purports to test virtue concepts through empirical methods. As such, it is highly relevant to my topic, and to the extent it has produced any concrete results, those results are important for philosophers to reckon with.

Moral philosophers have not yet really grappled with the results of positive psychology. As far as I can tell, this is because many do not think that there is much to grapple with yet. Annas, for example, criticizes Peterson and Seligman’s work to create a diagnostic system for studying virtue, calling it “premature and implausibly definite” (Annas 2011, 98 n. 14). On the other hand, she speaks approvingly of another major figure in the positive psychology movement, Mihalyi Csikszentmihalyi (Annas 2011, 70-71). Others claim that “few philosophers have made an effort to take account of the empirical research on this topic. In large part, this is because this psychological research, particularly the burgeoning field of ‘positive psychology,’ is so new that philosophers have not had a chance to address it” (Tiberius and Plakias 2010, 403-405).²⁶ While I agree with some of these claims, the field of positive psychology is actually not, I think, so new as to excuse philosophers from addressing

²⁶ Tiberius and Plakias do note exceptions to this rule (403 n. 2). Most intriguingly, they cite an unpublished manuscript by Peter Railton.

it. The label ‘positive psychology’ might be new, but empirical psychologists looking at these sorts of questions have been around for much longer. Indeed, in the next chapter I will extensively discuss the work of Ellen Langer, whose decades of work cover a similar conceptual landscape to the positive psychologists. There is plenty of empirical work in this field for my purposes here, which are to show that the empirical data on human character is not as univocal as situationists would have us believe.

I will use empirical work on gratitude as a representative example of the work done in positive psychology. Peterson says: “Certainly, we know that the habitually grateful among us are happier than those who are not” (Peterson 2006, 33). But at best, the psychological research supports the modest claim that most people experience a greater sense of well-being the more gratefully they act. This research points to a suggestion that, I think, undermines the basic situationist argument: imperfect virtue could still make us happy. Even if Doris were right about the impossibility of the generality thesis, and the other claims he thinks he has refuted, there is reason to think that trying to be virtuous is still a reasonable path to well-being and happiness. And the fact that there is empirical evidence indicating that subjects experience increased subjective senses of well-being by simply trying to be more grateful indicates that pursuing virtue could certainly be an “ethically desirable” life, in Doris’ parlance (Doris 2002, 109), even if we took as true Doris’ claim that a general character trait of gratitude is unachievable. Doris, and the

situationists in general, have lost sight of this basic structure of a virtue theory argument in their situationist critique.

The claim I want to take from the gratitude research is summarized by, somewhat surprisingly, a cognitive psychotherapist:

“First, research now suggests strongly that forgiveness and gratitude are relevant for psychological, physical, and relational well-being. Studies also indicate that forgiveness and gratitude can both be facilitated experimentally through relatively simple psychological interventions. Finally, studies show that forgiveness and gratitude are not only realistic psychotherapeutic goals that can be attained with existing psychological interventions, but also, by encouraging people to experience forgiveness and gratitude, they may experience gains in other areas of their lives.” (Bono & McCullough 2006, 7)

I would not generally take recommendations for psychotherapeutic practices as important for moral theory. But the claim I am trying to address here is one about the effect or value of pursuing a virtuous disposition for humans. Doris is claiming that this pursuit is damaging, because a general disposition of virtue is impossible (Doris 2002, 108), and therefore it is highly relevant when cognitive psychotherapists say that helping people try to be virtuous makes them happier, regardless of whether such a state of character is possible. That question does not even enter into the concerns of a cognitive psychotherapist: if a patient is happier and experiences more well-being, then the therapy is successful.

The research that Bono and McCullough rely on is extensive. A study by McCullough and others serves as one representative example (McCullough, Tsang and Emmons 2002). The authors’ general summary of the series of studies described in this paper is,

“Compared with their less grateful counterparts, grateful people are higher in positive emotions and life satisfaction and also lower in negative emotions such as depression, anxiety, and envy. They also appear to be more prosocially oriented in that they are more empathic, forgiving, helpful, and supportive than are their less grateful counterparts.” (McCullough et. al. 2002, 124)

It is important to note that the authors are reporting statistical correlations here – people with a higher measure of the grateful disposition, using both self- and other-reporting measures, also tend to have higher measures in the various positive emotions and lower measures in the various negative categories. Nothing in the study shows that gratitude produces, or causes, these effects. The grateful disposition and other measured qualities could all be effects of some other, more fundamental cause.²⁷

Other studies have looked at the effect of inducing subjects to practice gratefulness on their subjective well-being (Emmons and McCullough 2003; Watkins et. al. 2003). In these sets of experiments, the researchers asked a portion of their test subjects to think about and write down things that they were grateful for (Emmons 2003, 379), even to the extent of asking subjects to write letters of thanks to those to whom the subjects were grateful (Watkins 2003, 445). In all of the studies discussed in these papers, the experimental subjects with the gratitude inducements condition

²⁷ The study authors did try to analyze whether the effects of the grateful disposition could be understood in terms of other already accepted psychological traits. For instance, they concluded that the “Big Five personality taxonomy” explained some of the variation found, but “no more than 40% to 45% of the variation in the disposition toward gratitude” (McCullough et. al. 2002, 124). No amount of such testing, however, could show that all of the evidence measured is not possibly the result of some further, undiscovered personality factor.

experienced an increase in measures of subjective well-being (Watkins 2003, 448; Emmons 2003, 386).

None of this research shows that a grateful disposition exists, or that being grateful leads to Aristotelian *eudaimonia*. Psychological measures of subjective well-being and happiness fall far short of the standard of Aristotelian *eudaimonia*. But these studies do serve as confirmation,²⁸ albeit a weak confirmation, of a virtue moral psychology. For if the existence and practice of a virtue trait, in this case gratitude, led to a decrease in the happiness and sense of well-being of an agent, then something would be very wrong with virtue moral psychology. Being virtuous can require doing painful things, but on my theory at least, being virtuous is also constitutive of the good life. And the good life could not plausibly require one to cultivate a trait that systematically undermined one's subjective well-being. Therefore, the fact that gratitude seems to promote subjective well-being is consistent with, and confirmatory of, a virtue moral psychology.

It is worth dwelling a bit more on this claim, for it is important to my argument that these studies of gratitude confirm a virtue moral psychology, not just that they are consistent with it. Mere consistency with a virtue moral psychology would be interesting, but not an argument against Doris and the situationists. There

²⁸ The concept of confirmation is a complex one. In this section, I mean “confirms” to be synonymous with something like “eliminates alternatives to”. That is, a study confirms virtue moral psychology if it eliminates alternatives to that moral psychology. I take this to be a sort of colloquial version of Bayesian confirmation theory. I do not intend to suggest or argue that Bayesian confirmation theory is right, but rather to explain my argument in terms of one rigorous theory of confirmation. In this case, the gratitude studies confirm virtue moral psychology because they eliminate particular ways that virtue moral psychology could be wrong. They are inconsistent, for example, with the hypothesis that gratitude makes humans systematically miserable, which would go against the basic claims of virtue theory, in my view.

are many experiments consistent with clearly false hypotheses. Psychological experiments about gratitude are consistent with, for instance, a theory that the world is flat. But it would be absurd to claim that therefore these psychological experiments confirm a flat-world hypothesis.²⁹ They are beside the point. If the gratitude studies were merely consistent with virtue moral psychology, they would provide no response to Doris. It would only seem that they do, because the substantive area of study is so similar to the contested issues.

But these studies do confirm a virtue moral psychology in the same manner that the situationist studies confirm Doris' position, that there is no such thing as a general character trait. The situationist studies eliminate certain possible ways that general character traits could exist.³⁰ The gratitude studies eliminate certain possible ways that our moral psychology could fail to fit a virtue moral psychology. Both are relevant to the issue, but the empirical jury is still out, so to speak, as to how human moral psychology matches the claims of virtue theory.

²⁹ The notion of relevance in Bayesian confirmation theory can be explained as: an event E is relevant to the probability of some X if the probability of X given E, $\Pr(X/E)$, is not equal to the probability of X, $\Pr(X)$. E is irrelevant to X in the case that $\Pr(X/E)$ is equal to $\Pr(X)$; in other words, E has no effect on the probability of X.

³⁰ This is, in essence, a way to restate my argument from the first section. Doris fails to realize that the situationist studies at best *confirm* his thesis about character traits, they do not eliminate virtue moral psychology as a possibility. The situationist studies are inconsistent with particular theses about the virtues, such as a thesis that the virtues are widely held. To disprove a virtue moral psychology, the studies would have to falsify a necessary claim of that moral psychology, which it has been the argument of my first section that they fail to do. In this section, I am offering further evidence that the jury is still out on this empirical question by pointing to studies that eliminate some of the possible ways that virtue theory could be *wrong*. Of course, to prove virtue moral psychology, we would have to disprove *all* the ways that virtue moral psychology would be wrong, or at least all the ways we find sufficiently plausible.

In addition to the fact that the empirical evidence is more ambiguous than Doris indicates, positive psychology poses a substantial challenge for Doris and other philosophical situationists based on the fact that acting virtuously increases subjective well-being.³¹ This is the case whether or not those actions stem from an enduring disposition, and therefore aspiring to be virtuous, even if virtue does not exist in any human, might still be conducive to human happiness and flourishing. Note that this is not a claim about the particular character traits that are or are not possible for humans, it is rather about what sorts of human actions make us happier. And virtue theory might be consistent with the claim that a fully integrated, perfectly virtuous human will never exist but that the pursuit of virtue is still constitutive of human happiness. The gratitude research unanimously shows that aspiring to be grateful increases human subjective well-being. If seeking gratitude leads to increased well-being, it seems intuitive to suppose that seeking virtue in general could lead to increased well-being, which is consistent with the virtues being a part of happiness and the good life. Indeed, investigating this possibility is part of the research program of positive psychology.

If this train of thought is correct, then Doris' argument turns out to be tangential to the real issue, which is: what are we to do to be flourishing humans? Aristotle thought that learning to be virtuous involved learning the characteristic pleasures one gets from doing good (NE 1104b5-1105a17). Positive psychology

³¹ Annas places aspiration central in her account of virtue. A major focus of her book is comparing virtue to skills that involve "the *need to learn* and the *drive to aspire*" (Annas 2011, 16). See Chapter 4 for more in depth discussion of the relationship between aspiration and virtue proper.

supports this claim. But even if, for the sake of argument, Aristotle were wrong and Doris were right about the impossibility of general virtues, this research suggests that trying to be virtuous could still increase subjective well-being. The virtuous acts increased well-being, whether there is a stable psychological state to be labelled a “virtue” or not.

2.5 Conclusion

I have argued that Doris’ basic mistake is to confuse the goals of an Aristotelian virtue ethicist with those of a certain type of descriptive social or psychological scientist. The Aristotelian is trying to describe, as I argue above, human potentials, the conditions of human flourishing. The descriptive psychologist is interested in examining how people actually act under experimental conditions. A failure to find human flourishing among a wide-spread section of the population, while disheartening, is hardly fatal to a virtue theory.

But Aristotelian virtue theory does make some claims that are amenable to scientific examination. In particular, I claimed that the virtues are constitutive of human flourishing, that a part of living a flourishing human life is being virtuous. If that is the case, it seems that humans who embody, or seek to embody, the virtues should experience more flourishing, more well-being. And in fact, the positive psychology program seems to show that such increases in well-being in fact do exist, and are correlated with the attempt to embody virtues. The psychological research program of positive psychology seeks to empirically examine exactly these sorts of claims, and has given us evidence that exactly these sorts of aspirations are important

for human happiness. In addition, the positive psychology program explicitly responds to some of Doris' claims. Doris says, "central ethical practices can survive, and indeed will benefit from, the revisions I advocate [eliminating traditional notions of character]" (Doris 2002, 108). The gratitude research I have cited, however, indicates that giving up on the aspiration to gratitude, whether or not its achievement is possible, would not benefit us at all.

Doris seems to assume that because virtue is rare, or hard, aspiring to it will inevitably lead to shame, sadness, or disappointment for failing to achieve some ideal state. As he says, "Commitment to globalism threatens to poison understandings of self and others with disappointment and resentment on the one hand and delusions and hero-worship on the other" (Doris 2002, 169). But he fails to appreciate the evidence that aspiring to virtue increases well-being, it does not undermine it. Positive psychology provides empirical support for this contention, but frankly, a closer reading of virtue ethicists, such as Annas, also provides the theoretical underpinning for these claims. Her theory, for example, does not teach us to be ashamed of our imperfection, but to see value in the state of learning to be virtuous and not only in the perfectly virtuous state itself (See, for example, Annas' discussion of learning to be virtuous; Annas 2011, 52-58).

Situationism is an important body of research for our moral lives, for it illustrates a particular way in which we are weak. It does not show that virtuous aspiration is in vain, it rather points out a powerful way in which our virtuous aspirations are susceptible to being thwarted. Situationism should be a part of moral

education, not an argument against it. In fact, this is where Doris' later work suggests a path for potential reconciliation between situationists and virtue theorists, as I will try to argue in Chapter 3 on mindfulness.

My criticisms of Doris having been voiced, it is important for me to note and discuss Doris' own suggestion, near the end of his work, of a line of argument similar to mine (see Doris 2002, 149-153). He says "Perhaps the foregoing comparison misconstrues approaches to deliberation emphasizing character. The examples I have given concern the *description* of oneself under which one deliberates, while virtue ethics may be thought to concern the *ideal* under which one deliberates" (Doris 2002, 149). I take this to be an acknowledgment of something like my competence/performance argument laid out above, that his charges of descriptive adequacy may not go through against a properly modest virtue moral psychology. I accept for the sake of argument that my view of virtue is, in some sense, concerning the "ideals" under which one deliberates.

But Doris' discussion of this objection only serves to reinforce my argument that he has not understood the virtue theorist properly. He goes on to contrast two models of this ideal-based virtue theory, an "emulation" model and an "advice" model (Doris 2002, 150). On the emulation model, the model Doris says he has been assuming up until this point in the book, an agent tries to do what a virtuous person would do in any particular situation, to "approximate the psychology and behavior of the moral exemplar... [b]ut there is also the possibility of an *advice* model, where deliberation involves consulting the advice of the ideally virtuous agent" (Doris 2002,

150). The distinction is made clear through situationist examples. Doris posits an agent that is invited over to dinner at the home of an acquaintance to which the agent is inappropriately sexually attracted. The agent is not perfectly virtuous, and so fears the situational pressure of an intimate dinner might overwhelm the agent's ability to resist sexual temptation. On an emulation model, the agent should do their best to go to dinner and not be tempted, for a virtuous person would be free to go to dinner and resist any such temptation. But on the advice model, the ideally virtuous advisor might be able to take into account the agent's weakness and recommend that the agent avoid a situation that will unduly test their ability to stick to ethical commitments (Doris 2002, 105).

This seems to me a false dichotomy. For the proper emulation question is not: what would a virtuous agent do if we imagined them in my situation. As Annas points out, for instance, this is not always even a possible question, because there are situations a virtuous agent would simply avoid in which we might find ourselves. As Annas says, "it would be problematic to say that a truly virtuous person could be in [such a] position, since the truly virtuous person has the understanding not to be in the situation in the first place" (Annas 2011, 44). The question, rather, is how an agent in the challenging situation ought to act to be virtuous. And this is not impossible for us to ask, simply because a perfectly virtuous person would not get themselves into such a pickle, because no agent, no matter how virtuous, is free of situational factors. Being virtuous is being sensitive to situational factors and properly taking them into account in deciding how to act. Emulation is not trying to copy someone free from

causal influences in the real world, situational or otherwise. It is trying to act as someone who aspires to virtue in all cases. And in a particular case, say with our dangerous dinner, the virtuous agent would have to take into account their weakness to such a temptation in deciding whether or not to go. It might be best if the agent could enjoy the dinner without succumbing to inappropriate temptations, but their desiderative make-up might be such that this is not an open option for them. The case is no different, on my view, from one in which some scope of action is closed off to the agent by different factors, factors that we generally do not consider a challenge an agent's virtue, such as physical impairment, logistical limitations, and the like. A fully virtuous agent might not intentionally get themselves into a situation where they are inebriated without a ride home, but that does not mean that we must appeal to some sort of "advice" fiction in order to ask: what would be virtuous in this situation? For a virtuous agent could certainly end up in this situation, if for example a nefarious actor forced them to imbibe alcohol. The virtuous person would still act in this situation, and we should have no problem considering what they would do.

Doris also claims a basic victory if it be agreed that "character ethics should favor psychologically realistic ethical reflection of the sort I've advocated" (Doris 2002, 151). This is a statement I certainly agree with, at least up to the word reflection. But a large part of my criticism is that the sort of reflection Doris advocates is based on only part of the psychological evidence available to us. I suspect most every virtue ethicist would agree that character ethics should favor

“psychologically realistic ethical reflection”. But I dispute that Doris has captured this category in his book.

3. A Virtue Ethics Take on Situationism: Mindfulness as a Virtue

Despite the weaknesses in Doris' position, I fundamentally agree that modern social psychology is relevant to virtue ethics. Empirical social psychology is not just a basis for challenging the empirical adequacy of virtue ethics, it can also be a source of inspiration and insight. In particular, the empirical evidence reviewed in this chapter supports the claim that mindfulness, roughly speaking the disposition to be non-judgmentally attentive to one's own thoughts and feelings, is a good candidate for a virtue, for it can ameliorate the effects of some of the automatic cognitive processes that cause immoral actions in the situationist experiments.³² This chapter will argue in particular that not only is the data that concerns Doris not a problem for virtue ethics (as I argued in Chapter 2), it constitutes *prima facie* motivation to consider mindfulness a virtue.

Section 1 lays out the empirical data that describes a problem with the way humans act: in many morally important situations, we act on judgments and beliefs upon which we did not deliberate. This lack of deliberation contributes to many moral errors. This problem pervades the situationist research data and some other important psychological data, particularly work on System 1 and System 2 and implicit bias.

Section 2 then asks: is there a virtue trait that addresses this shortcoming? One classic

³² I understand that some readers might be inclined to protest that the situationist experiments do not capture *necessarily* immoral actions. There is room to argue this claim, but my own sense is that the actions of the experimental subjects in experiments like Milgram's was immoral. I will assume this is the case in this chapter, both because I think this is the right view and also because this puts my work in direct contact with Doris'. My goal here is to show how a virtue ethicist could look at the same data as Doris but draw different conclusions; it is a separate debate whether or not the data is as Doris and I both see it. Of course, if the Milgram experiments do not capture cases of immorality, then Doris' argument using them does not get off the ground.

way of thinking about virtues is that they are corrective of certain temptations or patterns of thought that are characteristically difficult for humans to resist. Acting selfishly or wrongly out of fear, for example, is a universal temptation for humans to overcome, and therefore courage is an important virtue.³³ The automatic thinking described in Section 1 is a universal human trait, and it often leads us to act in morally problematic ways. Given this, it is plausible that a trait which is corrective of this flaw is a virtue. Section 2 argues that the trait of mindfulness is such a corrective by looking at the example of implicit bias in detail. Section 3 discusses further empirical evidence, particularly the work of Ellen Langer, which supports the notion that mindfulness is a virtue because it benefits its possessor, another classic hallmark of virtue. This chapter does not establish that mindfulness is a virtue, though I hope it can serve as the beginning of such an argument. But the main focus is to argue that mindfulness deserves the sort of careful consideration that many other traits have received in the literature on virtue ethics, and to show that this claim is motivated by the same empirical evidence others use to attack virtue ethics.

3.1. Doris Revisited: Automatic and Controlled Thinking

There is a large literature discussing automatic cognitive processes. For example, there is a large literature on dual process cognition, which divides cognition roughly into System 1, consisting of fast, automatic cognitive processes and System

³³ As Philippa Foot said, “it is only because fear and the desires for pleasure often operate as temptations that courage and temperance exist as virtues at all. As things are, we often want to run away not only where that is the right thing to do but also where we should stand firm; and we want pleasure not only where we should seek pleasure but also where we should not. If human nature had been different there would have been no need of a corrective disposition in either place, as fear and pleasure would have been good guides to conduct throughout life” (Foot 1978,170).

2, consisting of slow, deliberative cognitive processes. (see, e.g., Kahneman 2011; Stanovich 2011). System 1 makes snap judgments we might describe as intuitive. System 2 is made up of our slow, reflective, deliberative mental processes. In many cases, System 1 works without the influence of System 2 when we make intuitive judgments and act upon them. In other cases, System 2 intervenes and deliberates regarding our course of action or formation of a belief, allowing us to act and think contrary to our intuitive System 1 judgments. System 1 and System 2 are explanatory tools and are not intended to capture actual distinct systems in the human mind or brain. Doris, in particular, states the contrast as one between automaticity and control. This section is concerned with the way in which System 1 automatic responses can be mediated, the way in which we can intervene in the causal process linking System 1 judgments to actions. Whether this mediation occurs through System 2, as it seems to in Doris, or some other system,³⁴ is not important for these arguments. What matters is that such mediation is possible, and it is possible in situations where it is morally important.

A system of automatic and controlled cognitive processes provides a plausible explanatory framework for much of the situationist research used by Doris to ground his argument against virtue ethics. This framework explains cases where agents are acting in situations with no notice they are in an experiment, for example where they have no apparent conscious awareness of a link between finding a dime and helping

³⁴ Stanovich, for instance, distinguishes more than one type of System 2 thinking: the algorithmic and reflective minds (Stanovich 2011).

an experimental confederate who dropped his papers (See Isen and Levin 1972). It explains a set of experiments on implicit bias, which Doris cites in his discussion of control and automaticity (Doris 2012, 374-375). It also explains experiments where Princeton seminary students on their way to give a sermon on the Good Samaritan were drastically less likely to stop to help a stranger in need if they were in a hurry. It could be the case that the students' prior commitment to give a good sermon dictated the way they perceived and categorized the moral demand of a person in need (See Darley and Batson 1973). While less obvious than the previous cases, this schema can even make sense of the more extreme cases of moral mistake in the situationist literature, particularly the Milgram experiments and Zimbardo's Stanford Prison Experiment (Milgram 1963; Zimbardo et. al. 1973). In both cases, there is good reason to think that many of the experimental subjects found themselves locked, so to speak, into a non-reflective mode of relating to the experimental stimulus that led them to engage in behavior they would never reflectively endorse. In Langer's terminology, discussed in detail in Section 3, they related to their situation mindlessly (Langer 1989). Mindlessness is a state characterized by being entrapped by certain categorizations of the world, behaving automatically without reflection, and acting from a single perspective (Langer 1989, 10). The non-conscious System 1 processes at work are perhaps harder to see in these cases because they involve actions that appear to involve conscious choice: to shock an apparent innocent into silence and to torture an innocent prisoner. But many of the experimental subjects in these cases appear to be trapped within a context of limited autonomy, where they did not

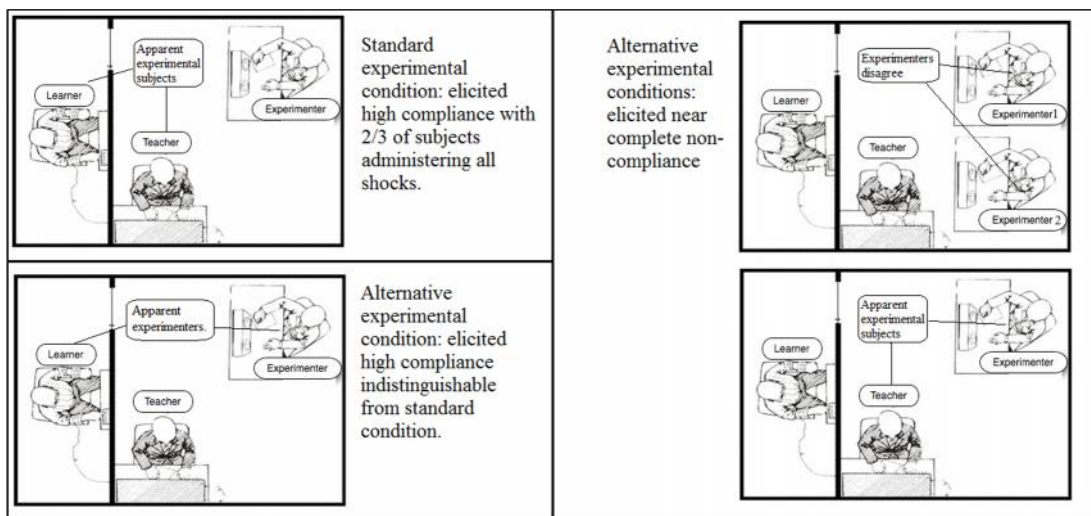
reflectively endorse any option available to them but failed to see that other responses, less morally objectionable responses, were available. And it is plausible that this perception of limited autonomy is related to automatic processes by which we make sense of the world prior to reflecting on how we can act in it. If this is correct, then the situationist literature could be seen as almost entirely a subset of the literature on automatic and controlled cognitive processes.

Doris gives a good defense of this line of argument in relation to the Milgram experiments, in particular (see Merritt et. al. 2012, 383-385). He notes that by looking at the different permutations of the Milgram experiment, we can see what features of the famous original might have generated the disturbing level of compliance. All of the Milgram experiments involved an experimental subject taking on the role of teacher and instructing a learner, who was supposed to memorize a set of names or objects. When the learner made a mistake, the teacher was told by an experimenter to give an increasingly severe shock. Teachers were told that the study was about the relationship between learning and punishment. In fact, the study was about how many shocks the teacher would deliver under the experimenter's orders. Everyone other than the teacher was an experimental confederate, the responses of the learner were pre-recorded, and the language used by the experimenter was tightly scripted. In the basic experiment, Milgram reported that two-thirds of the teachers delivered shocks that they had good cause to believe could be fatal, though the experimenter would continually assure the teacher that the shocks would do no permanent damage.

This basic experiment is quite famous, but most people are not aware of the many permutations Milgram ran to try to isolate the causal factors involved.³⁵ In the initial experiment, the teacher came into the experiment room with another apparent subject who would be the learner. An apparently random process assigned the test subject to the teacher role, but the process was not random, and the true experimental subject was always the teacher. When Milgram changed the process to appear to randomly assign roles of teacher and experimenter to fellow-subjects, and the person running the scenario played the role of the learner who was to be shocked, “every subject respected the experimenter-learner’s demands to stop” (Doris 2012, 383). Subjects did not just stop delivering shocks, “many subjects literally leapt to the aid of the [learner], running into the other room to unstrap him” (Milgram 1974, 103). In another permutation, two apparent experimenters stayed in the room with the teacher and, when the learner protested the shocks, the two experimenters began to disagree about what the teacher should do. Again, all of the teachers in this permutation chose to stop giving shocks (Doris 2012, 384-385). Interestingly however, when the test involved two apparent experimenters who were randomly assigned the roles of experimenter and learner, compliance by teachers was as high as in the basic experiment. The standard and alternative set-ups are depicted in Figure 1 below.

³⁵ Indeed, a recent book has made headlines and radio news programs by asserting that Milgram hid the full results of his studies (Perry 2013). It is an interesting book, and the popular media may have seized on the basic result, but it seems unfair to Milgram to accuse him of hiding results he actually published.

Figure 1



Doris takes these results to imply that the behavior of test subjects in the initial experiment was based on the “relative physical position and social status” of the experimenter telling the test subject what to do (Doris 2012, 385). He claims that these two factors could dictate how the test subject’s “other-oriented attention” is directed. Doris is clear that these factors do not provide a comprehensive explanation, and points to other social and non-social factors that could influence behavior. But the key point for Doris is that “the determinative cognitive processes occur unreflectively and automatically, cued by morally arbitrary situational factors” (Doris 2012, 387). None of the explanatory factors suggested provide a good reason, in the sense of a morally defensible reason, for action that would justify the test subjects’ choice to shock the apparent learner.

Indeed, Doris suggests that this dynamic provides a general explanation of the situationist data. He says, “we present empirical support for the claim that many important cognitive and motivational processes proceed without intentional direction,

and result in cognitive and behavioral outcomes inconsistent with the agent's (often reflectively affirmed) evaluative commitments" (Doris 2012, 371). That is, Doris argues that our failure to mediate System 1 responses in morally significant situations explains many of the results of the situationist experiments (Doris 2012, 372).

Unmediated automatic processes lead us into moral errors in certain situations, as in the Milgram experiment, and do so because of certain features of our cognition, such as with implicit racial bias.

Doris describes this process as one of incongruency between our normative commitments and our automatic processes. He says,

"behavior is influenced by cognitive processes that, if they were accessible to reflection, the actor would not endorse as acceptable reasons for action. This *incongruency*, as we shall call it, involves a relation between (1) automatic processes likely to influence a subject's behavior on normatively significant occasions of action, and (2) normative commitments of the subject, such that (3) if the subject knew about (1), he or she would reject (1) in light of (2)" (Doris 2012, 375).

Doris then argues that "to the extent that incongruency is pervasive, it renders the virtue-ethical model of practical rationality problematic." (Doris 2012, 375). This reiterates Doris' main argument against virtue ethics, covered in detail in Chapter 2, but recasts it here as an issue of incongruency. Doris describes how he sees this data as problematic for virtue ethics:

"The normative ideal of virtue ethics places a heavy executive burden on agents' reflective supervision of their behavior. Too heavy, we suggest, to be effectively borne by human cognitive equipment on many morally important occasions for action. The virtue-ethical model emphasizes selecting the right reason on which to act, such as 'She needs help,' or 'It's hurting him, this has got to stop'... This requires being able 'to see, on each occasion, which course of action is best supported by reasons'... But the trouble with moral dissociation, on the explanation we have proposed, is that it occurs largely

through cognitive processes that bypass the type of reflective supervision that is supposed to characterize acting for reasons” (Doris 2012, 387-388).

Virtue ethics cannot account for this behavior because virtue ethics cannot account for action that bypasses reflective deliberation and supervision, or so Doris says.

The main purpose of the remainder of this chapter is to challenge these implications Doris draws from the situationist data and argue that virtue ethics can accommodate these empirical results. Two brief criticisms at this point will set the stage for the rest of this chapter. First, Doris has not established that the cognitive processes at issue necessarily bypass reflective supervision; one of the most interesting aspects of the work on mindfulness described below is the ways in which we can bring these cognitive processes into reflective consideration. In simpler contexts, such as in Kahneman’s ‘bat and ball’ example,³⁶ we can learn to answer such questions correctly despite the automatic cognitive process at work. In particular, we can learn to answer these questions correctly while still feeling the force of System 1 intuitions. The System 1 intuitions do not necessarily “bypass” reflective supervision, though they sometimes may do so. There is no reason to think that a similar dynamic cannot happen in cases of moral incongruency. Second, even if such cognitive processes always bypassed contemporaneous reflective supervision, this does not imply that those processes are incorrigible. Virtue ethicists since

³⁶ In this experiment, Kahneman would give people the following scenario: “A bat and ball cost \$1.10. The bat costs one dollar more than the ball. How much does the ball cost?” Most people have the number \$0.10 jump into their mind as the answer when they read this problem, even though the correct answer is \$0.05. Also interesting: even when you know the problem, many can still feel the temptation of answering “\$0.10”. I, for example, still feel my System 1 telling me the answer is \$0.10, even though I know this is incorrect (Kahneman 2011, 44).

Aristotle have considered actions which spring directly from a good disposition, and not conscious deliberation, to be a hallmark of virtue. The most virtuous among us spring directly to action in a situation that demands it. There is a space in virtue theory, therefore, for the idea that we must train our automatic processes to respond rightly to the circumstances of our lives.

3.2 Mindfulness as a Solution: Implicit Bias

While Doris discusses many automatic processes which could cause moral incongruency, in order to make my discussion manageable, I will focus here on implicit bias and argue that the disposition of mindfulness could serve as a corrective to the morally troubling activity of implicit bias. I chose implicit bias because Doris considers it important, and also because it is the topic of an interesting and useful body of psychological research. It is also a good example because it is an automatic process that we have good reason to believe leads to morally problematic actions. Similar arguments to those presented in this section could be used in support of the claim that mindfulness is a corrective to moral incongruency caused by other sorts of automatic cognitive processes.

Implicit bias is bias that is not a part of one's conscious experience. One can be biased implicitly against African Americans, for example, while being consciously committed to racial justice and equality. A wide body of research shows the existence of such implicit bias. One of the most common tests used to show implicit bias is the Implicit Association Test, or IAT.³⁷ This test asks the subject to sort two kinds of

³⁷ An online version of the IAT is available for anyone to try at <http://implicit.harvard.edu>.

things, as quickly as possible, into two columns. The theory behind the test is that if the test takers ask the subject to sort things that are associated in their minds, even implicitly, they will sort more quickly than if those things are not associated. To use the race example, experimenters might use a set of words that evoke goodness (things like pleasure, happy, smile, love) or badness (hate, cruel, angry, mad) and mix that with a set of pictures of faces of black Americans and white Americans. Then they would ask the subjects to classify the words and pictures according to different sets of rules. First, they ask subjects to sort white faces and goodness terms to the left, and black faces and badness terms to the right. Then they ask subjects to sort white faces and badness terms to the left, and black faces and goodness terms to the right. The IAT measures the time it takes to sort the faces and terms accurately and reports this result. The claim is that the more terms are associated in our minds, the faster we will be at sorting them into the same column. The less they are associated in our minds, the more our consciousness has to work to put them in the correct column and the slower we will sort them accurately. One who is implicitly biased to associate African Americans with badness, for example, would sort more quickly if the black faces were to be sorted with badness terms, and they would sort less quickly if the black faces were to be sorted with goodness terms.

This research is fascinating in its own right, but the degree to which our brains associate faces of different colors with different normative categories is not obviously salient to our moral character. Implicit bias exists independently of overt bias, and it is plausible that a majority of harmful bias in the world is overt bias. Luckily, a

number of social scientists have begun looking at the relationship between implicit bias and morally significant decision-making that might be discriminatory on the basis of sex, race, or other categories. The results so far give good cause to think that implicit bias is morally problematic.

In one study, 127 science professors reviewed application materials that were identical but for the gender of the applicant (Moss-Racusin et. al. 2012). Half of the professors thought the applicant male, half thought the applicant female. As might be expected in a generally sexist society, the professors rated the male applicant as more competent, more hireable, offered to mentor them more than the female applicant, and recommended a higher starting salary. This study also asked participants to take the Modern Sexism Scale, a scale that “measures unintentional negativity toward women, as contrasted with a more blatant form of conscious hostility toward women” (Moss-Racusin et. al., 16476). The measure of faculty on this scale showed a statistically significant correlation with their judgments of competency ($p < 0.01$), hireability ($P < 0.05$), and mentorship ($P < 0.001$). That is, the more implicit bias the faculty exhibited, the lower they ranked the woman applicant, to a strongly statistically significant degree. This indicates that aside from the likely overt sexism among study participants, their implicit gender bias influences hiring decisions.

The work on race is even stronger. In particular, a number of studies have been done on the relationship between the implicit racial bias of physicians and their recommended course of treatment for patients of different ethnicities (Green et. al.

2007; Sabin 2012).³⁸ In one study, participant physicians were measured for implicit bias using the IAT (Green et. al. 2007). They were also measured for explicit racial bias. Then, they were asked to render a treatment recommendation for a pair of patients who were identical except for their race; one was racially black, one white.³⁹ Interestingly, explicit bias was not correlated with treatment decisions, but implicit bias did have significant effect on treatment decisions (Green et. al., 1235). For two out of the three measures of implicit bias,⁴⁰ the more implicitly biased the physician was against black people, the more likely they were to recommend a particular treatment to white patients and less likely to recommend it to black patients (Green et.

³⁸ This study has come under some criticism (Dawson 2009). I am no medical expert, but to an educated layperson the criticisms seem weak, and the original authors stand by their work (Green 2009). Part of the criticism focuses on the fact that the doctors in the original study were not asked to give a differential diagnosis for the patient, and therefore the original study did not capture the fact that there were many possible causes for the presented symptoms, and the tested treatment (thrombolysis) would be beneficial in some, harmful in others. Green points out that this is beside the point: “At least five studies (not conducted by any of us) have demonstrated that African-American patients are less likely to receive thrombolysis under this circumstances [sic]. Our concern was not with the role of race in the recommendation of thrombolysis. Rather, we were interested in the degree to which racial bias as measured by a test of implicit race attitude (the IAT) would predict physicians’ recommendation of thrombolysis to black versus white patients” (Green 2009, 141). The fact that African Americans do not get this treatment as much as Caucasian Americans is already established; Green et. al.’s study was focused on looking at whether implicit bias might explain this difference, not to justify the difference in the first place.

³⁹ The faces were not African American and Caucasian. The authors took a set of photographs with neutral facial expressions of various races and morphed them, using software, into a set of new patient images. The best 21 quality images were chosen and then 19 independent evaluators chose the two white and two black faces most closely matched in apparent age and attractiveness (Green et. al., 1233).

⁴⁰ The study measured implicit bias along three spectra: good-bad, general cooperativeness and uncooperativeness, and “medical” cooperativeness and uncooperativeness. The good-bad spectrum is the classic race-based IAT test. It asks test-takers to sort faces of different ethnicities into columns along with normatively good or bad words. The second and third IATs were developed for this study. The second changes the good or bad terms into ones representing cooperativeness and uncooperativeness, and the third changed the terms to more specifically represent medical cooperativeness and uncooperativeness (Green et. al., 1233).

al., 1236). Implicit bias was correlated strongly with racially disparate treatment decisions.

While this work linking implicit bias with morally significant discrimination is relatively new, one of the most important parts of Green's study was the way in which it links implicit bias with well-known discriminatory practices. That is, Green et. al. aims to explain the already established fact that black Americans receive thrombolysis treatment in a discriminatory manner, as shown by many other studies (Green 2009, 141). Importantly, the study showed no link between physicians' explicit bias and treatment decisions (Green et. al. 2007, 1235). It is plausible, therefore, that many known discriminatory practices and effects are caused, or partly caused, by implicit bias. For example, recent MacArthur Foundation grant winner Jennifer Eberhardt's work is a vivid illustration of the potential moral significance of implicit bias. Eberhardt has done a wide array of work examining the ways in which racial and gender bias could influence morally significant aspects of modern life. For example, she showed that a hypothetical criminal defendant was more likely to be sentenced to death the more "stereotypically black" that defendant appeared (Eberhardt et. al. 2006). In this study, Eberhardt and colleagues chose 44 black defendants who were death-eligible and advanced to the penalty phase of their trials. They obtained photos of these 44 defendants and presented them to a set of "naive raters", people who did not know who the pictures were of or what the purpose of the study was, and asked those raters to rate the stereotypical blackness of each picture. After controlling for a number of variables covering the circumstances of the crimes

committed and features of the defendant and victim, Eberhardt found a substantial, statistically significant correlation between stereotypical blackness and death penalty sentencing. Black defendants rated in the lower half of the stereotypicality index were only 24.4% likely to receive the death penalty whereas those rated in the upper half of that index were 57.5% likely to receive a sentence of death (Eberhardt et. al. 2006, 384).

3.3 Mindfulness as a Corrective to Implicit Bias

The evidence is very suggestive that implicit bias contributes to seriously flawed, morally significant decisions.⁴¹ In particular, implicit bias challenges our capacity to embody the virtue of justice, because it can influence our decisions about what medical treatment, or criminal punishment, someone is owed even when we have overt commitments to make unbiased decisions. This is a troubling case of moral incongruency. I take it as a given that we have a clear duty to struggle against any morally problematic implicit bias we know, or have good reason to think, we might be operating under. For example, if we know that we are implicitly biased against African American men, in that we implicitly associate them with criminality or animality, as Eberhardt's research suggests, we have a moral duty not to act on that bias. I will not defend this view here, both because it is beyond the scope of this

⁴¹ More research surely needs to be done on the link between implicit bias and biased actions. But the existence and prevalence of implicit bias has robust empirical support, as does the existence of biased decision-making. As I argued in the introduction, it is important for psychologists to consider the prior probability of their hypotheses before inferring causation. This is a good case where it seems highly plausible even before empirical data that implicit bias might manifest in morally serious ways.

dissertation, but also because I think most accept that such a defense is unnecessary. Acting in a biased manner on the basis of race, or gender, is immoral.

Given this, we must consider what it would take to correct for this implicit bias, if such correction is possible. There are at least two ways we might correct for it: eradicate the bias or find a process which allows us to interrupt the chain of causation between implicitly biased judgments and actions based on those judgments. It seems likely that these are not two distinct projects but related, for I suspect that interrupting the chain of cause between implicit bias and action may weaken implicit bias itself. Regardless, changing our implicit bias seems a noble, worthy goal. We might, for instance, stop consuming media which portrays African Americans in a criminal, animal light. We might seek to stop others from consuming this media, and lead a boycott of it. But given the ubiquity of implicit biases, and the way our automatic cognitive system functions, it is implausible to think that we could eradicate our implicit biases. Unsurprisingly, Doris agrees with this view (Merritt 2012, 388-389). Therefore, regardless of the degree to which we can change our levels of implicit bias, a disposition to interrupt the chain of causation between biased judgments and actions appears necessary for us to act rightly. This disposition to interrupt the causal chain is a constituent part of the virtue of mindfulness.

In the parlance of System 1 and 2 thinking, part of mindfulness is the intervention of System 2 between a System 1 output and action.⁴² System 2 overrides

⁴² In Stanovich's terms, "One of the most critical functions of Type 2 processing is to override Type 1 processing" (Stanovich 2011, 20). Overriding involves both "interrupting Type 1 processing and suppressing its response tendencies" and supplying a better response in the place of the System 1

System 1, sometimes merely preventing action and other times providing an alternate course of action. In Ellen Langer’s terminology, mindfulness is a contrast state⁴³ with mindless thinking (Langer 1989; Langer 2009). Mindless thinking is defined by automatic, unreflective responses to situations. Mindlessness is also characterized by the unreflective acceptance of given categorizations for things. For example, one of Langer’s famous experiments tested whether people given an object labeled as a “dog’s chew toy” would recognize it could also be categorized, and used, as an eraser. Mindfulness is not necessarily limited to System 2’s control over the action-guiding power of System 1, for recategorization could happen entirely within System 2, but control over System 1 is a part of mindfulness. Since implicit bias is a System 1 phenomenon, this control over System 1 is relevant for fighting the influence of implicit bias on our actions, and so this control will be the aspect of mindfulness considered in this section.

This argument is analogous to an argument regarding rational belief. We might think that one’s beliefs ought not to form any contradictions,⁴⁴ but we all have so many beliefs that it is practically guaranteed that we hold, and will always hold, contradictory beliefs. I will not argue this here, but imagine the Sisyphean task it would be to pro-actively police our beliefs for consistency. A more practical standard

response (Stanovich 2011, 21). These two functions are split in Stanovich’s terminology into the “reflective” and “algorithmic” minds, respectively.

⁴³ I discuss mindfulness as a “disposition” and refer to a mindful “state” in this chapter. I use mindful “state” to refer to the state of being mindful in a particular moment or situation. I use mindful “disposition” to refer to potential virtue, a state that endures over time and across situations.

⁴⁴ For a summary of defenses of the Principle of Non-Contradiction, and objections to it, see the Stanford Encyclopedia of Philosophy’s entry on dialetheism (Priest 2013).

for rationality might be to strive to notice and resolve, or at least refrain from acting on, such contradictions when they appear in conscious thought. To do this, we must have a disposition to look for and respond to contradictions that arise in our thinking. This disposition works in at least two ways: we try to notice and resolve contradictions when they arise, and we also try to notice when we are motivated by contradictory beliefs and refrain from acting on them until they are resolved.⁴⁵ The disposition is to place a halt, so to speak, on the connection between contradictory beliefs and action, at least until the contradiction can be resolved. Analogously, eradicating implicit bias completely seems likely to be a Sisyphean task. An analogously dual disposition might be a solution for implicit bias: a disposition to notice the effects of implicit bias as they arise and to try to correct for them, and a second disposition to interrupt the chain of causation between an implicitly biased judgment and action. The disposition of mindfulness includes this second disposition, to enable us to monitor and control the effect of contradictory beliefs and implicitly biased judgments on our actions. This aspect of mindfulness is a good candidate for a virtue, therefore, because it serves as a corrective to implicit bias, a characteristic challenge humans face to acting rightly.⁴⁶

⁴⁵ This description corresponds generally to the algorithmic and reflective minds in Stanovich's tripartite system. (Stanovich 2011).

⁴⁶ The example of rational belief is not randomly chosen. In fact, I think that the virtue of mindfulness includes the disposition to pay attention to our beliefs as they come into our minds and influence our actions, and to intervene in the causal chain between our beliefs and our actions where we have contradictory, or incoherent, or otherwise flawed beliefs. This point would require an argument that contradictory beliefs are problematic for moral action, however, which is outside the scope of this section. This section tries to show that mindfulness is a palliative for at least one characteristic problem humans must confront, implicit bias. It could also be, and I suspect is, a solution to other characteristics challenges humans face, such as dealing with conflicting or contradictory beliefs.

I am not aware of any contemporary virtue ethicist taking mindfulness as an important trait in relation to this sort of automaticity, but this move has been recently suggested by psychologists, and it may be in line with ancient Stoic thought.⁴⁷ In the words of a recent paper, “We suggest that though distinct, all these mechanisms of mindfulness have a universal effect of deautomatization, a process in which one’s previously established tendency to effortlessly and unconsciously engage in maladaptive behaviors becomes conscious and controlled” (Kang et. al. 2014, 168). The work of psychologists helps cast a light on the sorts of mechanisms the disposition of mindfulness might involve that allow us to interrupt the causal chain between implicit judgments and action. In this paper, the authors linked four aspects of mindfulness to deautomatization. These aspects are:

1. Awareness: “having conscious knowledge of one’s internal and external experiences, including bodily sensations, thoughts and emotions, and external events such as sights and sounds.
2. Sustained attention: “involves placing one’s attention on the ongoing stream of internal and external stimuli.”
3. Focus on the present moment: “involves directing one’s attention, with or without effort, to the internal and external phenomena occurring at each moment of awareness.”
4. Nonjudgmental acceptance: “involves experiencing thoughts, sensations, and events as they are at the moment they enter one’s consciousness, without judging them as being good or bad, desirable or undesirable, important or trivial” (Kang et. al. 2014, 170-171).

⁴⁷ Thanks to John Bowin for pointing out to me the many ways in which the discussion of mindfulness in this chapter shares a kinship with Stoic thought. I hope to pursue the similarities and differences between mindfulness and ancient Stoic thought in future work.

These qualities capture some intuitive qualities of mindfulness, including what has been considered mindfulness in Eastern religious traditions, and the core of what has come to be seen as effective therapeutic practices in Western psychology, and they have been the subject of empirical study by social psychologists. While they likely do not capture everything important about mindfulness as a potential virtue, they illustrate how this disposition might ameliorate the morally significant consequences of implicit bias. For purposes of the argument in this section, therefore, mindfulness is a disposition that includes at least these four qualities.

How do these factors help us deal with the specific challenge of implicit bias? Implicit bias exists despite, and generally in conflict with, overt, deliberate bias. Bias in the automatic cognitive system of a person who is explicitly biased is not so much implicit bias as it is implicit agreement with overt thoughts. Overtly biased people need to be convinced not to be biased; the problem is not just, or not centrally, their automatic cognitive bias. The important cases for virtue, therefore, are the ones where someone is: (1) implicitly biased in (2) a morally serious area, and (3) they reject such bias explicitly. In the above cases, these four factors could be immensely helpful, perhaps even necessary, to fight actions that stem from implicit bias. These factors are all ways of attending to one's own thoughts and judgments so as to bring the contents of one's own internal experience, including thoughts and emotions, into conscious awareness. This awareness of one's own thoughts and emotions as they arise in the mind is necessary to interrupting the causal chain between those thoughts

and emotions and action; therefore, it seems that we must have these qualities as a disposition in order to combat implicit bias.

One might wonder in particular about quality 4, which at first seems at odds with the notion that the mindful person intervenes between an implicitly biased judgment and action. Quality 4 is a disposition of “nonjudgmental acceptance” of our thoughts; but, I have just argued that mindfulness is important because we can be judgmental about, and refuse to act upon, thoughts that arise from implicit bias. First, it is important to see that the concept of mindfulness does not dictate how we act in light of implicitly biased judgments, only that we interrupt the chain of causation. In any given case, what we ought to do is governed by other virtues and interests; mindfulness only helps us act on those other right reasons despite implicit bias. As defined here, mindfulness is essentially a structural virtue like courage, to use Robert Adams’ terminology (Adams 2006). Structural virtues are virtues of “the way one organizes and manages whatever motives one has... of ability and willingness to govern one’s behavior in accordance with values, commitments, and ends one is for” (Adams 2006, 33-34). Given what we know about automatic cognitive processes, and our reactions to them, it is psychologically realistic that nonjudgmental acceptance of implicitly biased judgments is an important step on the way to discriminating control over the causal link between those judgments and action. In order to decide when and how to intervene, we have to have an honest picture of the implicit judgments we are making, or at least as close to honest as we can get. But a common initial reaction to the concept of implicit bias, particularly after taking the IAT and seeing that one has

implicit bias, is revulsion and denial that one could have, or be motivated by, such prejudice. It is tempting to introspect on one's own thoughts and self-servingly find no bias, despite the result of the IAT. This temptation means that it is likely important for us to cultivate this disposition of nonjudgmental awareness if we are to get anywhere close to an honest picture of our own implicit bias.⁴⁸

The case here is analogous to the other structural virtues, such as courage. Fear is a universal human experience. Courage as a virtue involves always making a choice in the right way, for the right reasons, despite fear. It is the disposition to reliably act in accord with practical reason regardless of how much fear one experiences. In order to do this, we must gain some form of psychological distance from fear, in the sense that we are always ready to act despite fear in the cases where practical reason demands this. It is plausible that this demands the sort of non-judgmental awareness of fear that mindfulness describes, for wherever we are unknowingly experiencing fear, we run the risk of acting with cowardice. This is a plausible psychological explanation, for example, of many cases of homophobia. There is psychological evidence that those who have homophobic views are more likely to be aroused by homosexual acts (Adams et. al. 1996). The psychology of the relationship between internal orientation and external expression of homophobia is complex (see, e.g., Weinstein et. al. 2012), but in the framework of virtue ethics, the

⁴⁸ The degree to which we have control of and responsibility for our automatic cognitive systems is an underlying issue in this discussion. I think that this is a very important issue for virtue ethics, and it is an issue that deserves its own treatment, at great length. In this section, I have tried to toe a middle line on this issue: we surely have some responsibility for the contents of our automatic cognitive systems, and we surely do not have complete or full responsibility for them.

basic explanation is one of fear: agents are motivated by a fear of their own internal desires, of being discovered, of having a bad character. This fear motivates them to acts of great injustice, often without conscious awareness that they are doing these acts out of fear. Nonjudgmental awareness of our fears, just as of our implicit biases, is important for virtue.⁴⁹

Neither mindfulness nor courage requires that we never act without reflection from implicit bias or fear. Sometimes fear is a reliable motive to good action, and practical reason need not intervene. One of the things I am most afraid of is that my small daughter will be harmed. This fear motivates me to be intensely attentive to her environment and therefore able to notice, and respond to, dangers that I might not otherwise notice. Of course, this fear could also lead me to make terrible decisions and limit my daughter's autonomy in unjustifiable ways. Courage demands that I not let this fear dominate my decision-making, but it does not demand that I eradicate this fear or that I ignore it. Likewise with implicit bias and mindfulness. Implicit bias is a sort of automatic cognitive association, and automatic cognitive associations are not all bad motives for action. For example, take the case of chicken sexer discussed in philosophy. This person is excellent at determining the sex of chicks, but he cannot teach anyone else how to sex because he does not know why he thinks any particular

⁴⁹ The argument in this paragraph serves as another one in favor of mindfulness as a virtue, though I cannot develop it in full here. Briefly, Julia Annas has suggested the unity of the virtues thesis as a plausible test for whether a particular disposition is a virtue. If we can plausibly describe some disposition as required in order to fully practice a cardinal virtue, like courage, then that disposition is a good candidate for a virtue itself. The argument in this paragraph is essentially that mindfulness, in the form of nonjudgmental awareness of fear, is a necessary part of the psychology of courage, because courage demands that we relate to fear as an input to practical reasoning; it requires that we not identify with the fear personally, that we have some psychological distance from it.

chick is of either sex. In psychological terms, his expertise appears to be completely contained within his automatic cognitive processes.⁵⁰ In this case, it is not a problem for his trait of mindfulness that he acts on these automatic judgments.

There is at least one important potential conceptual objection to considering this trait of mindfulness a virtue as an Aristotelian. Aristotle famously argued that virtue involves an alignment between an agent's orectic, or appetitive, and deliberative faculties. The continent person is the one who brings their desires into line with their reasoned judgments; the virtuous person's desires are in harmony with their judgments. On this view, one might think that mindfulness is no virtue, that it is at best a trait of the continent person striving for virtue, because it shows a flaw in our orectic faculties. The virtuous person's automatic system would not be implicitly biased, or flawed in any other way, so mindfulness is not a virtue. In other words, the virtuous person is one who has really accomplished the Sisyphean task of eradicating implicit bias.

This is a tempting line of argument, but it is not necessary to read Aristotle this way. In particular, something like Sarah Broadie's account of this issue is both a plausible reading of Aristotle, and also accommodates the arguments and evidence presented here (Broadie 1991, 63-67). The relationship between desire and reason in a virtuous agent is complex. Broadie argues that the virtuous person is not someone in which desire and reason happen to coincide perfectly, with two sufficient causes for

⁵⁰ This example is apparently famous and widely used in certain debates. The example of the chicken sexers is a real phenomenon (Pylyshyn 1999, 31; Biederman & Shiffrar 1987).

right action in every situation. Such a picture is fanciful, because it makes it seem as if the virtuous person's faculty of desire is a homunculus of that person's virtuous reason. But this objection assumes this picture because it characterizes the virtuous person as having two sufficient causes for right action. What Broadie argues instead is that desire relates to reason as an ideal child to a parent; it has a mind of its own, so to speak, but is perfectly obedient to the parent's wishes. On this view, an agent's virtue is not in jeopardy because she has desires that conflict with virtue, as long as the desiring part of the agent immediately, obediently follows the dictates of practical wisdom. As Broadie says, "in the virtuous person such feelings appear only as feelings and not as incipient actions" (Broadie 1991, 66).

It might be the case, for instance, that a virtuous agent has a powerful overriding desire to do what is right, and so the judgment of practical wisdom coupled with that desire is sufficient to "silence" all contrary desires and reasons for action, to use McDowell's terminology (McDowell 1995, 335). The notion of silencing captures something about the moral psychology of the virtuous. Virtue does not consist in balancing competing reasons in each case and thereby coming to the best overall solution, as with a utilitarian balancing test. Rather, the virtuous person's correct reason for action "is apprehended, not as outweighing or overriding any reasons for acting in other ways which would otherwise be constituted by other aspect of the situation (the present danger, say), but as silencing them" (McDowell 1995, 335). In this way, the virtuous person does not have to actively overcome a contrary desire to do the virtuous thing, because that desire is silenced in light of the right

desire, but they also do not need a virtuous homunculus in their faculty of desire, independently generating the same answer as their practical wisdom at all times.

This position makes sense of the virtue of mindfulness as the disposition that defines how practical wisdom relates to the outputs of our automatic cognitive processing. The mindful person is one who polices the boundary between automatic judgments and practical wisdom and always goes with practical wisdom where they conflict in conscious thought. They still make automatic judgments, and they still take them into account in decision-making in the very many morally appropriate contexts for utilizing automatic decision-making. Indeed, virtue demands that the automatic decision-making system be in charge in certain situations of imminent danger. If my daughter and I were enjoying the view from the ocean cliffs near our house and she began to lose her balance, it would be an act of virtue for me to grab her in a brief, fearful panic and save her from falling. A mindful person might let some outputs of their automatic cognitive system translate into action without interference. The mindful person is not in control of their automatic cognitive system, but they are in control of the chain of causation leading from their automatic cognitive system to action.

Mindfulness is a disposition that could be fundamental to addressing the moral problems of implicit bias. But Doris has something to say about the possibility of a response from virtue ethics to automaticity that deserves consideration:

“By way of remediation, then, the traditional virtue-ethical approach can prescribe, in effect, an agenda of deliberate self-improvement: e.g. ‘Note to self: pay better attention to how the other person is doing’ or, ‘If I encounter someone in distress and other bystanders are around, do not look to the other

to figure out what to do'. But given the ubiquity and power of behavior-influencing cognitive processes that may resist reflective supervision, and the limitations on the cognitive resources required to implement such supervision, there is good reason to doubt whether a program of deliberate self-improvement will alone suffice to alleviate tendencies toward moral dissociation" (Doris 2012, 388-389).

But the virtue ethicist is not limited to advocating a program of explicit, conscious and deliberate rule following to combat these automatic cognitive processes, as Doris suggests. Rather, as argued in this section, cultivating a disposition to respond to these automatic thoughts by creating space between the automatic judgments and action, as mindfulness does, is both prior to and more important than the specific, contentful rules one must follow instead of automatic judgments. The structure is, again, analogous to courage: fear is a behavior-influencing cognitive process that resists reflective supervision. It is a common human experience, for example, to remain afraid of something that on reflection we believe is not worthy of fear, and the courageous person acts rightly regardless of this fear reaction. But courage does not independently generate practical prescriptions; it opens up a space between the feeling of fear and action for the other virtues, the ones which dictate right action in the fearful situation, to tell us what to do. Mindfulness is similarly the disposition to open up a space between the output of automatic cognitive systems and action, in order that the mindful person can intervene before acting on morally problematic automatic judgments.

Doris might still protest that this notion of mindfulness is too demanding on our cognitive equipment and my argument is not a realistic response to the problems of automatic processing. This topic will be taken up in detail below, when discussing

the issue of whether and how mindfulness benefits its possessor. But it is worth briefly noting how some research on implicit bias has shown promising results in cases where test subjects are on notice that a particular implicit bias could be influencing their decisions. In the study on implicit racial bias in treatment recommendations, for example, the study authors adopted an interesting methodological tactic: rather than discarding the data from participants who reported a suspicion that the study was about racial bias, they kept that data separate and compared it with the main study sample. They found that this sub-population exhibited an inverse effect from the main expected effect:

“Also, those physicians who were aware that the study had to do with racial bias, and who had higher levels of implicit prowhite bias, were more likely to recommend thrombolysis to black patients than physicians with low bias—the opposite of the study’s main effect. This suggests that implicit bias can be recognized and modulated to counteract its effect on treatment decisions. These finding [sic] support the IAT’s value as an educational tool.” (Green et. al. 2007, 1237).⁵¹

The study’s authors suggest, and the data seems amenable to this claim, that awareness of the possibility of implicit bias can influence the effect that bias exercises on the decisions physicians make. Awareness did not reduce the implicit

⁵¹ The study conclusion may seem hasty and unconvincing here. In particular, if the participants were aware that racial bias was at play in their decision-making, they were likely motivated to act in accord with their explicit non-bias. There does not seem to be any requirement that they noticed or responded to their own *implicit* bias. But it is critical to remember that the treatment recommendations *were* correlated with the subjects’ implicit bias, but the correlation was the inverse of the general study finding. That is, when the subjects thought that racial bias was relevant to their decision-making, the more implicitly biased they were against a particular class of people, here African Americans, the more favorable were their recommendations for treatment for members of that class. There are multiple plausible explanations for this effect, but the fact that it varied with the participants’ IAT scores indicates that it is reasonable to infer that implicit bias was involved in this decision-making in some way. Since their decisions went in the opposite direction of their bias, it is suggestive of some possibility for correcting for the morally problematic effects of implicit bias.

bias itself, only the decisions that were otherwise affected by the bias. That is, when primed to be mindful that their automatic judgments were biased, subjects were able to intervene in the causal chain between implicitly biased judgments and action.

Finally, there is a potential objection to this line of argument from the point of view of skepticism about the veracity of introspection. How accessible to introspection are any cognitive processes, let alone implicit biases? And if these processes are not accessible to introspection, it is hard to see how we could interrupt the causal chain between these processes and action (see Pronin et. al. 2002; Pronin et. al. 2004; Pronin 2009; Schwitzgebel 2008). I have a number of responses to this line of argument and research, none of which I can develop fully here. But I will try to lay some preliminary responses out, and briefly show how this line of argument is actually friendly to my view, not hostile to it.

I will begin with three comments about the psychological research on which this argument is based. First, the psychological research on which some of this objection is based, like all such psychological research, tests for statistically significant differences in the means between control and experimental groups of study participants. Mindfulness, if it is a virtue, is likely a difficult trait to instantiate, and so there is no more reason to expect widespread mindfulness than there was to expect widespread honesty in Hartshorne and May's studies. Pronin's many interesting studies, therefore, do not show anything about whether introspecting bias is possible, only that such introspection is not a widely shared trait.

Second, Pronin's claims that introspection is an illusion or impossible seem to be largely based on the asymmetry between the way we assess ourselves vs. the way we assess others. This is apparent in the four factors of the "introspection illusion":

- "1. *Introspective weighting* (heavy weighting of introspections when assessing self)
2. *Self-other asymmetry* (absence of above when assessing others)
3. *Behavioral disregard* (disregard of behavior when assessing self but not others)
4. *Differential valuation* (asymmetric valuation of own versus others' introspections)." (Pronin 2009, 5)

This asymmetry is interesting, but makes the whole line of argument weaker than Pronin seems to realize when she writes as if introspection is an illusion. All she has established is that we generally give more credence to introspection in judging ourselves than we do to judging others. In one paper, she describes it along philosophically familiar lines: we all believe in our own free will but think that others are making decisions because of situational, deterministic forces (Pronin 2009, 41-44). This asymmetry seems to be more about the perception of introspection rather than the possibilities for introspection.

Third, there seems to be a plausible explanation for at least some of Pronin's results that does not depend on whether the cognitions in question are subject to introspection. In one study, the traits claimed to be unavailable to introspective access were also those judged to be socially undesirable (Pronin 2002). When Pronin set up a more complex set of traits that varied across spectra of introspective access and social desirability, she found that low introspective access and low social desirability

were both correlated with low self-ratings of bias (Pronin 2002, 373-374). She was unable to disaggregate the effects of these two spectra. In her own words,

“we failed in this disentangling attempt, largely because none of the biases assessed by our raters as highly available also were assessed to be highly negative. We do not think that this ‘confounding’ reflected a failure in our methodology. Rather, we think that awareness that one is susceptible to a given shortcoming makes one try to eliminate it and, after failing to do so, to assume it is both ubiquitous and forgivable. As a result, we conclude that cognitive and motivational factors reinforce each other in producing the illusion that one is less susceptible to bias than one’s peers.” (Pronin 2002, 379)

This picture of the relationship between availability and social desirability is actually consistent with the account of mindfulness developed here. In particular, it suggests that awareness of a shortcoming is an impetus for a changing relationship to that shortcoming. While awareness may lead us to think that a shortcoming “is both ubiquitous and forgivable”, awareness of a shortcoming is also consistent with striving to overcome that shortcoming. Indeed, among the qualities tested are a number of different forgivable qualities that one might strive to overcome: procrastination, susceptibility to the Gambler’s Fallacy, and fear of public speaking, to name some examples. Seeing these shortcomings as forgivable and human is consistent with striving to overcome them. If mindfulness can function to make us aware of and attentive to the workings of implicit bias, and other automatic cognitive processes, it might lead us to appreciate how widespread and forgivable they are. And we might also strive to fight their influence, and act rightly despite them.

Eric Schwitzgebel puts these sorts of concerns into a philosophical context (Schwitzgebel 2008). His paper is fascinating, and deserves a fuller response than I

can give here, except to briefly sketch a convergence between his skeptical view and the virtue of mindfulness. Schwitzgebel is generally skeptical that introspection, of the sort that might be required for mindfulness, is a particularly reliable process. The empirical basis for this suspicion is, as far as I can tell, based on empirical data like that discussed above. But Schwitzgebel's paper is also taken up with a detailed consideration of the phenomenology of introspection. To consider the phenomenology of mindfulness, however, takes us down a very different path than the empirical research that has been the focus of this section falls out of relevance. In fact, if we look at the phenomenology of mindfulness in historical traditions, it turns out to be very close to Schwitzgebel's phenomenology of introspection. In particular, Schwitzgebel talks about the evanescence, the un-graspability, of conscious experience and the self. He says:

“The teetering stacks of paper around me, I'm quite sure of. My *visual experience* as I look at these papers, my *emotional experience* as I contemplate the mess, my *cognitive phenomenology* as I drift in thought, starting at them—at these, I'm much less certain. My experiences flee and scatter as I reflect. I feel unpracticed, poorly equipped with the tools, categories, and skills that might help me dissect them. They are gelatinous, disjointed, swift, shy, changeable. They are at once familiar and alien.”
(Schwitzgebel 2008, 267)

This sort of report is precisely in line with language about *anatman* (literally “no-self” in Sanskrit) and mindfulness in the Buddhist traditions.⁵² As an example, the changeability Schwitzgebel talks about is echoed in many passages of the *Dhammapada* on the mind, such as “So difficult to perceive, exceedingly subtle,

⁵² For further similarity to the language of the Buddhist texts, see footnote 60 below.

alighting, as it does, wherever it desires” (Wallis 2004, 10). None of this is problematic for my account, it is rather an indication of what the virtue of mindfulness might look like from the inside, if we were to strive to achieve it. Schwitzgebel has not presented an argument against mindfulness,⁵³ he has simply laid out clearly some of the challenges we all face in striving to be mindful.

3.4 The Benefits of Mindfulness

Another classic hallmark of virtue is: virtues are traits of character that benefit their possessor (see e.g. NE 1106a17-24). This quality is almost definitional for Aristotle, as virtues are those qualities of character that are a necessary part of *eudaimonia*. It is important to keep this quality in mind, however, because there are many non-virtuous traits that might ameliorate flaws. A classic example is the relationship between brashness and fear. Fear is a characteristic challenge to and temptation for humans, and brashness, understood as the trait to ignore fear or actively do the opposite of what fear suggests, ameliorates it. But brashness is not a virtue because, among potential other flaws, brashness does not benefit the possessor. Brash people take careless risks and unjustifiably expose themselves to harm in myriad ways. In another example, the disposition to starve oneself might ameliorate our temptation to gluttony, but starvation is not good for the possessor.

In each case a virtue term, courage and temperance respectively in the previous paragraph, names not just any trait opposing a characteristic flaw but rather

⁵³ In fact, Schwitzgebel seems open-minded about the possibility of mindfulness meditation as the sort of training that would make us better at introspection (Schwitzgebel 2008, 259 n.21).

a proper or reasonable trait of opposition. This is the core insight of Aristotle's doctrine of the "mean" or intermediate state that virtues occupy (NE 1108b11-15). Virtue is an intermediate state between two vices, a state that strikes the right balance between competing interests. And part of how we judge the right balance is by asking what state benefits the possessor. This shows that the mean is relative to the actor, as in Aristotle's famous example of Milo the athlete, who needs a lot of food, as compared to the beginner, who needs little (NE 1106b2-5). The proper or reasonable trait of temperance in this case, and in any other, consists in a disposition to seek the proper amount and type of food so as to benefit the possessor.

For many virtues, benefit to the possessor is basically assumed, or argued for only in conceptual terms. With courage, the trait is required in order to pursue the goals one values in life in the face of fear. We all also have various goals and projects in life that we find valuable, and that give life meaning, but that involve risks or sacrifices of which almost everyone is afraid. Therefore, that in order to pursue those projects and goals we need the proper relationship to fear, and we need to be able to act well in the face of it. On this view, courage just must benefit the possessor, because it helps the possessor live the life they seek to live, a life of meaning and value.⁵⁴ There may be a similar sort of argument with regard to mindfulness, because

⁵⁴ One might object that this passage makes the virtues look instrumentally, rather than intrinsically, valuable. I think that virtues have both kinds of value, and do not mean to deny their intrinsic good. But their instrumental value makes the argument presented here more intuitively forceful. It is not clear to me what we would think about the virtues if they did *not* have instrumental value, if humans did not need courage, to continue the example here, in order to pursue what we find valuable in life. This is a counterfactual, however, because we do need courage, and this need helps illustrate why the virtues benefit their possessor.

the forces of our automatic cognitive system systematically undermine our various goals and projects in life.⁵⁵ This argument for mindfulness, it seems to me, has less intuitive force than with courage. But the good news for mindfulness is that there is strong empirical evidence from social psychology that it benefits the possessor, regardless of the intuitive link between mindfulness and flourishing. While there is widespread discussion in current psychology about mindfulness, this section focuses on the work of Ellen Langer, who developed a psychological account of mindfulness with an accompanying, impressively large set of empirical data. Her view is well developed, and based in an experimental framework rather than a clinical one. These factors make her work more appropriate than most for consideration by virtue ethicists.

The 1979 title experiment of Langer's latest book Counterclockwise illustrates her concept of mindfulness (Langer 2009, 4-10). Langer was interested in the question "what effects turning back the clock psychologically would have on people's physiological state", hence the name the "counterclockwise" study. In order to test her theories, she tried to recreate the world as it was in 1959 and see how elderly test subjects reacted. She and her students recruited two groups of participants. The experimental group attended a week-long retreat at a location Langer made to look like 1959. It was populated with media from 1959, their days were populated with

⁵⁵ One might object that our automatic cognitive systems are adaptive, and that they let us do many good things we could not otherwise do, and the like. This is, of course, true! But fear, and bodily desire, are also adaptive, and also necessary for a flourishing human life. Courage is not fearlessness, and temperance does not demand starvation or celibacy. Likewise, mindfulness is a proper disposition to act in light of our automatic cognitions. Mindfulness is not non-automaticity, it is proper automaticity.

discussions and activities based on it being 1959, and the participants were told to speak and act as if it were actually September, 1959. They were specifically instructed not to reminisce about 1959 but to try to act as if it were actually 1959. In her words, “they should return as completely as possible in their minds to that earlier time.” In particular, they wrote biographies for each other written as if it were 1959, and they shared photos of themselves as they were in 1959. The control group went to the same location for a retreat the following week, living in exactly the same surroundings and doing the same activities as the experimental group. The only difference with the control group is that they acted as if 1959 were twenty years in the past. They wrote biographies in the past tense, they shared photos of their 1979 selves, and they were instructed to “*reminisce* about the past and thus largely keep their minds focused on the fact that it was not 1959.” Langer measured all of the test subjects before and after their retreats with an impressive battery of tests.⁵⁶

What happened? First, participants in both the experimental and control groups showed measurable improvements: both groups had better hearing and memory and both gained weight and grip strength. On many other measures, however, the experimental group, which was asked to behave as if it were really 1959, improved much more: “The experimental group showed greater improvement

⁵⁶ Interestingly, Langer contacted leading geriatricians to seek hallmarks of aging and was told that there are no consistent proxies for chronological age. So she and her students tried to measure as many things as they could: “In addition to weight, dexterity, and flexibility, we planned to measure vision with and without eyeglasses, for each eye separately and together, as well as sensitivity to taste. We would give potential participants intelligence tests, to assess how quickly and accurately they could complete a series of paper-and-pencil mazes, and we would test their visual memory. We would also take photographs that we could later evaluate for the *appearance* of changes. Finally, they would each be asked to fill out a psychological self-evaluation test” (Langer 2009, 6).

on joint flexibility, finger length (their arthritis diminished and they were able to straighten their fingers more), and manual dexterity. On intelligence tests, 63 percent of the experimental group improved their scores, compared to only 44 percent of the control group... Finally, we asked people unaware of the study's purpose to compare photos taken of the participants at the end of the week to those submitted at the beginning of the study. These objective observers judged that all of the experimental participants looked noticeably younger at the end of the study" (Langer 2009, 10). The objective observers were particularly important to Langer—these observers knew nothing of the study and were not shown comparative pictures, but only one picture, either the before or after, of each participant. "They rated the appearance of the participants for whom we turned the clock back as being more than two years younger" (Langer 2009, 179).

For Langer, "[m]indfulness... is the simple process of actively drawing distinctions. It is finding something new in what we may think we already know" (Langer 2009, 182). She notes two hallmarks of mindfulness that are apparent in the counterclockwise study. First, mindfulness involves re-categorization, reclassifying current concepts or things and continually creating new categories (Langer 1989, 63). That is, being mindful involves both knowing that our conceptual categorization of the world is not inevitable and also knowing how that categorization dictates our interactions with the world. In a related study, Langer has shown that the relationship we have as children with elderly relatives is strongly correlated with our own health as we get older. (Langer 2009, 154-155) The implication is that our own cognitive

associations with old-age are strongly correlated with our experience of old age. The counterclockwise study suggests that this correlation is not immutable, and that we can experience more vigor in old age if we can re-conceptualize what it means to be old (see particularly Ch. 9 “Mindful Aging” in Langer 2009, 153-176). A simpler example of this effect that Langer repeated in many different contexts involves giving people an object and either suggesting to them what it could be or telling them what it is. Subjects that are told an object could be one thing are mindful that the object could also be something else, and are more likely to use the object creatively as needed, for example using a relaxation technique for a part of the body not specifically mentioned in the instructions (Langer 2009, 140) or using a dog’s rubber chew toy as an eraser (Langer 2009, 100).

Second, “a mindful state also implies openness to new information” (Langer 1989, 66) and an “openness... to different points of view” (Langer 1989, 68).⁵⁷ By situating the experimental subjects in a virtual 1959, Langer induced them to notice information they were blind to before and see themselves in the way they saw themselves twenty years earlier. This change of perspective had a profound effect on their well-being after only one week. This aspect of mindfulness is illustrated poignantly through another study by Langer and her student Ali Crum on hotel room attendants (Langer 2009, 111-124; Crum 2007). Langer and Crum wondered to what extent the positive effects of exercise depend upon the mindset of the exerciser. They

⁵⁷ This video <http://www.youtube.com/watch?v=Ahg6qcgoay4> is a priceless example of the mindless mode.

focused on hotel room attendants. These people have a job that meets or exceeds the surgeon general's requirements for a healthy lifestyle, but "room attendants appear at risk with respect to blood pressure, body mass index, body fat percentage, total body water percentage, and waist-to-hip ratio, all important indicators of health." If room attendants were simply informed about the positive benefits of exercise, and that they were in fact exercising at work, would there be a measurable difference? In order to test this, Langer and Crum handed out a notice and gave a short presentation that explained the benefits of exercise and also explained how the daily housework hotel room attendants do "satisfies the CDC's recommendations for an active lifestyle". The control group was treated exactly as the experimental group, except they were not given information about how their work was exercise. After four weeks, participants in the experimental group lost an average of two pounds, and increased body water percentage (which shows an increase in muscle mass), and reduced body fat percentage. The control group actually gained body weight and body fat. Simply realizing that what they were already doing counted as exercise seemed to have a distinct physical effect on the test subjects.

Langer relates a litany of experiments that show beneficial effects like these when experimental subjects are directed to pay attention to, and reconceptualize, their circumstances and actions (Langer 1989, 2009). Based on the totality of her work, she concludes that our health, our well-being, and even our physical capacities are influenced by actions and attitudes that are characteristic of the disposition of

mindfulness.⁵⁸ For purposes of the arguments here, Langer's summaries of this work are sufficient. The goal is not to establish clear necessary and sufficient conditions for mindfulness or to defend a necessary empirical link between the mindful disposition and benefits to the possessor, it is only to show that there is strong and varied experimental evidence for both the existence and benefits of mindfulness. Put another way, there is good evidence that mindfulness is a disposition that benefits the possessor.

An interlocutor might point out that most of Langer's work involves manipulating the situation so as to induce, so to speak, people into mindfulness. This may be true, although it is not clear that manipulating the way people cognize a situation is the same as manipulating the situation itself. But if this is true, does it mean that her work is actually situationist in nature? Has Langer only succeeded in describing the qualities of situations that are conducive to good behavior rather than bad? There are at least two ways that Langer's studies are potentially inconclusive as to whether we can have the disposition to be mindful without a situational stimulus. First, she does not give evidence regarding whether mindfulness is a characteristic that exists independent of experimental prompting and second, if it is such a characteristic, whether mindfulness is something that can be learned. To measure the former, it would take a reliable metric for determining the mindfulness of study participants. Langer does not discuss such a possibility, and her mode of discussing

⁵⁸ See particularly her current efforts to redo the counterclockwise experiment with cancer patients to see if the size of their tumors are reducible (Grierson 2014).

mindfulness does not lend itself to creating a battery of tests to measure mindfulness. But there is also no indication in her work that we cannot be mindful ourselves about how we act in the world; indeed, she wrote her popular books largely to encourage people to be more mindful. As I will describe briefly below, Langer clearly believes we can cultivate and possess an independent disposition to be mindful. She also does not give any evidence about whether situational manipulations encouraging mindfulness actually encourage a mindful disposition in the participants, as opposed to eliciting mindful behavior in that situation alone.

This being said, it is important to see that whether or not the trait of mindfulness is under conscious control is not necessarily relevant to its classification as a virtue. Aristotle famously believed that some virtues were not available to everyone, that they were contingent on the proper upbringing, resources, and circumstances of an adult life. Many people find this doctrine in Aristotle unattractive. They like to think, perhaps, that all virtue is available to us throughout human life. Perhaps this is influenced by a Kantian, or a Stoic, conception of will. However, it is clearly part of a traditional view of virtue, and even of common sense about the limits of our will, that our characters are fully formed as adults. But if there are mindful people in the world, and if mindfulness is a character trait that opposes a human flaw, then this is enough to consider it a virtue. For whether it is a trait we can acquire as adults or not, it is one we ought to try to inculcate and develop in children inasmuch as it benefits them and makes them better human beings.

Langer herself actually briefly addresses arguments like those made by situationists, saying “my social psychology colleagues are fond of saying that behavior is context-dependent. I am saying that if we are mindful, we can create the context” (Langer 2009, 183) and citing an interesting paper she published in 1980. In that paper, she generally argues that mindlessness represents the acceptance of apparent, given structures in the environment whereas mindfulness is the activity of structuring the environment. In her words, “the apparently structured environment *suggests* certain modes of humanly minded engagement, but it does not *dictate* that mode. Memory and imagination are the only limits of the activity” (Langer 1980, 102).

Langer provides some compelling everyday examples illustrating how we have the power to dictate our own context. In her words,

“In this light [of her concept of mindful control], the response that George Mallory gave to the question ‘Why do you climb Mount Everest?’ is not the *non sequitur* that it is often taken to be. ‘Because it is there’ is a very revealing, emotional response. It only seems mundane because Mallory neglected the importance of the second half of the answer, ‘... and I am here.’ I (the mountain climber) saw it (the mountain to be climbed) as a challenge. It is not just anybody who sees the mountain as a challenge. A tourist would not feel the urge to climb the mountain. The sight might strike the tourist as something to be photographed. Nor does the mountain represent a singular challenge. For other developed selves, the mountain calls out other challenges. For the skier, Yuichiro Miura, Mount Everest is seen as a challenging mountain to be skied down” (Langer 1980, 107-108).

Langer is pointing out what should be basically obvious to us: that the context of a situation is at least partly determined by the perspective and attitude of the person experiencing it. The situationist experiments take advantage of contexts in which many or most of us see situations as already structured in such a way as to restrict our

ability to interact with them. But there is nothing about the situation that forces us to act deplorably. If we were able to cultivate a mindset of mindfulness, or of always aspiring to virtue in general, then it seems possible that we might avoid many of the pitfalls situationist experimenters put in front of us.

This line of reasoning indicates that this mindfulness could serve as a counter to the interaction between automaticity and control that gives rise to moral incongruency in the situationist experiments. In the Milgram experiments, for example, a mindful person experiencing the discomfort of compliance might pause and re-think the options presented to them. As Doris points out, “the only power the experimenter holds, initially, is power the subject agrees to cede to him, with the option (in principle) of revoking it at any point” (Doris 2012, 386), and yet the context of the Milgram experiments seemed to encourage subjects to surrender far more power than they would have surrendered up front. The mindful person critically examines the situation around them and is open to new information and new conceptual categorizations, and they might use these to maintain control and power in the situation of the Milgram experiment.

3.5 Conclusion

One of the oddest rhetorical moves in the situationist literature is the manner in which psychological evidence of moral incongruity is taken as evidence of some immutable fact about humanity rather than as evidence of a dysfunction or defect needing cure. It seems akin to taking the discovery of some mechanism of cancer as a cause for resigning ourselves to early death as opposed to an opportunity to

understand how to combat a particular type of human disease. An optimist about virtue can take the same research as Doris, and even take his description and characterization of that research as a given, and see that research as a cause for hope about virtue theory. This way of thinking takes the psychological evidence on moral incongruity as diagnosing a problem for moral theory to overcome, and I hope this chapter has shown that this way of thinking is reasonable, and fits in well with the virtue ethics tradition.

Ultimately, Doris' recent work may be consistent with what is argued here. In concluding their recent article, Merritt and Doris say "While there is still an important role for the strategic application of individual powers of deliberation, self-monitoring, and self-control, it must be supplemented with systematic attention to the way 'internal' cognitive processes interact with environmental factors such as interpersonal relationships, social and organizational settings, and institutional structures" (Doris 2012, 392). This description fits Langer's account of mindfulness well, where the mindful agent is attentive to all of these details about the environment in which they act. The disagreement might then focus on the power of individual mindfulness to solve this problem by itself, which a virtue ethicist can endorse, as opposed to the need to deliberately seek out social structures and situations that encourage right action, as Doris appears to believe we must do. But this could be a dispute over the details and not over the basic dynamic at work.

One possible area of further exploration suggested by my arguments is to run some of the most well-known situationist experiments on experienced meditators.⁵⁹ This class of people is a recently popular subgroup of experimental subjects. There is at least anecdotal evidence in the psychological literature that expert meditators are less emotionally reactive (Ortner et. al. 2007) and have greater cognitive flexibility [many citations to pick though at <http://www.apa.org/monitor/2012/07-08/ce-corner.aspx>]. These factors could lead to different results in the situationist experiments. Studying people who have deliberately tried to cultivate mindfulness could help us learn more about the mindful disposition, and also more about the automatic processes that Doris suggests give rise to incongruency.

Finally, while this section focused on the empirical evidence available from psychology, it is important to consider the prior probability that a particular trait is a virtue before looking to empirical evidence. Mindfulness is exactly the sort of ancient, rich, and important concept that is a plausible candidate for a virtue.⁶⁰

⁵⁹ There is a classic problem of the difficulty of selecting a set of test subjects on unusual criteria without alerting them to the fact that they are being studied. Of course, many have thought it would be interesting to study the one third of people who do not comply in Milgram's base case. Do they resist other situational pressures? But it is so difficult as to be nearly impossible, and morally problematic, to try to repeatedly test the same people's behavior without tipping them off as to what is happening. Given the popularity of studying experienced meditators in neuroscience and psychology, however, I suspect that a clever experimenter would not have a problem designing an effective experiment.

⁶⁰ It is difficult, and beyond the scope of this work, to try to capture the varied, rich discussions of mindfulness in the Buddhist tradition. In order to give some sense of the notion of mindfulness in the Buddhist tradition, and hopefully motivate the intuition that Buddhist mindfulness is very close to what has been discussed here, I present here two passages from primary Buddhist sources on mindfulness. I hope to have the time to engage with the Buddhist tradition more fully in later work. First, from the *Sattipatthana Sutta*, an account of mindfulness that emphasizes accurate perception and understanding of the mind:

“(Contemplation of Mind)

And how, bhikkhus, does a bhikkhu abide contemplating mind as mind? Here a bhikkhu understands mind affected by lust as mind affected by lust, and mind unaffected by lust as mind unaffected by lust. He understands mind affected by hate as mind affected by hate, and mind unaffected by hate as mind

Dealing with the brain's automaticity is consistent with historical teachings of mindfulness. And those traditions are interested in precisely the sorts of problems of automatic cognition that Doris focuses on.

unaffected by hate. He understands mind affected by delusion as mind affected by delusion, and mind unaffected by delusion as mind unaffected by delusion. He understands contracted mind as contracted mind, and distracted mind as distracted mind. He understands exalted mind as exalted mind, and unexalted mind as unexalted mind. He understands surpassed mind as surpassed mind, and unsurpassed mind as unsurpassed mind. He understand concentrated mind as concentrated mind, and unconcentrated mind as unconcentrated mind. He understands liberated mind as liberated mind, and unliberated mind as unliberated mind." (Bhikkhu Bodhi 1995, 150)

Second, from the *Dhammapada*, a poetic take on the difficulty of mindfulness:

"Trembling and quivering is the mind,
difficult to guard and hard to restrain.
The person of wisdom sets it straight,
as a fletcher does an arrow.

...

It is good to tame the mind,
alighting, as it does, wherever it desires—
swift, resistant to restraint.
A tamed mind gives rise to ease.

So difficult to perceive, exceedingly subtle,
alighting, as it does, wherever it desires.
Let the person of wisdom guard the mind.
A guarded mind gives rise to ease." (Wallis 2004, 10)

4. Humility as a Virtue

“I tried to think of something clever to say, but before I could speak, [Dr. Martin Luther King] asked why I was studying for a Ph.D. in art history. He asked what I thought art could accomplish that other forms of communication could not. I remember that he said that he’d rarely discussed art, or even thought much about it. As I stammered an answer I cannot recall, he listened with the concentration of someone who genuinely wanted to understand. Never before, and rarely since, had I witnessed such authentic humility. It was so simple, so powerful a form of energy that for a few moments it freed me from bondage to myself” (Hood 2013).

Traditionally, a virtue must be a constitutive part of an ideal human, or of a perfectly virtuous being. But the ideal, perfectly virtuous human would never make an incorrect moral judgment. Thus, there is no opportunity for such a being to be humble. If humility is some form of lack of certitude in one’s own judgments, then it seems inconsistent with the character of the perfectly virtuous; humility could never be better than a false modesty for them. Another way to put this is that the virtue of humility is self-undermining. Humility may be appropriate for the non-virtuous, but as one gains in virtue, the importance of humility seems to fade, for what remains to be humble about?

These classic arguments about humility imply that either it is not a virtue, we are wrong about what the virtue humility is, if it is a virtue, or we are missing a resolution to these puzzles. This chapter makes two arguments in favor of the claim that humility is a virtue. First, it resolves these puzzles by looking at humility through the lens of the skill analogy. Then, it shows how there is strong empirical evidence supporting the claim that humility is a virtue. We do better if we have humility about all of our judgments, and the evidence is strong enough that it would be a serious

problem for a moral theory if it could not explain the value of humility. With these two arguments, I hope to show that we both have strong motivations to be humble and that humility is coherent as a particular virtue in the general context of virtue ethics.

4.1 The Concept of Humility

Any theory of humility as a virtue should seek to satisfy three desiderata:

- “(1) There is the thought that the modest person has *accurate beliefs* about (or presents herself accurately with respect to) herself or her accomplishments. This seems required for modesty to be a virtue.
- (2) There is the thought that the modest person believes (or presents herself as believing) that she, or her accomplishments, *rank low* on the relevant scale. This seems required for actual modesty. And,
- (3) there is the thought that the modest person may well have, indeed, even must have, *genuine accomplishments* to be modest about. Otherwise, as Bennett says, ‘it wouldn’t be modesty’” (Schueler 1997, 476-477) (*italics added*).

The first reflects the belief that virtues cannot require a serious epistemic flaw. The second reflects the everyday notion of what humility involves, a low self-evaluation. The third reflects the idea that a low self-evaluation is not humility if a person has not accomplished anything in the relevant area: a runner who comes in last is not being humble in saying they were the worst runner in the race.

While each of these claims captures a common belief about humility, taken in combination they present a puzzle. It seems impossible to satisfy all three at once (Schueler 1997, 477). If any two are true, the third seems to be false: if an agent accurately believes her accomplishments rank low, then she must have no genuine accomplishments; if she accurately believes she has genuine accomplishments, then she does not have a low opinion of herself; and if she has genuine accomplishments

that she ranks low, then it seems she has the epistemic flaw of ignorance, and lacks accurate beliefs, which does not seem like virtue to most.

This puzzle about humility has been a topic of debate at least since Aquinas. He stated the puzzle as: “virtue is ‘the disposition of that which is perfect’ (Phys. vii, text.17). But humility seemingly belongs to the imperfect: wherefore it becomes not God to be humble, since He can be subject to none. Therefore it seems that humility is not a virtue” (Aquinas 2008, §161). Much more recently, Norvin Richards says “if to be humble is to have a low opinion of oneself, there do appear to be people who could be humble only through self-deception or ignorance” (Richards 1988, 253; Cf. Statman 1993, 425-426). If virtues are the traits that all people need in order to flourish, humility cannot be a virtue because the best people do not need it for flourishing.

Authors in the literature have tried to rescue humility as a virtue by rejecting desideratum (1) and/or (2) (See Driver 1989; Richards 1988; Flanagan 1990; Button 2005). These answers are generally unsatisfying because they ask us to give up on one of the three desiderata that make humility an intuitively compelling virtue. In contrast to the literature, I will argue that we can make sense of humility as a virtue and satisfy all three desiderata if we think more clearly about how it is that a humble person engages in self-evaluation. In particular, I will argue that an analogy with practical skills illustrates how a virtuous person could have the accurate belief that their genuine accomplishments rank low on the relevant scale.

4.1.1 The Literature on the Puzzle

The oldest response to this puzzle in the philosophical literature is a rejection of desideratum (2). This position, particularly associated with Aristotle, asserts that the virtuous person has an accurate self-evaluation and genuine accomplishments and takes this to imply that the virtuous person has a high self-evaluation. This account implies the virtue of *megalopsuchia*, usually translated as “magnanimity” (See NE 1123a35-1125a35). The magnanimous person “thinks himself worthy of great things and is really worthy of them” (NE 1123b3-5). It is not a trait that is available to those who are not worthy of great things (NE 1123b5-6). Any agent who undervalues their high worth is pusillanimous,⁶¹ a characteristic vice opposed to magnanimity. Aristotle also discusses a nameless virtue associated with small honors, analogizing the relationship between this virtue and *megalopsuchia* to the one between generosity and magnificence (NE 1125b1-27). This nameless virtue is the closest he gets to explicitly discussing something like humility, but I think it is the wrong place to look for purposes of my argument in this section, because the puzzle arises when considering great, virtuous people, who are presumably worthy of great honor (NE 1124a1-2 says: “Magnanimity, then, would seem to be a sort of adornment of the virtues; for it makes them greater, and it does not arise without them.”). As Irwin points out (Irwin 199, 220, n. 3, §1) it is possible that a charitable reading of Aristotle’s writing on *megalopsuchia* and this nameless virtue is consistent with humility as a virtue,

⁶¹ The Greek word is *micropsuchia*, literally meaning “small-souled” and a clear antonym to *megalopsuchia*. Pusillanimous often has the connotation today of cowardice, which is not what Aristotle is talking about here. Irwin points this out (Irwin 1999, 220-221).

however, Aristotle seems to be widely taken as a representative of the position that a paragon of virtue is not humble or modest, and therefore that these traits are not virtues (see, e.g., Driver 377 n.4; Ben Ze-ev 239).

The classic contrast to this position is given by Aquinas:

“A thing is said to be perfect in two ways. First absolutely; such a thing contains no defect, neither in its nature nor in respect of anything else, and thus God alone is perfect. To Him humility is fitting, not as regards His Divine nature, but only as regards His assumed nature. Secondly, a thing may be said to be perfect in a restricted sense, for instance in respect of its nature or state or time. Thus a virtuous man is perfect: although in comparison with God his perfection is found wanting... On this way humility may be competent to every man” (Aquinas 2008, §161).

If the relevant scale of self-evaluation includes divine perfection, then it is clear that even the best human can accurately rank herself low on the scale in comparison to divinity. The best human still ought to be humble before God. This view is also reflected in modern religious virtue theorists. For example, Robert Adams believes “The quest for virtue cannot yield more than an always and necessarily incomplete and fragmentary approximation to a transcendent goodness (the goodness of God, I would say) which is its reference point” (Adams 2006, 173).

These responses to the puzzle are inadequate for purposes of this paper. First, the argument of this paper is to see how far we can go with an everyday notion of the virtue humility; a simple rejection of this everyday notion is beside the point. Second, the religious metaphysics of Aquinas do ground a clear response to this puzzle. However, I am interested in whether humility makes sense as a virtue for secular virtue ethics. Furthermore, while human humility seems morally required if God exists, it is not clear how it constitutes a virtue. That is, if humility simply means

acknowledging our human limitations in contrast to a divine being, it hardly seems a human excellence but rather something more like a minimum qualification to be a mature mortal agent.

In contemporary work, there are two lines of response to the puzzle. First, Julia Driver has an intriguing argument that humility just does involve the epistemic flaw of ignorance about one's own worth (Driver 1989). Driver believes that a modest agent "if he speaks, then he understates the truth, but he does so unknowingly. This entails that the modest person is ignorant, to a certain degree, with regard to his own self-worth" (Driver 1989, 376). She favors this account because she thinks it offers a good explanation of the oddity of the phrase "I am modest", which she claims is self-defeating, akin to Moore's Paradox (Driver 1989, 375-376). Driver uses this account of modesty to develop a detailed criticism of Aristotelian virtue ethics, and argue instead for a consequentialist virtue ethics (Driver 2001).

Driver's arguments are interesting, but I generally agree with criticisms that have been leveled against her position (see, e.g., Flanagan 1990; Statman 1992, 423-424; Schueler 1997, 468-470; Winter 2012, 537-539). The two most problematic parts of Driver's account are that it has (1) virtue require an epistemic defect and (2) the possession of virtue be involuntary (Statman 1992, 423). It is possible that certain virtues would be consistent with particular epistemic defects, but it is strange to suggest that any virtue requires an epistemic defect. This strangeness is exacerbated by the fact that the epistemic defect in question is one that is curable: if the humble agent comes to believe the truth about herself, then she is no longer humble.

Therefore, on Driver's account humility seems to have a quality of involuntariness, insofar as we cannot generally voluntarily choose what we believe to be true. Virtue traits are more robust and stable than this, at least in the Aristotelian tradition. These problems, in addition to others identified in the extensive critical literature, are sufficient motivation to consider alternative accounts of humility.

Other contemporary scholars, particularly Driver's critics, have tried to redefine humility in such a way as to avoid the force of the puzzle. The first move, and apparently most popular, is to soften desideratum (2), the low self-evaluation aspect of humility. These authors argue that humility involves the disposition not to overrate one's abilities, or to keep one's achievements in perspective, rather than a disposition to rate oneself lowly on the relevant scale. For example, Norvin Richards defines humility as "an inclination to keep one's accomplishments, traits, abilities, etc., in perspective, even if stimulated to exaggerate them" (Richards 1988, 256; Cf. Flanagan 1990). Mark Button says, "I define democratic humility as a cultivated sensitivity toward the incompleteness and contingency of both one's personal moral powers and commitments, and of the particular forms, laws, and institutions that structure one's political and social life with others" (Button 2005, 841). These definitions involve a re-evaluation of humility as not underestimation of one's own value but proper estimation, or "non-overestimation" (Flanagan 1990).

These redefinitions are unsatisfying because they fail to capture the force of the intuition that a low self-evaluation is an important aspect of humility. Many agents exhibiting these traits would not be described, in everyday language, as

humble. For example, in a case where the agent was truly great, it seems like this account collapses into Aristotle's magnanimity: the agent would have nothing to do but declare her greatness. This is because keeping accomplishments "in perspective" is a vague standard. What is the proper perspective to take on accomplishments that are at the pinnacle of human achievement? Flanagan suggests the world's fastest human might remain modest because he "might think that being the world's fastest human is not so important *sub specie aeternitatis*" (Flanagan 1990, 425) (Cf. Ben-Ze'ev 1993, 237: "our talents and accomplishments are of less importance when related to the role and place of each one in the universe."). But I do not see how this sort of justification could ground a virtue, or how it can justify humility. With regard to a virtue, it seems that recognizing one's smallness in the light of the universe is not a particularly excellent human achievement, but rather a basic sign of mature ethical reflection. It is not excellent to know that one is a finite, mortal being. With regard to humility, it does not seem to be humble to declare oneself the greatest artist the world has ever seen, but then acknowledge that human artistic achievement is not so grand *sub specie aeternitatis*. My account makes sense of a great artist being self-critical, of focusing on how they can be a better artist as opposed to how art figures in relation to all of the universe.

The second redefinition move is to understand humility as the lack of particular desires or dispositions. Schueler says the humble person is "someone who doesn't care whether people are impressed with her for her accomplishments. That is, she lacks a certain desire or set of desires, namely, that people be impressed by her

for what she has accomplished” (Schueler 1997, 478-479).⁶² Winter defines humility as “having a disposition to refrain from putting forward one’s accomplishments” (Winter 2012; 537). Schueler and Winter use different terms, “desire” versus “disposition”, but both focus on how the humble person relates to public discussion of their accomplishments. The differences between desire and disposition are not important to understand why their accounts fail, so the two terms are lumped together here.

Schueler and Winter underestimate the extent to which self-evaluation, and not just an attitude toward other-evaluation, is an intrinsic part of humility. We can only make sense of humility as a virtue if it involves some aspect of the way we view ourselves, and this point about self-reference is a widely shared intuition about humility.

Winter’s compelling example of Bill Gates illustrates the shortcomings of his position, and foreshadows my argument below. Winter relates an episode in which Gates goes to great lengths to describe how his success was dictated by early access to computers and experience with programming, a set of experiences due to luck. When asked how many people in the world he thought had similar lucky circumstances to build upon, Gates says “if there were fifty in the world, I would be stunned” (Winter 2012, 538). This example is best understood by appreciating how Gates really sees himself differently than we do, rather than by explaining his conduct

⁶² Richards suggests something similar, but does not develop it in detail. He says, “Notice, however, that shunning the praise and attention one serves need not betray an improper sense of oneself or of what one has done” (Richards 1989, 256).

as a disposition not to boast about something (Winter 2012, 538). Gates does seem to show noble modesty, but Winter's understanding of modesty as refraining "from putting forward one's accomplishments" does not capture this episode very well. Gates' modesty is noble and impressive because it reflects a genuine judgment that luck really did influence his success in profound, and likely to be underestimated, ways. He is not expressing just a reluctance to harp on his own accomplishments; he has perspective on his accomplishments and is telling us what he sees, and how much he appreciates the role luck genuinely played. This is clear if we reflect on how we would view Gates if we knew he really thought luck did not play a big role in his success and he was just expressing a disposition to avoid boasting. Gates does not seem humble at all in such a case, but rather particularly polite or tactful. Such a disposition not to boast makes modesty more a quality of etiquette than ethics.⁶³

This is enough reason to reject both accounts. But Schueler's account is even more problematic, because the humble person ought to care, at least to some extent, how other people evaluate their accomplishments (Cf. Winter 2012, 537). The humble person can care about truth, and can have friends and loved ones whose opinion she cares about, and be hurt when a loved one objectively underestimates or under appreciates the humble person's achievements. Take the hypothetical case of a person who just earned her Ph.D. in the humanities. She feels proud of her accomplishment, of all the effort she put into earning the degree, and of the years of

⁶³ Note the similarity here to Aristotle's discussion of truthfulness, where boastfulness is a vice (NE 1127a15-1127b30). I am not certain that truthfulness makes sense as an Aristotelian virtue, despite what Aristotle wrote, but non-boasting is at best the lack of a vice for Aristotle, not a virtue itself.

dedication it took. But she is accurately humble, she keeps her accomplishment in perspective, and she does not brag. Also assume that her parents consider her accomplishment no accomplishment at all. They consider her degree frivolous and they wish she had been an engineer, or doctor; they constantly express their disappointment in her decisions. It seems an extreme, and odd, requirement to say that in order to be humble she must be indifferent to the gross injustice of her parents' view. As this case indicates, Schueler's account gives too little attention to the way that injustice can generate legitimate caring, even in a humble person. A desire to be judged justly should not be a disqualifier for humility (Cf. Richards 1989, 258). Indeed, while the position given below is inconsistent with Aristotle's magnanimity, his articulation of pusillanimity (*micropsuchia*) compellingly describes a vice that the Ph.D. student would have if she demanded so little respect for herself (see particularly NE 1125a20-30).

4.1.2 A Solution: The Skill Analogy Argument

"I do the same thing over and over, improving bit by bit. There is always a yearning to achieve more. I'll continue to climb, trying to reach the top, but no one knows where the top is!" (Jiro Ono, Jiro Dreams of Sushi)

The problem of making sense of humility as a virtue centers on the asymmetry between self- and other-evaluations of the humble person. All agree that the humble person must have genuine achievements about which they are humble. But this implies that either the humble person accurately understands their genuine achievement and does not rank themselves lowly, and thus they do not seem humble, or that the humble person has an epistemic flaw and does not think they have made

genuine achievements, and thus they do not seem virtuous. We seem forced to choose between a virtue of accurate self-evaluation that is not humility or a humility that is not a virtue. But there is a better way to deal with the second horn of this dilemma. The key is to notice that the intuitive force of objections to Driver's account do not require rejection of her observation that there is an asymmetry between self- and other-evaluations of the achievement of a humble person. Driver privileges the other-evaluation and labels the self-evaluation as false, as symptomatic of an epistemic flaw, but we need not follow suit. We can instead privilege the self-evaluation because the scales of evaluation being used in each case are different: the humble person is evaluating themselves on a different scale than other people are evaluating them on. The humble person sees a wider and deeper range of possibilities for relevant actions, and they therefore rank their own achievements much lower on the relevant scale than those with a less developed moral sense would rank their achievements. That is, the evaluations of the humble person's activity are achievement relative, because the scale of evaluation changes with one's achievement in the relevant area.

I will make this argument by way of the skill analogy. The skill analogy is the claim that the acquisition and activity of virtue is analogous to the acquisition and activity of practical skills, including artisan and artistic skills. This style of reasoning about virtue has roots in Aristotle, but has also received serious recent attention in Julia Annas' book, Intelligent Virtue. As she puts it there, the skill analogy claims that,

“exercising a virtue involves practical reasoning of a kind that can illuminatingly be compared to the kind of reasoning we find in someone exercising a practical skill. Rather than asking at the start how virtues relate to rules, principles, maximizing, or a final end, we will gain by looking at the way in which the acquisition and exercise of virtue can be seen to be in many ways like the acquisition and exercise of more mundane activities, such as farming, building, or playing the piano” (Annas 2011, 1).

As an example of this style of reasoning, Aristotle analogizes the acquisition of a virtue with the acquisition of crafts:

“Virtues, by contrast, we acquire, just as we acquire crafts, by having first activated them. For we learn a craft by producing the same product that we must produce when we have learned it; we become builders, for instance, by building, and we become harpists by playing the harp. Similarly, then, we become just by doing just actions, temperate by doing temperate actions, brave by doing brave actions” (NE 1103a31-1103b2).

This theme runs throughout the Nichomachean Ethics, where Aristotle compares the acquisition and nature of ethical *aretê* with other forms of excellence.

The skill analogy is useful because practical skills commonly display the achievement relative judgments that ground this account of humility. This discussion focuses on two particular classes of skill for illustration, artistic performance and scientific research, but there are many other skills which have this feature. Annas herself includes a skill analogy argument along similar lines in her book. In relation to the question of how to explain to someone who is not virtuous what enjoyment there is in being virtuous, particularly when many have a common sense notion that doing the moral or ethical thing is somehow unpleasant, Annas says: “In the case of skills the answer is clear. We don’t, and don’t expect to, have any idea of what it is like to exercise a skill with mastery, and thus with enjoyment, unless we have some grounding in the skill. Watching a master potter, or sports star, gives no idea of what

it is like for them to create pottery or score goals” (Annas 2011, 80). We do not understand how a master artisan experiences their craft unless we are masters, and we likewise do not understand how a master evaluates their own achievements unless we ourselves have a degree of mastery. This argument is therefore inspired by Annas’ treatment of the skill analogy, but extends it to an area, achievement relativity, which Annas does not consider to solve a problem about a virtue, humility, that Annas does not discuss.

Achievement relativity is intuitively plausible for artistic performance skills. For example, when I hear Yo-Yo Ma play, his tone sounds basically perfect to me; I cannot identify any flaw in his playing. If he were to complain that he played poorly one day, it would seem like false modesty to me, like he was somehow ignorant of how well he performed. But I think that the best explanation for the disagreement he and I might have about his performance would be in terms of our different sensitivities to the possibilities for excellence in cello performance. My hypothetical Ma has a sensitivity to cello performance that makes it possible for him to make a true, humble judgment about his performance that seems ignorant or false to me but is in fact not ignorant nor false. Thus, he can make an accurate, self-deprecating, judgment about his own admittedly high level of accomplishment.

One of the characteristic experiences of becoming proficient in any art seems to be an enhanced perception of the potential for flaws in that art. The better you are at singing, the more attuned your ear becomes to slight variations in pitch, tempo, sound shape, and the like which you can use to judge a performance. These tools are

necessary to become a good performer, because you have to be able to discern subtle details to improve the subtlety of your performance. But they also make apparent perfection recede into the distance. A non-singer might not be able to point out any flaws in a great singer's performance, but it is certain that the performing artist has things about that performance that she wishes she had done differently or better.

In general, virtuoso artistic performance is not only a matter of achieving a certain level of accomplishment in that art. Being an excellent artist is a practice, an activity (as Aristotle reminds us a virtue is) that is a constant thread throughout the performer's life. An excellent artist is always getting better or getting worse. Likewise with virtue: "virtue is essentially dynamic, not a static condition of the person but an aspect of him or her that is always developing for the better or the worse... Our work, in living virtuously, is never done" (Annas 2011, 163-164). This feature of artistic performance and virtue helps explain achievement relativity: the master's heightened sensitivity to excellence, and constant activity in pursuit of excellence, makes an appreciation of how the master can improve itself an intrinsic part of mastery. Therefore it seems plausible to expect that the master and non-master will have different views of the master's accomplishments.

To pick an example from a completely different skill area, the training for a Ph.D. in science also involves this sort of achievement relativity. In her dissertation defense, for example, my wife ended with a slide that contained a graph of "What I think I know" over time. As she proceeded from her first to her fifth year in graduate school, her knowledge went from "Frankly, an impressive amount" to "squat". The

slide itself does not establish my point, but the reaction the slide received from the Oceanography department at UC Santa Cruz was compelling: universal laughter and one of the faculty raising his hand and saying roughly “you cut it off; the graph keeps doing down as you get farther beyond your Ph.D.” That is, it seemed an accepted truism in her department that a part of training to be a professional oceanographer was learning how little about oceanography you actually know. Many other scientific fields share this dynamic. For example, take the contrast between proliferating uses of quantum mechanics in popular books and the famous quote from Richard Feynman, the famous physicist, that “I think I can safely say that nobody understands quantum mechanics” (Feynman 1964). As one becomes more expert in a scientific field, and thus as all would evaluate their knowledge as higher than the amount most people possess, one also appears to become more aware of how much scientific knowledge they lack. Judgments of scientific expertise are achievement relative.

It is likely the same, or even more pronounced, with virtue. If virtue is anything like a skill, it is a skill with as wide a range of subtle details, contextual clues, and sensitivities as any other practical skill. Most philosophers probably already have a sense of the complexity of moral decision-making. Our sense comes not from being any better with practical morality, but when you spend time reflecting seriously on moral concepts and judgments, you begin to notice characteristic details and arguments that do not occur to others. I have had the experience many times of influencing someone’s thinking because I could notice and highlight details that we both agreed were morally relevant. This sensitivity to the contours of moral decision-

making has been trained, and it is a sensitivity that is certainly a part of achieving a high level of virtue.

In fact, there is strong anecdotal evidence that people of extraordinary virtue do exhibit achievement relativity. Vanessa Carbonell catalogs some of these people (Carbonell 2012, 229-231). While the force of Carbonell's article is directed at the influence of moral saints on everyday moral obligation, she describes in detail cases of achievement relativity. Paul Farmer, one of her examples, is a doctor who founded Partners in Health, which he has grown into a worldwide organization "with an annual budget in the tens of millions of dollars." Farmer has made extraordinary personal sacrifices:

"Farmer himself treats countless patients directly, sometimes hiking for hours to make house calls. In the early years of the organization, Farmer alternated between living in a hut next to his clinic in Haiti and sleeping in the basement of his office in Boston. He had his entire salary deposited directly into the organization's budget, leaving the office staff to make sure his modest bills were paid. He didn't buy new clothes or take vacations, and he went for long periods without seeing his wife and daughter" (Carbonell 2012, 229).

Despite this list of sacrifices, Farmer "constantly feels that he should be doing more." In his own words, "I don't care how often people say, 'You're a saint.' It's not that I mind it. It's that it's inaccurate... people call me a saint and I think, I have to work harder. Because a saint would be a great thing to be" (Carbonell 2012, 230). Similar descriptions are "ubiquitous in the literature on Holocaust rescuers" (Carbonell 2012, 230). Carbonell claims "part of what is going on in these cases is that the moral saints and the observers have different perspectives on how much sacrifice the actions involve" (Carbonell 2012, 248). These individuals have a sense that their own

sacrifices in pursuit of moral ends are not as great as they are perceived to be by others, thus evaluations of their moral exemplariness are achievement relative.

Achievement relativity provides a direct response to one argument from Statman, and understanding this response may help explain achievement relativity.

Statman says,

“The fact that one can always be a better person in one or another respect does not rule out the possibility that one’s achievements might be truly impressive. Suppose human worth could be measured on a quantitative scale. And suppose one has gained 90 points out of 100. Why do we consider that the right perspective is to stress the ten missing points which hardly anyone has ever gained, and not the 90 which one actually gained while the human average is only 60?” (Statman 1992, 427)

Statman uses this line of argument against authors like Richards and Flanagan, challenging that in extraordinary cases, non-overestimation could still involve great esteem that does not look much like humility. But on the account given here, it is a misconception of mastery to think that there is a 100-point scale that all observers agree upon and against which all achievements are judged. The achievement relativity argument implies that while most people think they achieve 60 out of 100, masters believe most people achieve 60 out of 150. That is, they perceive a greater range of achievement than other people. To use numbers parallel to Statman’s, a master can accurately judge themselves as achieving 90 out of 150 while being perceived by others as achieving a 90 out of 100. The master is not making a mountain out of a molehill, so to speak; the master perceives that the molehill is in fact a mountain.

It is important to be clear that achievement relativity is not equivalent to or sufficient for humility. Achievement relativity only makes it possible to satisfy all

three desiderata in the puzzle outlined above. In practice, achievement relativity usually causes more pride and arrogance than humility. The heightened perception and judgment of the master can easily be used to demean and insult the less skilled. This seems a more likely result than turning such subtle perception on oneself, a requirement of humility.

This argument clears the ground for a humility that satisfies all three desiderata, and it provides a better ground for humility as a virtue than the religious, or *sub specie aeternitatis*, arguments referenced above. Humility before a divine, perfect, or transcendent being is an odd thing to consider a human excellence in the first place. It is hardly excellent for a human to admit imperfection measured in light of the divine, or in light of eternity; that seems rather a basic requirement to be taken seriously as a moral thinker. This being said, the achievement relativity account has a kinship with the religious defense of humility because both posit a scale upon which humans generally cannot achieve perfection. For the religious defense, it is strictly impossible to reach the top of the scale because it extends to divine virtue. For the achievement relativity account, it is generally practically impossible to reach the top of the scale because it is so difficult to have the scale clearly in view. The act of reaching higher levels of achievement upon the scale reveals further heights to which one can aspire, and leads one to continually re-evaluate how difficult it is to achieve virtue. It is this continually clearer perception of the difficulty of perfection that makes sensible the humble person's low self-evaluation, and reconciles that low self-evaluation with the other desiderata of humility as a virtue.

4.2 Humility vs. Modesty

“You don’t want modesty, you want humility. Humility comes from inside out. It says someone was here before me and I’m here because I’ve been paid for. I have something to do and I will do that because I’m paying for someone else who has yet to come.” - Dr. Maya Angelou (Reiss 2014)

In this paper so far, the terms modesty and humility have been used interchangeably. This follows most of the literature where authors discussing modesty (Driver 1989; Schueler 1997) engage with authors discussing humility (Richards 1988; Hare 1996; Milligan 2007; but See Ben-Ze’ev 1993, 240). But the closest argument to my achievement relativity argument in the previous section comes from the literature discussing modesty, and it is worth showing the difference between my argument and that argument, and how the differences show an important distinction between modesty and humility. This distinction will serve to show why the two standards account is unsatisfying as an account of an Aristotelian virtue.

In the modesty literature, some authors have endorsed a “two standards” account (Brennan 2007; Smith 2008; McMullin 2010). These accounts claim that modest people apply different standards to their own conduct than they apply to other people’s conduct, and that the standard they apply to themselves, the “ideal” (Brennan) or “aspirational” (Smith) standard, is more demanding. In this way, a modest person could rank themselves low on one standard while also knowing that they rank high on the public, general standard. As Smith says, “one who is capable of such levels of excellence, after all, has every reason also to apply extraordinary aspirational standards to his own activities” (Smith 2008, 31). As regards the conduct

of others, however, the modest person on this account applies the general standard to other people (Brennan 2007, 121).

The achievement relativity account differs from this two standards account in two important ways. First, it construes the self-evaluation of the humble person differently. On the achievement relativity view, what motivates a master to be humble is not a rational choice to apply a set of heightened standards, but rather the heightened perception gained through mastery. The person aspiring to virtue is always interested in an accurate, authentic self-evaluation, they simply learn more about what that involves as they gain mastery. The two standards account of modesty has the modest person choosing an aspirational standard for their own conduct because of their accomplishment. This aspirational standard does not involve any special knowledge or judgment on their part, just the decision to hold themselves to a higher standard that most people find unreasonably exacting. This two standards view is an unrealistic account of an Aristotelian virtue. It is telling that Smith gives examples of aspirational standards for common, goal-oriented activities such as playing tennis without double-faulting (Smith 2008, 32) and playing chess well (Smith 2008, 31). These activities are games with relatively unambiguous success criteria, but this simplicity makes them poor analogies for virtue. Virtue is not simply a matter of adhering to widely-known standards better than others, though in some contexts and cases this may be a part of virtuous activity.

Second, and perhaps more importantly, the two standards account of modesty does not capture the virtue of other-evaluation. The achievement relativity account

given above deliberately takes no stance on how virtuous people evaluate others while the two standards account claims that modest people evaluate others on the general, common, non-idealistic standards (Brennen 2007, 121). I do not think that this sort of modesty is particularly virtuous. It seems more, perhaps, a matter of social etiquette, or even less flatteringly a reification of modern, Western social norms. But why could a master not chastise others for failing to live up to excellent standards, and do so with humility? The Zen tradition, for example, has famous stories of students being forced to stay on the threshold of a monastery for days, weeks, or longer in extreme conditions to show the seriousness of their commitment to study. The humble yet harsh, demanding sushi chef in the splendid documentary Jiro Dreams of Sushi is another example.

This point is different from the claim that virtuous people ought to, and of course do, reserve judgment about others in many cases because they have incomplete information. This seems well-justified. Common sense and life experience suggests that we often do not know enough about others to judge the excellence of their conduct. But this is different from the claim of the two standards account of modesty, which is that even when complete information is available, a modest person applies a less demanding standard to others. This application of a less demanding standard seems to require a sort of dishonesty from a modest person, and this view is subject to the same sorts of problems outlined in the puzzle above, because it requires the virtuous person to be deceptive. It is also unclear how to reconcile this form of

modesty with the virtue of justice, which involves proper, fair judgments about other people's accomplishments, blameworthiness, and praiseworthiness.

The two standards account is particularly problematic for certain positions in our society. If a teacher, for example, were to embody this version of modesty, it would be difficult for her to demand excellence of her students. Teachers should certainly be compassionate, kind, and fair, but in many contexts good teachers also ought to be strict and demanding, to ask more of her students than they think they can give, and to judge them by a very high standard which they are trying to learn how to meet. The achievement relativity account makes sense of a humble, demanding teacher in a way that the two standards account of modesty cannot accommodate.

This distinction between the achievement relativity and two standards accounts illuminates a difference between the terms humility and modesty that makes humility the better term for the virtue at issue here. As will become clear in the next section's discussion of empirical evidence, the relevant virtue for this chapter is the one of self-evaluation. The puzzle in the prior section arises from, and has force in relation to, judgments of self-evaluation. And from the arguments and evidence in this chapter, it is not even clear that there is a virtue associated with other-evaluation, aside from the classical notion of fairness embodied in the virtue of justice. All of these reasons weigh in favor of the term humility and the classic self-regarding content of that virtue. This also allows us to avoid the morally problematic notions of other-evaluation in the term modesty, not to mention the sexual and sexist

connotations of the word “modesty” that have nothing whatsoever to do with Aristotelian virtue.

4.3 Empirical Humility: Motivated Reasoning and Cognitive Dissonance

Modern psychology provides a wealth of evidence supporting the claim that humility is a virtue: it is better for us, better for those around us, and corrects for a characteristic human flaw. This evidence suggests that humility is a virtue in the classic Aristotelian sense, despite Aristotle’s own apparently contrary view. Combined with the theoretical arguments in the prior two sections, this evidence shows that we ought to consider humility an important virtue.

The humble disposition, like all virtuous dispositions, involves a diverse set of traits. It is very difficult to list a set of necessary and sufficient conditions for any virtue, humility included. In fact, it shows a misunderstanding of virtue ethics to seek such a set of conditions, for virtues are not things that are well-captured by bright-line rules (NE 1094b12-23). Rather than strive for a complete definition of humility, this section focuses on a basic sub-set of humility’s qualities. To have humility, a person must at least:

- A. not lose sight of the possibility that their own judgments, including their moral judgments, are mistaken;
- B. appreciate the role of luck in their lives and achievements; and
- C. appreciate that in many cases, respectful engagement with those who have different opinions will help them make better judgments.

Not only must a humble person have these qualities, these are qualities not commonly associated with other virtue concepts. That is, they are important parts of humility and

they are most appropriately understood as parts of humility, rather than parts of some other virtue. The remainder of this section provides empirical evidence supporting the claim that these three traits are a part of human virtue.

The sort of empirical argument given here is briefly suggested in some of the literature (Flanagan 1990, 426; Schueler 1997, 484-485), but to my knowledge, no author has actually taken the ball and run with it, so to speak. Both Flanagan and Schueler make brief gestures to general features of this sort of argument, but the intention here is to be much more specific and detailed than they are and thereby hopefully show that their intuitions are sound. In particular, it is not obvious how we ought to approach empirical evidence for or against a particular virtue trait, as discussed in the introduction. This section takes a broad stance, arguing that if there is empirical evidence that a disposition satisfies three classic hallmarks of virtue, then this evidence constitutes *prima facie* evidence the trait in question is a virtue. These classic hallmarks are:

1. Virtues are good for their possessor. They are constitutive of their possessor's *eudaimonia*, happiness or flourishing.
2. Virtues are good for others. The virtuous are good for and to those they are around; they are the best friends, the best fellow citizens, the best family members.
3. Virtues function as correctives for characteristic problems that humans have to deal with to live good lives (Foot 1978, 9).

These three hallmarks are general, broadly accepted qualities that virtues must have, and they capture the central concerns of an Aristotelian virtue ethics.

4.3.1 Humility Benefits Its Possessor

Humility potentially benefits its possessor in many ways. This section focuses on the psychological evidence regarding cognitive dissonance as an illustrative case. Cognitive dissonance itself encompasses a large area of research, and so as has been the pattern in this thesis, I will particularly focus on the work of two prominent scholars in the field, Carol Tavis and Elliot Aronson, and their popular book Mistakes Were Made (But Not By Me) (Tavis and Aronson 2007), with reference to other studies where appropriate.

Cognitive dissonance “is a state of tension that occurs whenever a person holds two cognitions (ideas, attitudes, beliefs, opinions) that are psychologically inconsistent, such as ‘Smoking is a dumb thing to do because it could kill me’ and ‘I smoke two packs a day’” (Tavis and Aronson 2007, 13). This dissonant feeling seems universal: basically all humans feel this tension when holding conflicting cognitions. The troubling part of cognitive dissonance is not that the feeling exists, but rather how it serves as “the engine that drives self-justification” (Tavis and Aronson 2007, 13). When someone holds the two conflicting cognitions about smoking, to use Tavis and Aronson’s example, they must give up on one of them in order to resolve the dissonance. They could stop smoking, but they usually don’t. Dissonance works as an engine of self-justification when people choose to give up on the belief that they do dumb things by generating a justifying cognition, such as “but I haven’t smoked for long, so I’m not in danger. I’ll stop before I hurt myself.” The problem is not so much that this cognition is necessarily false, but that the motivation

for it stems from a simple desire to end the dissonant state, not because one has good reason to change one's beliefs. The belief change in this case is not based on evidence, but only on a desire to have a different belief.

The experience of dissonance is a part of the human condition; it is impossible to go through life without conflicting cognitions, and we encounter and resolve them every day. The problem with cognitive dissonance is that we are motivated to resolve the dissonance simply in order to avoid the feeling of cognitive dissonance, even when we have no good reason to resolve it. One common example encountered on college campuses, particularly among tech-savvy students, is that of regularly downloading copyrighted works without permission. Most students who admit to doing this are not planning to break into concerts without paying, or steal CDs from a music store, so on some level they believe that artists should be paid for their work. This generates competing cognitions such as "not paying artists is stealing, and it is bad to steal from artists" and "I am a good person". When confronted, most students have some argument that taking a copyrighted work without permission is not really stealing. And this resolution is startlingly strong—when pushed with arguments showing their lack of good reason for taking copyrighted works, very many students simply cannot bring themselves to see the copyright infringement that they engage in as wrong, even when they accept an analogously valid argument against other forms of intellectual property theft.

Given the data on cognitive dissonance, how would the virtue of humility help address this issue? Humility is not a character trait that changes our fundamental

cognitive machinery to make us no longer experience dissonance. But humility does involve the view that one's own ability to make correct judgments is fallible. A humble person, therefore, could experience less of the discomfort of cognitive dissonance because they understand how far they have to go to reach perfection. For example, rather than being strongly linked to self-concepts like "I am a good person" or "I am a just person", a humble agent would be committed to self-concepts like "I am trying to be a good person" and "I am trying to act justly". Factor A of the humble disposition, always keeping in mind the possibility that one's judgments of all sorts could be wrong, implies that humble people aspire to be good or just but are not strongly committed to the ideas that they are already good or just. These self-concepts of aspiration leave room for the idea that one is sometimes wrong, and therefore these cognitions no longer contradict cognitions that one is acting badly or wrongly in a particular case. Someone who is "trying to be good" can do wrong without dissonance.

One might argue that this only shifts the dissonant cognitions from those about one's static character traits, such as "I am a good person", to those about one's goals or desires, such as "I am trying to be a good person". On the description above, a humble person could experience dissonance if they have the cognition that they are not actually aspiring to be good after all. For example, if a humble person believes they are trying to be temperate, but they eat a box of donuts every day for weeks or months, at a certain point they could experience dissonance because they believe that "someone who eats lots of donuts every day is not trying to be temperate". I could say

that the humble person has to be humble even about their own desires and ends, such that they could be open to the idea that they are wrong about their own desires and ends. This move suggests an infinite regress, or dialectic, whereby dissonance can always be created between the humble person's judgment about their own cognitions and some evidence that this judgment is wrong. I am inclined here to say two things: first, it is true, and shows that the force of cognitive dissonance is basically inescapable. But second, the dynamic described above of the humble agent undermines the self-justifying power of dissonant thinking, where the agent is motivated to change beliefs purely to avoid the feeling of dissonance. This is because simple beliefs about being a good or bad person are put at risk in very many ways in our everyday lives. Beliefs about our goals and aspirations are beliefs about a course of conduct, or our intentions in acting, and these are put at risk in far fewer situations, or over long time periods, than our simple beliefs about goodness and badness. For each step in the infinite regress, the beliefs at stake become more abstract and much less likely to be challenged by a competing cognition in ordinary life. Thus, while cognitive dissonance seems inescapable, the humble disposition can drastically reduce the power of dissonance to motivate us toward self-serving, bad reasoning.

It is important to keep in mind that cognitive dissonance seems inescapable, as was vividly illustrated in a recent study on how post-choice preferences can influence evaluative measures even when an agent forgets the initial choice (Coppin et. al. 2010). The experimenters asked test subjects to rate the pleasantness of a battery of smells. Based on those ratings, they then asked test subjects to choose their favorite

smell from pairs that each particular test subject rated as similarly pleasant. About ten minutes after rating the pairs, the subjects were again asked to rank the pleasantness of the odors, and they were also asked which odors they smelled earlier in the paired portion of the study. The results showed that for the odors each test subject considered in pairs, they raised the pleasantness rating of the odor they picked and lowered the pleasantness rating of the odor they did not pick, as predicted by classic cognitive dissonance theory. However, the experiment also showed that this effect existed even when the test subject did not remember that a particular smell was one of the pairs they were given. This indicates that the classic cognitive dissonance explanation for this effect could only work if dissonance could work implicitly, which is unclear (Coppin et. al. 2010, 492).

For purposes of the argument for humility, however, what matters is that the subjective evaluations changed. Cognitive dissonance in this study is not just a matter of competing cognitions that could be the subject of conscious reflection; the study suggests that it also functions with non-conscious, or automatic, judgments of pleasantness. This suggests that humility does not only involve an attitude toward one's own reflective judgments, it must also include a humble attitude toward one's own brute preferences and desires. And at this level, humility can help improve our moral attitudes toward others. This is because "being wrong" in regard to our brute preferences is different from being wrong in our reflective judgments. It is odd to say one is "wrong" in a purely aesthetic judgment about the pleasantness of a smell, or the taste of ice cream, and the like. Being wrong here is not about there being some

better aesthetic judgment, but rather about being open to the idea that one's own preferences are not inherent, unchanged, necessarily *sui generis* parts of oneself, but that they could come from an unknown, or external, source (such as experimental psychologists). This part of humility is critical for progress on problems such as entrenched racism and sexism, for example.

4.3.2 Humility Benefits Others

Much of the evidence described in the previous section also has force for this facet of virtue. This section will focus on a particularly terrifying case of cognitive dissonance detailed in Tavis and Aronson's book: police interview techniques. As Tavis and Aronson show, the widespread misunderstanding of cognitive dissonance, and police interrogators' over-estimation of their own prowess, leads to a Kafkaesque system of interrogation, confession, and proof. While this discussion is limited to the particular context of police and suspected criminals, its strength illustrates the general potential benefits of humility for other people.

Tavis and Aronson provide many vivid examples of the harms of botched confessions. Here is one harrowing, famous case:

“On the night of April 19, 1989, the woman who came to be known as the Central Park Jogger was brutally raped and bludgeoned. The police quickly arrested five black and Hispanic teenagers from Harlem who had been in the park ‘wilding,’ randomly attacking and roughing up passersby. The police, not unreasonably, saw them as likely suspects for the attack on the jogger. They kept the teenagers in custody and interrogated them intensively for fourteen to thirty hours. The boys, ages fourteen to sixteen, finally confessed to the crime, but they did more than admit guilt: They reported lurid details of what they had done. One boy demonstrated how he had pulled off the jogger's pants. One told how her shirt was cut off with a knife, and how one of the gang repeatedly struck her head with a rock. Another expressed remorse for his ‘first rape,’ saying he had felt pressured by the other guys to do it, and

promising he would never do it again. Although there was no physical evidence linking the teenagers to the crime—no matching semen, blood, or DNA—their confessions persuaded the police, the jury, forensic experts, and the public that the perpetrators had been caught. Donald Trump spent \$80,000 on newspaper ads calling for them to get the death penalty.

And yet the teenagers were innocent. Thirteen years later, a felon named Matias Reyes, in prison for three rape-robberies and one rape-murder, admitted that he, and he alone, had committed the crime. He revealed details that no one else knew, and his DNA matched the DNA taken from semen found in the victim and on her sock. The Manhattan District Attorney's office, headed by Robert M. Morgenthau, investigated for nearly a year and could find no connection between Reyes and the boys who had been convicted. The DA's office supported the defense motion to vacate the boys' convictions, and in 2002 the motion was granted. But Morgenthau's decision was angrily denounced by former prosecutors in his office and by the police officers who had been involved in the original investigation, who refused to believe that the boys were innocent. After all, they had confessed." (Tavris and Aronson, 128-129).⁶⁴

This case has simple force for this section: if the police and prosecutors involved in prosecuting the "Central Park five" had more humility, they could have potentially avoided causing gross injury to innocents.

First and foremost, if the police and prosecutors had been more aware of the possibility that their judgments were false, they may not have pushed the five defendants so hard to confess, given the lack of physical evidence. While arresting the five initially made sense, given their physical proximity to the crime, the early push to confession in the absence of physical evidence seems unjustified. But it turns out that this practice reflects some basic police training practices, the Reid interrogation technique. Unfortunately, the Reid technique seems to be based on many improper and incorrect beliefs about human psychology, such as that innocent

⁶⁴ This case is also detailed in a haunting documentary, "The Central Park Five".

people never confess, and that nervousness, sweating, and avoiding eye contact are signs of lying. These claims are empirically false, and lead some naive, innocent subjects interrogated under the Reid technique to feel trapped. In Tavis and Aronson's words,

"The Reid technique is thus a closed loop: How do I know a suspect is guilty? Because he's nervous and sweating (or too controlled) and because he won't look me in the eye (and I wouldn't let him if he tried). So my partners and I interrogate him for twelve hours using the Reid Technique, and he confesses. Therefore, because innocent people never confess, his confession confirms my belief that his being nervous and sweating (or too controlled), or looking me in the eye (or not) is a sign of guilt. By the logic of this system, the only error the detective can make is failing to get a confession" (Tavis and Aronson , 144).⁶⁵

A confession retroactively justifies all of the inferences of guilt the interrogator makes, and so the pressure on the interrogator to get a confession is powerful, and it grows with every inference made in favor of guilt. The confession becomes not just a statement about a crime, but about the interrogator's own competence and rationality, and a non-confession conflicts with a wide variety of the interrogator's self-conceptions, including their belief that they have the right person, that they are competent investigators, and that it is wrong to berate innocent people. An awareness of their own potential fallibility could undermine this engine of self-justification in the minds of the interrogator, and thereby ameliorate the pressure from cognitive

⁶⁵ Reid and Associates have an extensive online document claiming to rebut the charges made by Tavis and Aronson, and others (available at <http://www.cbc.ca/thenational/includes/pdf/reidresponse.pdf>). I will not discuss this document in detail, except to note that the focus of the document is not on the moral or psychological case mounted against the Reid Technique, but on the way that courts have accepted confessions generated with the Reid Technique. The fact that courts have accepted Reid Technique confessions is at best *evidence* of its legality, and it is not relevant to its morality or empirical adequacy.

dissonance that leads them to push for confessions far beyond a reasonable point.⁶⁶

That is, humility about their own judgments, investigative practices, and inferential reasoning skills would help police interrogators be better to others, and better at their jobs.

Second, the police and prosecutors failed to appreciate the role of luck in their arrest of the Central Park five. While it may have been reasonable for the police to initially suspect the teens in the attack, reasonableness turned into certainty far too quickly. We now know that the five teens happened to be “wilding” in the park on the same night that Matias Reyes was committing his attack on the Central Park Jogger. The police, however, seem not to have considered the likelihood that two unrelated crimes would happen on the same night in Central Park. At the least, this sentiment is reflected in the prosecution’s statement to the court in favor of vacating the convictions after the evidence of Reyes’ guilt was uncovered, where they said “[c]ertainly, no one would have thought that as the defendants and their group were making their way through Central Park, a serial rapist was also at large” (Ryan 2002, 51). But this belief is unjustified, and underestimates the function of luck in these

⁶⁶ One might object that actual cases of incarceration based on false confession are rare, and that we know about the Central Park 5 because it was such an unusual and disturbing case of innocents being incarcerated. But I think we have little basis for such comfortable assumptions. First, it is fair to note that the Reid technique would not lead to most of us, or even many of us, to confess to a crime we did not commit. But if it would lead in a predictable way to some false confessions, and it does seem to do so, that is enough for the argument for humility to work. Second, and perhaps more importantly, this objection fails to appreciate the way that confessions work coercively within the context of our plea bargaining system. The overwhelming majority of criminal cases are disposed of through plea deals, where any confession is not adjudicated in the courtroom. And the pressure to accept a plea, given that the prosecution has your confession on tape, must be immense. Part of the reason that cases like the Central Park 5 are so rare is because most confessions are not vetted through a trial process, rather defendants agree to lesser sentences in a plea bargaining process.

sorts of cases. To take an example, there were 368 serious crimes committed in the Central Park precinct in 1990.⁶⁷ It is impossible, therefore, for there not to be at least two serious crimes happening in Central Park on three days during that year; it is extremely likely that there was far more clustering than this. Without getting into the probability arguments for this, consider that for every day that year that a serious crime did not happen in Central Park, another day must have had an extra crime. Since it is plausible that there were many days without serious crimes in Central Park, it must be the case that having more than one serious crime occur on a single day in Central Park is actually a common, rather than unusual, occurrence. Given these sorts of statistics, the inference that the police made from the coincidence of the five youths “wilding” on the same night as the attack on the Central Park Jogger should not have been very strong, particularly given the different details of each crime, the lack of physical evidence, and the inability of the Central Park Five to describe details of the attack on the Jogger.

Perhaps the most compelling evidence of the importance of humility, and the power of pride, is the fact that some police and prosecutors still refuse to admit that they erred in prosecuting the Central Park Five (Drizin and Leo 2005 , 899 n.38). Even with DNA evidence affirmatively showing that Matias Reyes was present at the rape, a year of investigation failing to show any connection whatsoever between Reyes and the Central Park Five, and a district attorney’s office who agreed with the defendants and filed a statement in favor of vacating their sentences, some of the

⁶⁷ This data available at http://www.nyc.gov/html/nypd/downloads/pdf/crime_statistics/cs022pct.pdf.

original investigators seem unable to come to terms with the overwhelming likelihood that they erred. This is particularly galling in light of the “beyond a reasonable doubt” standard of evidence, for no rational person could look at the state of evidence in this case and conclude that the Central Park Five committed this crime beyond a reasonable doubt.

One might object that humility is not necessarily the proper curative for these problematic police investigations. Eliot and Aronson, for example, focus on the initial framing of the case and the way that investigators move down a “decision pyramid”, where the initial decision to see a case in a certain way leads to a series of self-justifying inferences that reinforce that initial choice. This seems to be a problem of investigative technique. But this framing problem is in tension with the problem of base rates. Given that any investigative technique is imperfect, and necessarily only can increase the degree of belief in certain facts about a case, it is very important that the police only use these techniques on suspects whom they have good reason to believe could be guilty. That is, if they were to apply their investigative techniques without making a judgment about the prior probability of guilt, the general problem of false positives for low base rate situations would loom large.⁶⁸ This means that the police ought not to give up on their initial judgments about a case too quickly, at least as long as those initial judgments reflect their best reasoning about the facts of the case. This tension between not putting too much faith in initial judgments, but also

⁶⁸ This is why, for example, police ought not to set up a DUI checkpoint and give breathalyzers, or field intoxication tests, to every driver. The false positive rate of such a practice is so high that the police are likely to find more sober people who fail the test than intoxicated people.

letting those judgments guide the use of investigative techniques, is exactly the sort of situation for which a virtue is a good cure. Humility here could be an intermediate state between putting too much confidence in one's own judgments and putting too little confidence in those judgments.

It is tempting to damn the police and prosecutors who appear in this passage of Tavis and Aronson's book, but I think that the best view is that these people are in a position, in a career, in which it is exceptionally important and also exceptionally difficult to be humble. I have no doubt they failed in a profound, morally horrifying way in their prosecution of the Central Park Five. But the problem is not that some innocent people were convicted: this seems inevitable in a necessarily imperfect justice system. The problem is that the police and prosecutors deny the possibility that they were imperfect. It turns out to be the case that humility is a critical virtue for police investigators because of the circumstances of their job, in a manner very similar to how courage is critical for the soldier.

4.3.3 Humility Corrects for a Characteristic Human Weakness

The Coppin study is good evidence that cognitive dissonance is a characteristic human problem that will always confront us. If dissonance functions even on the level of unconscious, unremembered smells, it is difficult to see a theoretical account of cognitive dissonance that is not a characteristic human flaw. Other recent work, by Daniel Kahneman, is another powerful illustration of how widespread and fundamental the problems are for which humility corrects. Just as

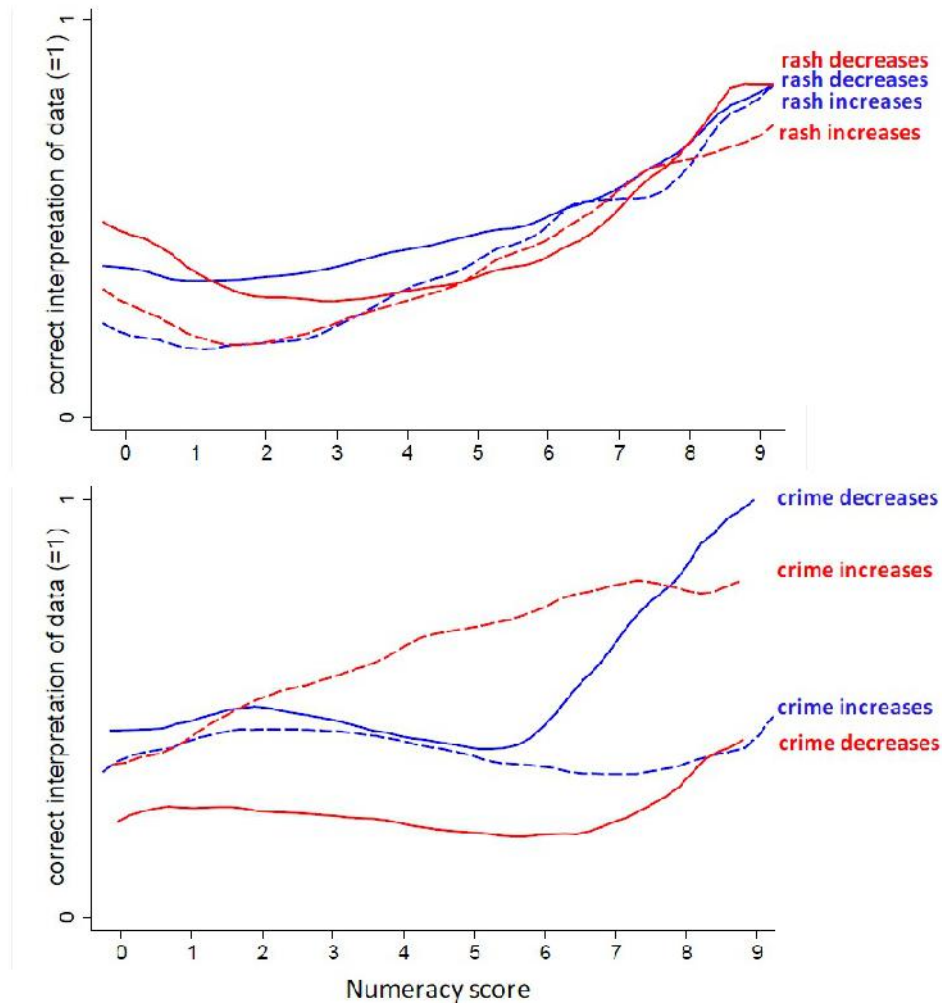
fear necessitates courage, cognitive dissonance and motivated reasoning necessitate humility.

Kahan's paper shows that on average, our basic mathematical skills are in some ways dependent on our political views (Kahan et. al. 2013). It received popular media attention when first released, for good reason (see, e.g., Mooney 2013). In this study, Kahan first scored participants on a numeracy scale, a psychological measure of the subjects' comfort using mathematical reasoning to solve problems. He also asked participants to self-identify as conservative or liberal. Then, he gave study participants one of four cases, two normatively neutral and two normatively loaded. Each of these cases was mathematically identical, and involved solving a reasonably difficult math problem about judging a correlation between some intervention and its effect on a problem. The normatively neutral cases concerned the effectiveness of a skin cream in treating a rash. In one case, the correct answer to the problem was that the cream was effective at treating the rash, in the other that the cream made the rash worse. The normatively loaded cases concerned the effectiveness of gun control policies on crime. One case had gun control laws decrease crime, the other had gun control laws increase crime.

Kahan's results are summarized in the following figure:

Figure 2

"Liberal Democrats" (< 0 on Conserv_Repub) "Conservative Republicans" (> 0 on Conserv_Repub)



The top graph shows the results for the normatively neutral cases, the bottom graph for the normatively loaded cases. The top graph is fairly simple: as expected, the higher the test subjects' numeracy score, the more likely they will get the correct answer, regardless of the case. The bottom graph diverges from the top impressively: there is a marked decrease in the abilities of high numeracy test subjects to correctly answer problems that are normatively loaded against their own presumed beliefs. In

the case where gun control laws decrease crime, self-identified republicans of a high numeracy score are not even as accurate as the lowest numeracy self-identified democrats. High numeracy self-identified democrats are not better in the case where gun control laws increase crime. In each case, a lower numeracy score person of opposite political beliefs is much more likely to answer correctly.

Kahan speculates that a partial explanation for this data is that numerate people are motivated to look for the right answer when their intuitive response goes against their political beliefs. He argues that high numeracy agents were only activating their System 2 reasoning when the System 1 answer challenged their political beliefs (Kahan 2013, 26-27).⁶⁹ This suggests that high numeracy individuals, were they somehow on guard against this sort of mistake, might be able to avoid it. For example, in a normatively loaded context the test taker could simply copy down the numbers and work the problem without the normative content. This is a powerful illustration of the importance of humility because it is not the case that someone who got the problem wrong would be unable to understand the correct answer. In fact, it is far more plausible that those who scored highly on the numeracy scale, but got the problem wrong, would readily see their error once it was pointed out. The problem is, after all, a case of making a mathematical computation correctly; there is no reason to think that these high numeracy people cannot understand the problem. If these test subjects were open to correction, if they were generally humble, then that correction

⁶⁹ If Kahan is right, this study is also important for my earlier discussion of mindfulness, where I suggest that the virtue of mindfulness involves, at least in part, the conscious activation of System 2 in morally important contexts.

would not be hard to make. Of course, not all of our biases are as easily shown to be wrong as a math mistake, but on the other hand, it is humbling in itself to realize that our dry, mathematical ability is subject to effects of this sort.

Kahan's and Coppin's studies are examples of the ways in which automatic, System 1 thinking is governed by cognitive dissonance and motivated reasoning. In fact, System 1 gets things wrong in many other ways as well, such as in the cases of implicit bias discussed above in Chapter 3. In fact, there is a sense in which the entirety of the empirical evidence in this dissertation is evidence that humility corrects for a characteristic human flaw, because each way in which our cognitive system characteristically errs is another context in which we ought to be humble about our own judgments and thought processes.

4.4 Conclusion

“[M]ost of the conscious thinking of a philosopher is secretly guided and forced into certain channels by his instincts” (Nietzsche, BGE, Part 1, 3).

The arguments in this chapter suggest that not only is humility a virtue, it is a cardinal virtue, so to speak. Even more than mindfulness, humility is the virtue demanded by modern research in social psychology. For while the interactions between automatic and controlled processes, the realm of mindfulness, are important, the overwhelming brunt of the research points at the systematic fallibility of all of our cognitive processes. It is an arrogant fantasy to think that any of us are not regularly misled by both our automatic thinking and our controlled reflections. And once we give up on the notion that we are capable of thinking perfectly, humility becomes a fundamental part of a well and honestly lived life. The Christian etymology of

humility is exactly right: man, created from the soil, cannot achieve divine perfection. This is no less meaningful in the absence of a divine metaphysics, for we are embodied, evolved beings made from, quite literally, dirt, and we do not always function as well as we like to think.

I would like to finish this chapter by discussing briefly some common moral intuitions that the account of humility given here can make sense of, for it is a strength of the account that it can explain and respond to some particularly troubling dynamics in modern society. We regularly hear complaints, and make them ourselves, about how much partisanship dominates the public conversation in our country. And it is a lack of humility, rather than a lack of education, or good intentions, that best explains the worst examples of public partisanship (though surely many such people lack other virtues as well). Humility is the virtue that we ought to cultivate and explicitly value in order to fight this public culture (Cf. Button 2005). There are a few reasons for this.

First, ascribing a lack of humility to someone explains their conduct in a way that does not ascribe evil intentions to them. For example, there is a strong vein of homophobic sentiment in our public culture in the United States. In California, recent debates over Proposition 8 are strong evidence of this. It is very easy, and potentially seductive, to impute bad or evil intentions to those who supported Proposition 8. But it is much better, and more accurate, to understand the failure to recognize the rights of same-sex couples as a failure of humility, rather than a general ethical failure.

Generally speaking, Proposition 8 supporters show not malice but a lack of one of the key features of humility highlighted earlier. They might:

- Overestimate the accuracy of their own judgments about what causes sexual orientation, what counts as a moral good, and what counts as good public policy. There are overwhelmingly strong arguments in favor of licensing same-sex marriages. The embarrassing lack of evidence Proposition 8 supporters could provide in the trial court is good reason to think that there is not any good evidence in favor of Proposition 8.⁷⁰
- Underestimate the role of luck. Generally speaking, people who supported Proposition 8 do not believe that sexual orientation is a matter of luck but of choice. They often believe that people with a homosexual orientation could choose to be heterosexual. They do not appreciate the evidence that any of us, as a matter of luck, could have non-normative sexual orientations.
- Fail to engage with opposing arguments. Our public culture very often induces people to exist in an echo chamber of mutually reinforcing rhetoric. People supporting Proposition 8 may fail to appreciate how engaging with the arguments their opponents give is likely to make their reasoning better. Serious engagement with opposing positions is sometimes seen as a threat to morality, not conducive to it.

⁷⁰ Indeed, since that trial a study has come out showing that children raised in same-sex households fare *better* than those raised in opposite-sex households (Crouch et. al. 2012).

These three factors are, I think, quite prevalent in the belief sets of those who supported Proposition 8. Of course, they overlap and incompletely map the territory of opposition to same-sex marriage. But their plausibility helps us those supporting Proposition 8 not as bigots, but as those who lack humility. And lacking a virtue is much different, and elicits more sympathy, than active evil.

The lacking humility explanation also fits evidence about those who oppose same-sex marriage. For example, opposition to same-sex marriage was shown to be most influenced, over the long-term, by who gives arguments rather than by the content of those arguments (LaCour 2014).⁷¹ In particular, even arguments that produce a large shift in opinion in the short term produce little to no long-term effects, unless those arguments are given by a self-identified gay canvasser. Subjects who spoke with a straight ally canvasser reverted to their previous beliefs fairly quickly; subjects who spoke with a gay canvasser had persistent effects even nine months after the conversation. As the study authors point out, these subjects were effectively shifted from the average Georgia voter to the average Massachusetts voter

⁷¹ In the month before submitting this dissertation, this study came under severe question based on a number of unexplained problems with the conduct of one of its authors, Michael LaCour. LaCour is alleged to have committed many wrongs, including fabricating data, lying about the survey company he worked with, and lying about paying study participants. The journal, *Science*, has retracted the publication (<http://www.sciencemag.org/content/early/2015/05/27/science.aac6638>). In his defense, LaCour posted an exceedingly strange non-explanation of his conduct at https://www.dropbox.com/s/zqfcmlkzjuqe807/LaCour_Response_05-29-2015.pdf?dl=0. I have left this paragraph in the dissertation for two reasons. First, rather than ignore this high profile retraction, this example serves as an example of one of my central claims about social science research in the introduction: it should be engaged with cautiously, and single studies that purport to show amazing effects should be treated skeptically. In the case of this study, I used it as an interesting example that fits a theory argued for by different means. The second reason to leave this paragraph in is that LaCour's co-author, Donald P. Green, believes that the effect could be real and wants to do the study again, properly. Much of the work alleged in the study was in fact done, and there is no allegation that Green was party to the fraud (See an interview with Green posted at <http://nymag.com/scienceofus/2015/05/co-author-of-the-faked-study-speaks-out.html>).

(LaCour 2014, 1369). This is fairly strong evidence that opposition to same-sex marriage is not a matter of widespread commitment to a counter-principle that implies such marriage is wrong, which is the sort of position that dominates our public discourse. Rather, it is much more plausible that people learn, upon meeting an openly gay advocate for same-sex marriage, that their beliefs about gay people are simply wrong.

Finally, dialogue about humility is a much more promising avenue for belief change than partisan fighting. This is particularly true for the issue of same-sex marriage, because humility is an important virtue in the Christian tradition, and much opposition to same-sex marriage comes from within certain heavily Christian communities. Humility is a moral virtue that we can share with these communities, and one that can ground a conversation between us about what it means to make judgments about the circumstances of other people's choices and lives.

5. Conclusion

This dissertation has grappled with the general question of what the findings of empirical social psychology mean, or ought to mean, for virtue ethics. In doing so, I criticized the major recent use of social psychology in virtue ethics, stemming from John Doris' *Lack of Character*. I argued that this use of social psychology was improper, and that in fact virtue ethics has much to gain from a more widespread engagement with social psychology. To illustrate the potential benefits of this engagement, I showed how social psychology could motivate virtue ethicists with new arguments for considering mindfulness and humility to be virtues.

The arguments in this dissertation imply some general claims about how virtue ethicists ought to use psychology, and how psychologists ought to think about virtue. First, the study of average human behavior is useful for descriptive virtue ethics, but not normative virtue ethics. Most social science research involves setting up experimental conditions that allow one to look for statistically significant differences between the means of two or more populations of experimental subjects. The idea is that, if such a difference in means exists, then you are very likely to have found some real phenomenon in your study, because the difference in means is unlikely to be due to chance. These studies are important for descriptive virtue ethics, for understanding what sorts of dispositions people actually have. But a particularly important point that too few psychologists seem to understand is: a difference in average behavior says very little about how we ought to be. Virtue ethicists are generally quite happy to admit that virtue is rarely, if ever, instantiated in a human

life. This rarity makes it difficult to use standard social psychological methods to put constraints on virtue.

That said, the second principle argued for is: the study of the psychology of possible behavior is important for normative virtue ethics. Ellen Langer, in particular, has worked to show that one can do psychological research not focused on average behavior, and this method is alive in the positive psychology movement. As Langer says, “in most of psychology, researchers describe what is. Often they do this with great acumen and creativity. But *knowing what is and knowing what can be are not the same thing*” (Langer 2009, 15). Social psychology can be incredibly informative insofar as it helps us understand what is possible for humans, for possibility is what matters when we are considering what a normative theory, like virtue ethics, can demand. The more that social psychologists can engage in rigorous analysis of human possibility, the more fruitful it will be for virtue ethicists to engage with them.

Finally, a third principle to keep in mind is: the psychology of virtue is not analogous to the psychology of vice. This captures one of the central mistakes Doris and the situationists continue to make, as detailed in Chapter 2. But it also infects much psychological work relevant to virtue ethics. Regardless of one’s account of the moral psychology of virtue, the moral psychology of vice is not anything but a lack of the moral psychology of virtue. Vice does not require the consistency of virtue, the harmony of virtue between different faculties, or the habituation or training of virtue. A major reason why the standard methods of social psychology are unhelpful for normative virtue ethics is because they rely on the assumption that there is a

regularity or consistency across the population, but there is no reason to think that the moral psychology of vice has this regularity. I think there is a sense in which the many social psychologists who purport to measure consistently bad actions have in fact measured a wide variety of vicious predilections. This would explain, for example, why Milgram's attempts to permute his famous experiment in myriad ways did not lead to any consistent explanation for why people behaved so deplorably in his initial, famous condition. Empirical moral psychologists would be well-served, I think, to spend more time considering the virtue ethics tradition.

There are a number of future lines of research that could spring from this dissertation. First, there is much work to do to develop accounts of mindfulness and humility as virtues. One strain of further thought might involve examining whether these virtues are intellectual or moral, following Aristotle's categorization. On the intellectual side, these virtues might be understood as facets of *phronesis*, or practical wisdom. But the breadth of empirical evidence surveyed here indicates, I believe, that these virtues involve morally significant dispositions beyond practical reasoning. Accordingly, more in-depth examination of these potential virtues might ground an argument against dividing the virtues into moral and intellectual categories.

These dispositions could also be seen as capturing aspects of virtue have been underemphasized, or ignored altogether, in recent Western, secular moral philosophy. Mindfulness reaches back to the ancient Stoic views, and out to non-Western philosophical views. Humility reaches back to religious traditions in the West, particularly following Aquinas and Christian theology, and out to developing lines of

thought in the psychological sciences about the fallibility of human judgments. Both virtues could generate a wide variety of future work for virtue ethicists.

References

- Adams H., L. Wright, and B. Lohr (1996). "Is Homophobia Associated With Homosexual Arousal?", *Journal of Abnormal Psychology*, Vol. 105, No. 3, 440-445.
- Aquinas, T. Transl. by Father of the English Dominican Province (2008). Summa Theologica. Available free online at <http://www.newadvent.org/summa/>.
- Aristotle. Irwin, T. (transl.) (1999). Aristotle: Nicomachean Ethics, Second Edition. Indianapolis: Hackett Publishing Company.
- Broadie, S. (commentary) and Rowe, C. (transl.) (2002). Aristotle: Nicomachean Ethics. Oxford University Press.
- (1934). Aristotle: The Nicomachean Ethics (Greek Text). Harvard University Press.
- Adams, R. M. (2006). A Theory of Virtue: Excellence in Being for the Good. Oxford University Press.
- Annas, J. (2011). Intelligent Virtue. Oxford University Press.
- "The Unity of the Virtues". Talk given at Stanford University, November 20, 2009.
- (2005). "Comments on John Doris's 'Lack of Character'." *Philosophy and Phenomenological Research*, Vol. 71, No. 3, 636-642.
- (1993). The Morality of Happiness. Oxford University Press.
- Anscombe, G. E. M. "Modern Moral Philosophy". *Philosophy* 33, No. 124 (1958). Reprinted in Crisp (1997), 26-44.
- Asch, S.E. (1955). "Opinions and Social Pressures". *Scientific American*, Vol. 193, No. 5. Reprinted in Aronson, E. (2004) Readings About The Social Animal, Worth Publishing, New York, 17-26.
- Becker, L.C. (1998). A New Stoicism. Princeton University Press.
- Ben-Ze'ev, A. (1993). "The Virtue of Modesty", *American Philosophical Quarterly*, Vol. 30, No. 3, 235-246.

- Biederman I. and Shiffrar, M. (1987). "Sexing Day-Old Chicks: A Case Study and Expert Systems Analysis of a Difficult Perceptual-Learning Task", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 13, No. 4, 640-645.
- Bodhi, Bhikku and Bhikku Nanamoli (1995). The Middle Length Discourses of the Buddha, Wisdom Publications, Boston, MA.
- Broadie, S. (1991). Ethics with Aristotle. Oxford University Press.
- Button, M. (2005). "'A Monkish Kind of Virtue'? For and Against Humility", *Political Theory*. Vol. 33, No. 6, 840-868.
- Carbonell, V. (2012). "The Ratcheting-Up Effect", *Pacific Philosophical Quarterly*, 93, 228-254.
- Connor, S. (2012). "The Bolt Effect: How Chasing Usain Bolt made the world run faster", *The Independent*, March 23, 2012. Available at <http://www.independent.co.uk/news/science/the-bolt-effect-how-chasing-usain-bolt-made-the-world-run-faster-7582603.html>.
- Cooper, J. "The Unity of Virtue." *Social Philosophy and Policy*, 15 (1998): 233-274.
- Coppin, G., S. Delplanque, I. Cayeux, C Porcherot, and D. Sander (2010). "I'm No Longer Torn After Choice: How Explicit Choices Implicitly Shape Preference of Odors", *Psychological Science*, 21:489.
- Crisp, R. and Slote, M. (ed.) (1997). Virtue Ethics. Oxford University Press.
- Crum, A.J. and E.J. Langer (2007). "Mind-set matters: exercise and the placebo effect," *Psychological Science* 18, 2, pp. 165-71.
- Darley, J.M., and C.D. Batson (1973). "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior". *Journal of Personality and Social Psychology* 27: 100-8.
- Dawson N. And H. Arkes (2009). "Implicit Bias Among Physicians", *Journal of General Internal Medicine*, 24(1), 137-140.
- Doris, J. (2002) Lack of Character: Personality Moral Behavior. Cambridge University Press.
- (2012). The Moral Psychology Handbook. Oxford University Press.

- Driver, J. (1989). "The Virtues of Ignorance", *The Journal of Philosophy*, Vol. 86, No. 7, 373-384.
- (2001). *Uneasy Virtue*. Cambridge University Press.
- Emmons, R.A. and M.E. McCullough (2003). "Counting Blessings Versus Burdens: An Experimental Investigation of Gratitude and Subjective Well-Being in Daily Life", *Journal of Personality and Social Psychology*, No. 2, 377-389.
- Fagles, R. (transl.) (1983). *Sophocles: The Three Theban Plays*. Penguin Group.
- Feynman, R. (1964). "Quantum Mechanics", *The Messenger Lectures*, MIT Press.
- Flanagan, O. (1990). "Comments and Criticism: Virtue and Ignorance", *The Journal of Philosophy*, Vol. 87, No. 8, 420-428.
- Foot, P. (1978). Virtues and Vices. *Virtues and Vices and Other Essays in Moral Philosophy*, pp. 1-18. Oxford University Press. Reprinted in Crisp (1997), 163-177 (All page references are to Crisp).
- (2001). *Natural Goodness*. Oxford University Press.
- Geach, P. (1977). *The Virtues*. Cambridge University Press.
- Green A.R. (2009). "Response to Dawson and Arkes", *Journal of General Internal Medicine*, 24(9) 1083.
- et. al. (2007). "Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients", *Journal of General Internal Medicine*, 22(9), 1231-1238.
- Grierson, B. (2014). "What if Age Is Nothing but a Mind-Set?", *The New York Times Magazine*, October 22, 2014.
- Hare, S. (April 1996). "The Paradox of Moral Humility", *American Philosophical Quarterly*, Vol. 33, No. 2, 235-241.
- Harman, G. (1999). "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error", *Proceedings of the Aristotelian Society New Series*, Vol. 99, 315-331.
- Hart, H.L.A. (1958). "Positivism and the Separation of Law and Morals", *Harvard Law Review*, Vol. 71, No. 4, 593-629.

- Hartshorne, H. and M. May (1930). Studies in the Nature of Character III: Studies in the Organization of Character. The Macmillan Company; New York, NY.
- (1929). Studies in the Nature of Character II: Studies in Service and Self-Control. The Macmillan Company; New York, NY.
- (1928). Studies in the Nature of Character I: Studies in Deceit. The Macmillan Company; New York, NY.
- Hood, W. (2013). “My Dinner with Dr. King”, The New York Times, April 4, 2013, A23.
- Hursthouse, R. (2012). “Virtue Ethics”, Stanford Encyclopedia of Philosophy, available at <http://plato.stanford.edu/entries/ethics-virtue/>.
- (1999) On Virtue Ethics. Oxford University Press.
- Isen, A.M. and P.F. Levin (1972). “Effect of Feeling Good on Helping: Cookies and Kindness”. *Journal of Personality and Social Psychology* 21:384-388.
- Kahan, D., E. Dawson, E. Peters and P. Slovic (2013). “Motivated Numeracy and Enlightened Self-Government”. The Cultural Cognition Project, Working Paper No. 116. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2319992.
- Kahneman, D. (2011). Thinking Fast and Slow. Farrar, Straus and Giroux; New York, NY.
- Kang, Y., J. Gruber, and J. Gray (2014). “Mindfulness: Deautomatization of Cognitive and Emotional Life”, The Wiley Blackwell Handbook of Mindfulness, John Wiley & Sons, 168-185.
- Korsgaard, C. (2008). The Constitution of Agency: Essays on Practical Reason and Moral Psychology. Oxford University Press.
- “Aristotle’s Function Argument”, in (2008), 129-150.
- “Aristotle on Function and Virtue”, in (2008), 151-173.
- LaCour, M. And D. Green (2014). “When contact changes minds: An experiment on transmission of support for gay equality”, *Science*, Vol. 346, Issue 6215, 1366-1369.

- Langer, E. (2009) Counter Clockwise: Mindful Health and the Power of Possibility. Ballantine Books; New York, NY.
- (1989) Mindfulness. Da Capo Press; Cambridge, MA.
- and B. Chanowitz “Knowing More (or Less) Than You Can Show: Understanding Control through the Mindlessness-Mindfulness Distinction”, in J. Garber and M. Seligman (eds.) (1980). Human Helplessness: Theory and Applications. Academic Press; New York, NY, pp. 97-129.
- Levin P. F. and A. M. Isen (1975). “Further Studies on the Effect of Feeling Good on Helping”. *Sociometry*, Vol. 38, No. 1 (Mar., 1975), 141-147.
- Louden, R. (2007). “Kantian Moral Humility: Between Aristotle and Paul”, *Philosophy and Phenomenological Research*, Vol. 75, No. 3, 632-639.
- MacIntyre, A. (1981) After Virtue: A Study in Moral Theory. Notre Dame University Press.
- McCullough, M.E., Kilpatrick, S.D., Emmons, R.A., & Larson, D.B. (1999). Gratitude as moral affect, *Psychological Bulletin*, 127, 249-266.
- , Emmons, R.A., Tsang, J. (2002). “The Grateful Disposition: A Conceptual and Empirical Topography”. *Journal of Personality and Social Psychology*, Vol. 82, No. 1, 112-127.
- , Bono, G. (2006). “Positive Responses to Benefit and Harm: Bringing Forgiveness and Gratitude Into Cognitive Psychotherapy”. *Journal of Cognitive Psychotherapy: An International Quarterly*, Vol. 20, No. 2.
- McDowell, J. (2010) “Brandom on Observation”, in Weiss and Wanderer (2010), 129-144.
- (1998). Mind, Value, & Reality. Harvard University Press.
- “Two Sorts of Naturalism”, in (1998), 167-197.
- “Virtue and Reason”, in (1998), 50-73.
- “Some Issues in Aristotle’s Moral Psychology”, in (1998), 23-49.
- “Non-Cognitivism and Rule-Following”, in (1998), 198-218.
- Merritt, M., J. Doris, and G. Harman (2012). “Character”, in Doris 2012.

- Mikhail, J. (2011). Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment. Cambridge University Press.
- Milgram, S. (1963). "Behavioral Study of Obedience", *Journal of Abnormal and Social Psychology*, Vol. 67, No. 4, 371-378.
- Milligan, T. (2007). "Murdochian Humility", *Religious Studies*, Vol. 43, No. 2, 217-228.
- Mooney, C. (2013). "Science Confirms: Politics Wrecks Your Ability to Do Math", *Mother Jones*, Sep. 4, 2013. Available at <http://www.motherjones.com/politics/2013/09/new-study-politics-makes-you-innumerate>
- Moss-Racusin, C.A., et. al. (2012). "Science faculty's subtle gender biases favor male students", *Proceedings of the National Academy of Sciences of the United States*, Vol. 109, No. 41, 16474-16479.
- Noubary, R. "What is the Speed Limit for Men's 100 Meter Dash". Available online at <http://mathaware.org/mam/2010/essays/NoubaryRun.pdf>
- Park, N., Peterson, C., & Seligman, M.E.P. (2004). Strengths of character and well-being. *Journal of Social and Clinical Psychology*, 23, 603-619.
- Perry, Gina. (2013) Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments. The New Press, New York, NY.
- Peterson, C. and Seligman, M. (2004). Character Strengths and Virtues: A Handbook and Classification. Oxford University Press.
- Peterson, C. (2006). A Primer in Positive Psychology. Oxford University Press.
- Priest, G. (2013). "Dialetheism", *Stanford Encyclopedia of Philosophy*, available at <http://plato.stanford.edu/entries/dialetheism/>.
- Pronin, E. (2009). "The Introspection Illusion", in *Advances in Experimental Social Psychology*, Vol. 41, Burlington: Academic Press, 1-67.
- , Gilovich, T., and Ross, L. (2004). "Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others", *Psychological Review*, Vol 111, No. 3, 781-799.

- , Lin, D., and Ross, L. (2002). “The Bias Blind Spot: Perceptions of Bias in Self Versus Others”, *Personality and Social Psychology Bulletin*, Vol. 28, No. 3, 369-381.
- Pylyshyn Z (1999). “Is Vision Continuous With Cognition? The Case For Cognitive Impenetrability of Visual Perception”, *Behavioral and Brain Sciences*, 22, pp. 341-423.
- Rawls, J. (1971). A Theory of Justice. Harvard University Press.
- Richards, N. (1988). “Is Humility a Virtue?”, *American Philosophical Quarterly*, Vol 25, No. 3, 253-259.
- Rorty, A. (ed.) (1980). Essays on Aristotle’s Ethics. University of California Press.
- Ryan, N. (2002) “Affirmation in Response to Motion to Vacate Judgment of Conviction”, *New York v. Wise et. al.*, filed in the Supreme Court of New York, Indictment No. 4762/89
- Sabin, J. And Anthony G. (2012). “The Influence of Implicit Bias on Treatment Recommendations for 4 Common Pediatric Conditions: Pain, Urinary Tract Infection, Attention Deficit Hyperactivity Disorder, and Asthma”. *American Journal of Public Health*, Vol 102, No. 5, 988-995.
- Sabini, J. and M. Silver. “Lack of Character? Situationism Critiqued”. *Ethics*, 115 (April 2005), 535-562.
- Schueler, G.F. (1997). “Why Modesty Is a Virtue”, *Ethics*, Vol. 107, No. 3, 467-485.
- Schwitzgebel, E. (2008). “The Unreliability of Naive Introspection”, *Philosophical Review*, Vol. 117, No. 2, 245-273.
- Sherman, N. (1997). Making a Necessity of Virtue: Aristotle and Kant on Virtue. Cambridge University Press.
- Sreenivasan, G. “Errors about Errors: Virtue Theory and Trait Attribution”. *Mind*, Vol. 111, No. 441 (Jan., 2002), 47-68.
- “Character and Consistency: Still More Errors”. *Mind*, Vol. 117, No. 467 (July, 2008) 603-612.
- Stanovich, K. (2011). Rationality and the Reflective Mind. Oxford University Press.

- Statman, D. (1992). "Modesty, Pride and Realistic Self-Assessment", *The Philosophical Quarterly*, Vol. 42, No. 169, 420-438.
- Tavris, C and E. Aronson (2008). Mistakes Were Made (but not by me). Harcourt; Orlando, FL.
- Thompson, M. (2008). Life and Action: Elementary Structures of Practice and Practical Thought. Harvard University Press.
- Tiberius, V. and Plakias, A. (2010). "Well-Being". In Doris, J. (2010) The Moral Psychology Handbook. Oxford University Press.
- Urmson, J. O. "Aristotle's Doctrine of the Mean". *American Philosophical Quarterly* 10 (1973), 223-230. Reprinted in Rorty (1980), 157-170.
- Vaillant, G.E. (2002). Aging well. New York: Little, Brown.
- Vranas, P. (2005). "The Indeterminacy Paradox: Character Evaluations and Human Psychology". *Nous* 39:1, 1-42.
- (2009). "Against Moral Character Evaluations: The Undetectability of Virtue and Vice". *Journal of Ethics*, vol. 13, 213-233.
- Wallis, G. (2004). The Dhammapada, Random House: Modern Library Classics, New York, NY.
- Watkins, P.C., K. Woodward, T. Stone and R. Kolts (2003). "Gratitude and Happiness: Development of a Measure of Gratitude, and Relationships with Subjective Well-Being", *Social Behavior and Personality*, 31(5), 431-452.
- Weiss, P. (1941). "Adventurous Humility", *Ethics*. Vol. 51, No. 3, 337-348.
- Weinstein N., W. Ryan, C. DeHaan, A. Przybylski, N. Legate, and R. Ryan (2012). "Parental Autonomy Support and Discrepancies Between Implicit and Explicit Sexual Identities: Dynamics of Self-Acceptance and Defense", *Journal of Personality and Social Psychology*, Vol. 102, No. 4, 815-832.
- Weiss, B. And Wanderer, J. (2010). Reading Brandom: On Making It Explicit, Routledge, London and New York.
- Williams, B. (1985). Ethics and the Limits of Philosophy. Harvard University Press.
- (1981). Moral Luck. Cambridge University Press.

— (1980). “Justice as a Virtue”, in Rorty (1980), 189-199.

Winter, M.J. (2012). “Does Moral Virtue Require Knowledge? A Response to Julia Driver”, *Ethical Theory and Moral Practice*, 15:533-546.

Zanna, M. (2009). Advances in Experimental Psychology, Vol. 41, Burlington: Academic Press.

Zimbardo, P. G., W. C. Banks, C. Haney, and D. Jaffe (1973). “The Mind Is a Formidable Jailer: A Pirandellian Prison”. *New York Times Magazine*, April 8, 1973.