Irresistible Reasons, Immovable Minds, and the Miracle of Rational Persuasion

by

Stephen Martin

Department of Philosophy
Duke University

Date: ____________________________

Approved:

____
Owen Flanagan, Supervisor

____
Karen Neander

____
Michael Ferejohn

____
Alexander Rosenberg

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Philosophy
in the Graduate School of Duke University
2014

Abstract

Irresistible Reasons, Immovable Minds, and the Miracle of Rational Persuasion

by

Stephen Martin

Department of Philosophy
Duke University

Date: ______________________

Approved:

____
Owen Flanagan, Supervisor

____
Karen Neander

____
Michael Ferejohn

____
Alexander Rosenberg

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Philosophy
in the Graduate School of Duke University
2014

Copyright © 2014 by Stephen Martin
All rights reserved

Abstract

My dissertation is about good arguments and why they fail to persuade. Besides being a common experience of everyday life, this is an old worry of Plato's that continues to motivate two contemporary lines of research. The first concerns what makes something a good argument, and the second concerns what a mind must be like to be moved by one. Together, these lines guide my project and divide it into two parts. Part I is about good reasons, specifically *epistemic* reasons. In my first chapter, I defend epistemic instrumentalism, the position that epistemic reasons are good reasons only relative to one's epistemic preferences. I acknowledge that epistemic instrumentalism opens the door to a terrible proliferation of incompatible preferences, but claim that this is merely a potential problem, and not an actual problem to be solved. In my second chapter, I discuss the nature of reasonhood, and argue, contrary to orthodoxy, that there is no compelling reason to accept the skeptic's claim that, because of the inconsistency of three very basic epistemic preferences, it is impossible for any position to be conclusively safe to hold. Part II is about immovable minds. Immovable minds are minds that are unpersuaded by good reasons. In my third chapter, I argue that for good reasons to be persuasive, the properties that make them good reasons must be identified, through habituation, with other desirable qualities like pleasure or success. Identifying the merits of good reasons with other rewards cultivates intellectual character, and intellectual character, as I argue in my final chapter, remains worth cultivating, notwithstanding *situationist* doubts about the existence of character and *intuitionist* concerns about human rationality.

To my mother

Contents

Abstract iv

List of Tables viii

Acknowledgements ix

1 Epistemic Preferences and the Realities of Intellectual Life 1

1.1 Socratic Intellectualism 2

1.2 Plato's Departure 4

1.3 Epistemic Preferences and the Pursuit of Truth 9

1.4 Epistemic Instrumentalism and Intellectual Coexistence 15

2 Reasonhood 22

2.1 Skepticism 25

2.2 Rational Protection & Foundationalism 29

2.3 Circularity & Coherentism 32

2.4 Finitude & Infinitism 37

2.5 Epistemology and the Evaluation of Epistemic Preferences 40

3 The Miracle of Rational Persuasion 44

3.1 Rational Persuasion is Postulated without Explanation 45

3.2 Rational Persuasion Demands an Explanation 48

3.3 Rational Persuasion Requires Intellectual Character 54

3.4 Intellectual Character Must Be Nurtured 59

4 **Two Empirical Challenges to Character** **62**

4.1 Background 62

4.2 Moderate Situationism and Moral Character 67

4.3 Automatic Intuitionism 73

4.4 A Perfect Storm 77

4.5 Conclusion 84

Bibliography **85**

Biography **97**

List of Tables

2.1 Good Reasons 24
3.1 Explanatory Options 51

Acknowledgements

There are many to thank: La Salle University and the La Salle philosophy department, especially Joe Volpe, Marc Moreau, and Mike Kerlin, who got me started; the Tufts philosophy department, especially Nancy Bauer, George Smith, Jody Azzouni, and Dan Dennett, who taught me that this wasn't going to be easy; Felipe De Brigard and Anthony DiClaudio, friends throughout; Duke University and the James B. Duke Fund, who made the life of a lowly graduate student one of relative abundance; Duke philosophy, where I made my home; Karen Neander, Michael Ferejohn, and Alex Rosenberg, who made this project theirs, too; my Duke officemates: Matt Braddock, who ate my lunch, Gordon Steenbergen, who heard everything here first, and Nate Gindele, who was there at every turn; my sister Rachel and my brother Thomas, who are unlikely to read much further; my father, who has taught me more than anyone else about what it means to argue unconvincingly; and my mother, Diane, who taught me to read and who taught me to write and who taught me mostly everything else. She taught me to do my best in school and even though she didn't make it to the end she still made sure that I did.

There is Owen Flanagan, who found me on the scrap heap and looked out for me and brought me to Duke and brought me along and saw me through. He taught me that writing is rewriting and rewriting and rewriting and that there is still a lot of rewriting to do. He pushed me and guided me and inspired me and he taught me in one way after another what a philosopher really is. I cannot thank him enough.

And there is Kristen Powell, my best friend, my college sweetheart, my partner, my wife.

Through everything–this school and that school, this job and that; applications, rejections; tough breaks, lucky breaks, fall breaks, winter breaks, spring breaks, summer breaks; moves north and south and north and south; Medford, Somerville, Boston, Durham, Waltham, Worcester, Knoxville; flights overseas and treks across borders; miles and miles of marathon miles; the cat in the backseat, the dog in the front, two hundred, three hundred, four hundred more miles; home for the hell of it, home for the holidays, home, the two of us, one of these days. Through everything Kristen has been constant and patient and forgiving and loving and true. I don't know if this project would have come to completion without her, but I know that it would not have been worth it if it had.

1

Epistemic Preferences and the Realities of Intellectual Life

The problem Callicles raises in the *Gorgias* is that you cannot make someone virtuous through reason alone unless that person actually prefers virtue most of all, and that some people, people like Callicles, do not prefer virtue. This problem recurs in epistemic form. It is that you cannot rationally persuade someone to believe something with a certain epistemic property unless that person prefers positions with that property, and that some people, in fact, do not prefer positions with certain epistemic properties. Both problems are consequences of instrumentalism about rationality, the thesis the rationality of one's decisions, including one's decisions about what to believe, is entirely relative to one's preferences. Following Plato's argument in the *Republic* and Aristotle's endorsement of it in the *Nicomachean Ethics*, I argue in this chapter that because instrumentalism is true, these problems must be acknowledged and managed, civilly and sometimes civically, to the best of our abilities.

First I outline the elements of Socratic intellectualism as they appear in the *Gorgias*. My intention in doing so is to reconstruct the principles from which Socrates both argues that no one really prefers injustice to justice, and explains why Callicles is unconvinced by his

argument. Second, I detail Plato's departure from Socratic intellectualism in the *Republic* and highlight what about his own view Aristotle embraces in the *Ethics*. My aim is to establish the resources for dealing with the potential proliferation of fundamental preferences for things other than virtue, the possibility of which Socrates denies and Plato and Aristotle accept. Third, I outline the elements of *epistemic intellectualism*. I argue that epistemic intellectualists make a mistake similar to that of Socrates intellectualists in supposing that, because every rational person prefers the truth most of all, that no two rational individuals could possibly differ with respect to their most fundamental epistemic preferences. Finally, I consider how we might manage the more troublesome consequences of epistemic instrumentalism, appealing to the *Republic* and the *Ethics* for insight.

1.1 Socratic Intellectualism

Bernard Williams writes that the argument of the *Gorgias* is "so unconvincing, in fact, that Plato later had to write the *Republic* to improve upon it" (2006, p. 22). That argument, long and winding as the *Gorgias* is, is straightforward.

1. Every rational person prefers the good most of all.
2. Justice, *qua* virtue, is good.
3. Therefore, no one really prefers injustice to justice.

Distinctively Socratic doctrines follow.

4. A lack of virtue reflects either a preference for something other than the good, or false beliefs.
5. From 1 and 4, a lack of virtue reflects false beliefs.

Suppose, for example, that I vote for war, and you do not. It appears that we prefer different things, but if Socrates is right, then we do not really differ in our preferences. Rather, we

differ only in our beliefs about the best means to achieve what we prefer most. Thus the discrepancy between us must ultimately be epistemic in nature.

Let us suppose further that war in this case is unjust, making my vote for it vicious and your objection to it virtuous. It is then clear that my becoming virtuous requires only that my position be brought into line with yours.

6. Virtue is achieved by abandoning false beliefs.

These doctrines have methodological consequences. Socratic questioning *exposes* false beliefs by their inconsistency with a preference for the good, and critical thinking *determines* which among one's inconsistent beliefs are rational and which are not.

7. Therefore, everyone capable of answering Socratic questions honestly and engaging in critical thinking is capable of virtue.

Crucially, if virtue is to be achieved by abandoning false beliefs, it must be one's actual beliefs that are under consideration. And to do that, one must say what one really believes when questioned. This conversational ground rule is the only rule Socrates asks his interlocutors to accept. Its function is to guarantee that his conversations expose actual false beliefs for elimination, thereby effecting actual moral improvement.[1] In the absence of such a constraint, one can save the appearance of consistency through cleverness and creativity alone.

The principles Socrates uses to try to convince Callicles that he must be wrong about justice in the *Gorgias*, with the addition of one methodological assumption, are the same principles he uses to explain why Callicles is unconvinced. That methodological assumption is assumption 8:

[1] Gregory Vlastos (1982) says that the requirement of sincerity alone distinguishes what Socrates is up to from mere competitive debate. Competitive debate tolerates, indeed recommends, advancing arguments without any belief in the truth of their premises if doing so is advantageous. For Vlastos, the sincerity constraint is Socrates' way of making sure his interlocutors are willing to be refuted.

8. Every epistemic discrepancy can be resolved by answering Socratic questions honestly and thinking critically.[2]

Because every rational person desires the good most of all, the inconsistencies in what Callicles says cannot really reflect a desire for something other than the good, if he is truly behaving rationally. And so they cannot really reflect a desire for injustice. Therefore they must reflect inconsistent beliefs. Because every epistemic discrepancy can be resolved by answering Socratic questions honestly and engaging in critical thinking, and because his claims are brought into line by reason without any real change in his mind, Callicles must therefore be behaving irrationally, or otherwise lying, joking, or somehow simply failing to express his beliefs accurately.[3]

1.2 Plato's Departure

The assumption that everyone prefers the good most leads to two of Socrates' major claims in the *Gorgias*. The first is that no one really prefers injustice, and the second is that Callicles is unconvinced that no one really prefers injustice because he is either behaving irrationally or not really saying what he believes. If that assumption is false, however, then these claims

[2] This assumption states a sufficient condition on the achievement of rational consensus. It says that if a dispute is epistemic in nature, then it is, in principle at least, resolvable. It is by no means obvious, and there are three objections to it that are worth mentioning. The first challenge is from the Pyrrhonian skeptics. Pyrrhonians claim that no epistemic dispute can be indisputably resolved without appeal to an indisputable standard of proof, and yet no such standard exists, meaning that no epistemic dispute can be indisputably resolved. The second challenge is from Kant. Kant claims that no dispute about things in themselves can be resolved by appeal to evidence from the realm of appearances, and yet the only evidence there is for belief is evidence from the realm of appearances. The third challenge is from Quine. Quine claims that no conflict between empirically adequate, mutually inconsistent theories can be resolved by appeal to empirical evidence, and yet the only evidence there is for a theory is empirical evidence, meaning that some epistemic disputes are unresolvable. Each of these arguments aims to prove that there are at least some some beliefs or sets of inconsistent beliefs that cannot be brought into line with one another. Each of these challenges amounts to the claim that there are some epistemic disputes–disputes over the truth of epistemic principles, disputes over metaphysics, disputes between equally empirically adequate theories–that cannot be resolved by answering Socratic questions honestly and thinking critically.

[3] As Socrates says, "These conclusions, at which we arrived earlier in our previous discussions are, I'd say, held down and bound by arguments of iron and adamant... And if you or someone more forceful than you won't undo them, then anyone who says anything other than what I'm now saying cannot be speaking well." (G. 508e-509a)

are unwarranted. And indeed there are good reasons for believing that not everyone prefers the good most of all.

Socrates' justification for the assumption that everyone prefers the good most in fact equivocates on what is meant by "the good." Socrates' argument (G. 467c-468e) shows that anyone who prefers anything must prefer something most of all. And indeed everyone, conventionally, calls that something "the good." But Socrates' mistake in the *Gorgias* is to move from the (perhaps) *uncontroversial* fact that every rational person prefers *something* most of all, which everyone may agree is *called* the good, to the *controversial* claim that everyone prefers the *same something* most of all. Socrates equivocates specifically in taking agreement about an objective, interpersonal, functional definition of "the good" as that which everyone prefers most of all to entail agreement about an objective, interpersonal, *specification* of "the good" as it is functionally defined. The question one must ask is whether the good as Socrates specifies it in 2, that is, as justice *qua* virtue, is the only thing capable of being what which someone prefers most of all.

Plato's answer in the *Republic* is that virtue is not alone in its capacity to function as that which someone prefers most. In fact, he argues, there are two other kinds of thing capable of doing that. The first kind includes things like victory, glory, honor, and fame. The second includes things like pleasure, wealth, and the satiation of physical needs. Both of these alternatives are capable of competing with and thereby taking over one's preference for virtue. Since not everyone prefers virtue most of all, not every lack of virtue reflects ignorance about the best means to the good, and since not every lack of virtue reflects ignorance about the best means to the good, not every lack of virtue can be remedied by answering Socratic questions honestly and engaging in critical thinking. Sometimes the only way to correct for a lack of virtue is to change what a person prefers most of all, that is, to create in a person a preference for justice.

There are more ramifications still. Though Plato, in the *Republic*, rejects the thesis that everyone fundamentally prefers the good, insofar as it is specified in terms of virtue, he

implicitly maintains both 9 and 10:

9. What a person prefers most of all cannot be changed by reason.
10. Disputes about what is most preferable cannot be resolved by answering Socratic questions honestly and thinking critically.[4]

Assumptions 9 and 10 express the idea that neither the motivational component of what we prefer most of all nor disagreements about what is most preferable flex under the pressure of rational argumentation.

For the Socrates of the *Gorgias*, 9 and 10–if they are true–are trivially true. If everyone desires the good most of all, and the good is specified in terms of virtue, then trying to change what someone desires most of all to make that person more virtuous, whether by reason or by any other means, is ill-conceived. It is ill-conceived because that person already desires virtue. Moreover, since everyone desires the good most of all, then there are no genuine disputes about what is most desirable to be resolved.

For the Socrates of the *Republic*, however, and so a key feature of Plato's apparent departure from Socratic intellectualism, the proliferation of potentially inconsistent fundamental preferences creates a novel problem. The problem is that a person might not prefer virtue most for entirely non-cognitive, conative reasons. The task is then to say how someone who truly lacks virtue–someone like Callicles, or at least someone like the tyrant that Callicles and later, in the *Republic,* Thrasymachus aspire to be–can acquire virtue if not by simply drawing out inconsistencies and abandoning false beliefs. The *Republic*'s novel solution to that problem, one Aristotle adopts as well, is that what a person prefers most of all depends on one's upbringing, which is in turn sensitive to conditions beyond that person's control, things like political climate, family structure, and education. It is these conditions that partially explain the acquisition of one's fundamental preferences, and in turn partially explain the capacity for Socratic questioning and critical thinking to result in moral improvement.

[4] This assumption is logically equivalent to the converse of 10.

Only if a person is habitually predisposed to desire virtue most of all is answering Socratic questions and thinking critically, or, more broadly, philosophy itself, capable of making a person's behavior more virtuous.[5]

It is worth taking stock. Socrates, Plato, and Aristotle each maintain that disagreements can be resolved by reason if and only if they reflect differences between beliefs and not differences with respect to non-cognitive, conative variables of individual preferences. Socrates, as he appears in the *Gorgias* and other early dialogues, does not believe it possible for fundamental preferences to be inconsistent for non-cognitive, conative reasons because everyone prefers the good most of all, which implies that every disagreement can be resolved by reason. Plato and Aristotle disagree. Both hold that fundamental preferences may be inconsistent for non-cognitive, conative reasons, that they sometimes are inconsistent for non-cognitive, conative reasons, and that these inconsistencies are the result of differences in upbringing. Only if two individuals have the same upbringing can what they desire most of all be consistent, only if they have a virtuous upbringing can they desire virtue most of all, and only if they desire virtue most of all can answering Socratic questions and thinking critically make them more virtuous than they would otherwise be.

Still, if someone can fundamentally prefer *general* things like victory, glory, honor, and fame, or pleasure, wealth, and the satiation of physical needs for non-cognitive reasons, the natural question is why someone could not prefer other more *specific* things most of all too: winning the presidency of the United States, summiting Mt. Everest, having sex with a particular person. That is to say, why believe that all of our many apparent preferences reduce to preferences for one of three fundamental things, ostensibly making it much more likely that any given disagreement is going to reflect a difference in beliefs and nothing else,

[5] Compare Plato's explanation of the acquisition of virtuous fundamental values with Mill's remarks about the acquisition of a taste for higher pleasures: "Capacity for the nobler feelings is in most natures a very tender plant, easily killed, not only by hostile influences, but by mere want of sustenance; and in the majority of young persons it speedily dies away if the occupations to which their position in life has devoted them, and the society into which it has thrown them, are not favourable to keeping that higher capacity in exercise. [Individuals] lose their high aspirations as they lose their intellectual tastes, because they have not time or opportunity for indulging them, and they addict themselves to inferior pleasures."

and thereby making it much less likely that any given dispute is going to be unresolvable?

The idea here echoes a key moment from the *Parmenides*. Why, Parmenides asks Socrates, if there are forms of goodness, justice, and beauty, can there not be forms of human beings, of fire, and of water, or, for that matter, forms of "hair, mud, dirt, or anything else which is vile and paltry"? (Par. 130a–e). If fundamental desires proliferate to include anything whatsoever, then it may be that every serious disagreement reflects a dispute about what is most desirable. And if every serious disagreement reflects a dispute about what is most desirable, then no serious disagreement can be resolved by reason. This problem is more familiar as the problem of instrumental rationality. If rationality is defined as a calculus of means, given one's ends, and if one's ends can include anything whatsoever–and not just a few basic things–then it is possible for two individuals to adopt subjectively rational, intersubjectively inconsistent means, about which it would therefore be irrational to come to any agreement.

Plato does not raise or address this problem exactly, but Aristotle raises something like it like it in the *Nicomachean Ethics*. Aristotle begins the *Ethics* by highlighting the Socratic idea that for every set of preferences, there is something a person prefers most of all, a *summum bonum* or *highest good*. This good generally goes by the name of *eudaimonia*, or happiness.[6] Unlike Socrates, however, Aristotle acknowledges that whatever the highest good is, that is, whatever happiness may consist in, may vary from person to person. This is because it is underspecified by the functional definition as that which everyone prefers most of all. Implicitly, the number of possible highest goods is logically unrestricted.[7]

---

[6] The idea of a highest good would be redundant for Socrates, as he appears in Plato's pre-*Republic* works.

[7] Aristotle writes: Verbally there is very general agreement; for both the general run of men and people of superior refinement say that it is happiness, and identify living well and faring well with being happy; but with regard to what happiness is they differ, and the many do not give the same account as the wise. For the former think it is some plain and obvious thing, like pleasure, wealth, or honour; they differ, however, from one another–and often even the same man identifies it with different things, with health when he is ill, with wealth when he is poor; but, conscious of their ignorance, they admire those who proclaim some great thing that is above their comprehension. Now some thought that apart from these many goods there is another which is good in itself and causes the goodness of all these as well. To examine all the opinions that have been held would no doubt be somewhat fruitless: it is enough to examine those that are most prevalent or that

Aristotle does not so much attempt to solve the problem of infinitely many competing highest goods as he gives reasons for not taking it to be a very serious problem. The ingenuity of his approach is this. Aristotle recognizes the difference between a potential problem and an actual problem, and the *potentially* unrestricted proliferation of non-cognitive, conative components to individual preferences is not an *actual* problem. It is not an actual problem because the non-cognitive, conative components of fundamental preferences are not *actually* unrestricted. They are not actually unrestricted because human nature–for Aristotle–is in fact well-defined. Human beings are rational featherless bipeds, biological, psychological, political animals with *de facto* constraints upon the sorts of thing that actually motivate them.[8] It is a matter of fact that the range of human motivators, far from being unrestricted, divides roughly into the three categories identified by Plato. And not insignificantly, since the non-cognitive, conative element of what is most preferable does not vary much from person to person even if it does vary at all, neither should openness to reason between persons, meaning that actual cases of irresolvable disagreement should be few and far between.

1.3 Epistemic Preferences and the Pursuit of Truth

Socrates, Plato, and Aristotle each maintain that disagreements can be resolved by reason if and only if they reflect differences between beliefs and not differences with respect to the non-cognitive, conative component of what is most preferable. This thesis is about preferences in general. But it has an interesting consequence. It if its true, then by extension:

11. Disagreements can be resolved by reason if and only if they reflect differences between beliefs and not differences with respect to what is most *epistemically* preferable.

seem to have some reason in their favour. (1984, 1095a16-30)

[8] Mill defends the principle of utility on similar grounds, arguing that people do not in fact differ in their desire for happiness above all else and that what this happiness consists in is a function of human psychology.

Epistemic preferences are preferences for beliefs with certain properties. If two individuals cannot agree about which sorts of beliefs are most preferable, then they may be incapable of settling any given dispute about what to believe.

In fact, human beings differ greatly with respect to their fundamental epistemic preferences. Mathematicians prefer proven beliefs. Religious people prefer beliefs that resonate with scripture. Scientists prefer empirically tested beliefs. Pragmatists prefer warrentedly assertable beliefs. Conformists prefer beliefs that fit in. Contrarians prefer beliefs that stand out. Sensationalists prefer beliefs that create a stir. Wishful thinkers prefer beliefs that make them feel good. And philosophers, at least as far as other people are concerned, prefer absolutely indubitable beliefs (for themselves, of course, they are willing to settle for much less).

One thing I have left out as an object of epistemic preferences is truth. This is not an oversight. When philosophers talk about various epistemic preferences, their discussions tend to begin and end with the recognition of preferences for true beliefs, as opposed to preferences for anything else.[9] Besides being downright unimaginative, there are two major reasons that these discussions are misguided. The first, of course, is that truth, like existence, is not a property. The truth predicate does not designate a universal, something that every true belief has in common. Rather, it functions as a linguistic device for endorsing the content of a belief without expressing that content. It is a mechanism of what Quine (Cf. 1960, p. 251), with Tarski's work in mind, calls *semantic ascent*. The *ascent* involved is that of stepping "up" into a metalanguage from the object language in which the content of a belief is expressed. From the standpoint of a metalanguage, one can endorse something in the object language by referring to it by name–mentioning it–as true, rather than by expressing it–using it–to assert what one wants to assert. The rub is that since truth is not a property of beliefs, it cannot be the property that makes beliefs preferable.

[9] See Kelly

The second major problem with thinking of truth as the object of epistemic preference is that doing so effectively reiterates, in epistemic terms, the Socratic thesis that every preference ultimately aims at the good.

12. Every rational person prefers true beliefs most of all.

This thesis sometimes takes the form of the claim that beliefs *aim at* or *aspire to* truth, that the rational person is engaged in the *pursuit of truth*. And like the thesis that every rational person prefers the good most of all, it can be understood in a trivial way, in which the nature of truth is unspecified, or in a substantial way, in which the nature of truth *is* specified, just as Socrates does by identifying justice, *qua* virtue, with the truth in the *Gorgias* (at 2 above).

There is a problem, though with any attempt to specify truth by identifying it with some other epistemic property. The problem is that it cannot be done. That it cannot be done for mathematical provability in particular has been known since Gödel proved that it cannot be done for anything beyond first-order logic. That it cannot be done for other epistemically desirable properties in general, though perhaps unproven, is intuitively plausible for reasons Hilary Putnam (1978) gives in "Reference and Understanding."[10] There Putnam considers and resists Michael Dummett's *analysis* of truth into warranted assertability by pointing out that for any identification of truth with a given epistemic property, there are modal statements of the form "*S* might have property *P* and still not be *true*."[11]

Here is the idea. Take any epistemically desirable property you like, entailment by scientific consensus, for example. If the property is identical with truth, then the following true:

13. *P* is true if and only if *P* is entailed by scientific consensus.

[10] See also Rorty (1986) and Azzouni (2000) for discussions and endorsements of Putnam's argument.

[11] These modal statements, Putnam argues, are both (a) themselves plausible, and (b) indispensable to scientific practice. On Putnam's view, the plausibility of these modals owes to their indispensability, and their indispensability is a function of a need to account for a margin of error in our epistemic practices.

But not only must 13 be true if truth is identical with entailment by scientific consensus, 13 must be *necessarily* true–because of the necessity of identity statements–if truth is identical with entailment by scientific consensus. And yet *P* might be entailed by scientific consensus and still not be true, and so truth cannot be identical with entailment by scientific consensus. Run the same procedure again, but for beliefs that resonate with scripture:

14. *P* is true if and only if *P* resonates in scripture.

Again, *P* might resonate in scripture and yet still not true. If Putnam is right, then any attempt to identify truth with an actual epistemic property, desirable or not, is bound to fail. In short, it is misguided to place any weight on the claim that every rational person prefers true beliefs most of all, or that rational beliefs aim at the truth, or that the rational person is engaged in a pursuit of truth. Insofar as truth is not identical with any given epistemic property, it is not something to prefer or to aim at or to be pursued, making it false to say that every rational person prefer true beliefs most of all; insofar as truth is an underspecified placeholder for whatever it is that rational persons do prefer most of all, it is trivial to say that every rational person prefers true beliefs most of all.

I make these points because philosophers continue to focus on truth in discussing their qualms with any instrumental conception of epistemic rationality. Call these philosophers *epistemic intellectualists*. Take Tyler Burge (1993), for example, who Thomas Kelly (2003) quotes favorably:

> Reason has a function in providing guidance to truth, in presenting and promoting truth without regard to individual interest. That is why epistemic reasons are not relativized to a person or to a desire. (p. 475)

In the passage Burge states that reason has a function, and that its function is to provide guidance to truth. But if this is right, then it must be false that the dictates of reason–epistemic reasons–are not relative to a person or a matter of preference. For the person

who does not desire truth is therefore under no obligation to follow the dictates of reason. This is the problem of the *Gorgias* all over again. If morality has the function of providing guidance to virtue, in presenting and promoting virtue without regard to individual interest, then the person who does not desire virtue is under no obligation to follow the dictates of morality.

But perhaps there is a problem with this interpretation. For though a rational person can, *pace* Kant, ask why one ought to desire virtue most of all, no rational person could seriously deny that truth is most epistemically desirable. In this spirit, Kelly argues as follows. If the dictates of reason are binding only insofar as they are means to the truth as one's ultimate epistemic end, then no one has any obligation to believe something without a preference for truth about it. But we have obligations to believe things if we come across evidence for believing them, regardless of whether we want the truth about them or not. And so no one cannot rationally deny that truth is most epistemically preferable. Thus the dictates of reason, even if they are relative to a preference for truth, are categorical because the rationality of desire for truth is unquestionable.

But this argument is problematic. It proceeds by slipping in–as the source of epistemic normativity–a preference for beliefs supported by evidence of some sort (what sort of evidence that is, exactly, is left unspecified). With this preference in place, it indeed follows that if you come across something supported by that evidence, then you ought to believe it–*whether you want to or not*. And to be sure it is the "whether you want to or not" that gives the obligation to believe what is supported by the evidence you come across the appearance of categoricity. This clause states explicitly that the obligation to believe what you have come across is not relative to some desire. In fact, however, the "whether you want to or not" clause is redundant. It is redundant because we have already stipulated, tacitly, that you want beliefs supported by certain evidence. The appearance of categoricity would disappear if "whether you want to or not" were replaced by "if you want to be consistent" or "if you really desire beliefs supported by certain evidence most of all." Far from being

a situation that "fits poorly with the instrumentalist conception of epistemic rationality," (2003, p. 625) this objection exemplifies epistemic instrumentalism.

The mistake in this argument parallels the one Socrates makes in the *Gorgias*. Socrates conflates an objective, interpersonal, functional definition of "the good" as that which everyone prefers most of all with an objective, interpersonal, *specification* of "the good" as it is functionally defined. This argument conflates a trivial conception of truth as that which is most epistemically preferable (a conception no one could rationally question *because of its triviality*) with a substantial conception of truth in terms of some epistemically desirable property (a conception which is always–in principle–questionable because of the reasons Putnam gives, and which may in fact be worth questioning if there is in fact disagreement about it).

This point warrants elaboration. To say that the preference for truth is rationally unquestionable is to say that it is incoherent to desire anything but the truth most of all. And, to be sure, it *is* incoherent to ask why one should prefer true beliefs, so long as what it means to be a true belief is left unspecified, just as it is incoherent to ask why one should prefer the good, so long as the good is left unspecified. The problem for Burge and Kelly, however, is that if what it means to be a true belief is left unspecified, then there are no grounds from which they can object to someone's fundamental preference for beliefs that resonate in scripture, for example, or for beliefs that have withstood the test of time. Once Burge and Kelly fill in what they mean by truth, however–once they identify the most preferable of beliefs–it is not incoherent at all to question the fundamental desirability of truth. Specifying the most preferable sorts of beliefs requires committing to epistemic properties that someone can actually, rationally disagree with.

I suspect that the motivation for philosophical focus on truth, by both those who defend epistemic instrumentalism and those who object to it, is that it neatly allows for a division between those who think rationally and those who do not. The fundamental desire for truth, which is one, contrasts nicely with the desire for non-truth, which is many. Philoso-

phers are happy to map non-truth onto the diverse array of less philosophically reputable, mutually incompatible epistemic desires, preferences like those of the person of faith and the conformist and the contrarian and the sensationalist and the wishful thinker. But when the veil of truth is pulled away, one finds that the more reputable epistemic preferences behind it–preferences for proven beliefs, for empirically tested beliefs, and for absolutely indubitable beliefs, among them–are often equally incompatible. Any class of philosophically upstanding, conceptually distinct epistemic preferences, when push comes to shove, will exhibit differences with respect to what these preferences oblige us to believe and what they do not.

It is a mistake to overlook the great diversity and incompatibility among even reputable epistemic preference on the false assumption that all of them manifest a deeper desire for truth. It is on par with Socrates' mistake in the *Gorgias*. Similarly, it is what makes it plausible to believe that no two rational individuals can *really* prefer beliefs with different properties, since the truth is what any rational person is after. But this gets things exactly backwards. Truth is a device for endorsement, not a reason for it. We do not consider certain beliefs to be most epistemically desirable because they are true; rather, we call beliefs true because we consider them to be most epistemically preferable.[12]

1.4 Epistemic Instrumentalism and Intellectual Coexistence

The enduring lesson of the *Gorgias* and Plato's transition from it to the *Republic* is that because you cannot make someone virtuous through reason alone unless that person actually prefers virtue most of all, and because there are some people who do not prefer virtue

[12] There is another problem worth mentioning. Operating under the illusion that epistemic reasons are categorical, or that it is irrational not to desire effectively denying that Callicles presents any problem at all–is why some philosophers downplay or altogether ignore the weight of practical considerations in deciding how to determine what to believe. If the dictates of reason are categorical, then they make the same demands on everyone, regardless of circumstance. But any theory that neglects practical considerations runs the risk of practical irrelevance, and in fact there is no reason to suppose that what is most epistemically desirable, all things being equal, is *actually* most desirable when practical considerations are taken into account.

most of all, there are some people you cannot make virtuous through reason alone. Callicles appears to be one of them. The problem he raises is that moral instrumentalism–the view that what one ought to do is a function of the non-cognitive, conative element of one's preferences–combined with the proliferation of human motivators, means that there may be multiple, mutually inconsistent ways to settle moral disagreements.

As Williams (2006) points out, the ramifications of Callicles' behavior in the *Gorgias* for the categoricity of moral reasons are abundantly clear, as they were to Plato. If given the choice between personally beneficial injustice or flat-out inconsistency, one can always opt for inconsistency. Denying that epistemic preferences are preferences like any other, practically measurable against other interests one might have, suggests an unrealistic optimism about the cognitive supremacy of epistemological values that historically required a great deal of intellectual labor to discover, defend, and assimilate.[13]

The problem Callicles poses in the *Gorgias* recurs in epistemic form. Epistemic instrumentalism, like moral instrumentalism, is a form of relativism, whereby epistemic rationality is defined relative to one's epistemic preferences. The proliferation of fundamental epistemic preferences means that there are multiple ways to settle disagreements about what to believe. This proliferation is obscured by the assumption that upstanding but apparently incompatible epistemic preferences (preferences like those for provability, empirical testability, absolute indubitability, warranted assertability, and so forth) manifest a more fundamental desire for truth. But this assumption, I have aimed to show, is untenable, meaning that the consequences of epistemic relativism must be faced. The chief consequence of my concern is that you cannot rationally persuade someone to believe something because of its possession of a certain epistemic property unless that person prefers beliefs with that property to those without, and that some people, in fact, do not prefer beliefs with certain epistemic properties. The appropriate question at this stage is not about the coherence of

[13] Experiment and observation are examples of values that required a great deal of intellectual labor to discover, defend, and assimilate.

epistemic instrumentalism, but rather about how serious we should take the worst of its consequences to be.

First things first, disagreement about what is most epistemically preferable does not entail disagreement about what is more generally most preferable. The evolutionary biologist and the biblical literalist, for example, likely differ wildly in their deepest epistemological commitments. But there is nothing about the mutual inconsistency of these commitments that precludes agreement about the value of things like minimizing suffering, for example, or of making the best life possible for everyone's respective children. If the evolutionary biologist and the biblical literalist can agree about these things, and these things are more important to the both of them than their respective preferences for certain sorts of beliefs, then their differing epistemic obligations can be measured against their common values. For example, if the evolutionary biologist's preferred method of settling epistemic disputes leads to less suffering or better lives for everyone's respective children, then that success alone weighs in favor of the evolutionary biologist's fundamental epistemic preferences.

If it stretches the imagination to believe that the evolutionary biologist and the biblical literalist might have common interests beyond whatever they consider to be most epistemically desirable, consider the diversity of fundamental epistemic desires among university professors, a diversity that is much less reflected in their otherwise homogeneous social and political lives. University professors, let alone professional philosophers, may be incapable of reaching consensus about what is most preferable to believe, and yet they are nonetheless capable of peaceful coexistence, presumably because they share other values that are more fundamental than what they desire of their beliefs.

Still, it seems naïve to simply *hope* that two individuals who differ drastically in their fundamental epistemic preferences might ultimately agree at a deeper level about other things they want in life. It seems naïve because even though disagreements about fundamental epistemic preferences do not *entail* disagreements about non-epistemic desires, they so often go hand in hand. Consider the following:

> At a hearing on a Louisiana law prohibiting the promotion of creationism in science classes, a state senator asks a school teacher to provide a single experiment demonstrating Darwinian evolution. Enthusiastically, the teacher describes an experiment in which subcultures of *E. Coli* are frozen at generational intervals to visualize its evolutionary stages.
>
> "They evolve into a person?" the senator asks.
>
> "No, they, they–"
>
> "OK. I'm just asking, how, how to get there. I'm just trying to figure out how to get there."[14]

Let us assume that evolution by natural selection is a fact of nature. The senator's difficulty accepting it is not so much a misstep in a field of truth as it is a misconception surrounded by falsehoods on every side. By prompting his question about whether *E. Coli* has been shown to evolve into a human being, the school teacher's argument reveals their implicit disagreement about evolution by natural selection. Bu more likely than not, it also reveals the senator's wholesale rejection of what Sellars describes as the scientific image of human beings and their place in the world, an ideology organized around the epistemic priority of, and fundamental preference for, empirically tested beliefs. If this is right, then, given epistemic instrumentalism, the senator's false belief cannot be corrected by any mere argument, which is why he is unpersuaded by the reasons for believing in evolution that the school teacher gives him. The reasons that make it a good argument are only good reasons within an image of the world he does not share.

If anything, it is doubtful that the school teacher and the state senator differ only in their most fundamental *epistemic* preferences. The stark differences between their images

[14] The exchange occurred between Louisiana State Senator Mike Walsworth and Darlene Reaves. Phil Plait of *Slate* writes about it here: `http://www.slate.com/blogs/bad_astronomy/2013/01/23/creationist_louisiana_legislators_show_contempt_and_cluelessness_about_science.html`. Plait's article includes video of the exchange.

of the world, and between the fundamental epistemic preferences that shape them, more likely than not reflect differences in more fundamental non-cognitive, conative motivators. For someone who believes in the practical necessity and actual possibility of rational persuasion, this is the worst of all possible worlds. There is a divergence in both (a) what these individuals desire most of all–a divergence in the desires around which one's ethical and political values are organized, and (b) what sorts of beliefs they prefer most of all–a divergence in the epistemic preferences that determine what it is rational to believe.

If the most dire consequences of epistemic instrumentalism cannot be dismissed on grounds of incoherence, perhaps they can at least be managed as a practical, political problem. As we have seen, this is Plato's strategy in the *Republic*, which can be read as one extended attempt to delineate the conditions under which one can be expected to profit, morally speaking, from philosophical inquiry. Plato's approach is one that Aristotle embraces so completely that he simply asserts it, concisely and without argument, in the *Ethics*:

> Presumably, then, *we* must begin with things familiar to *us*. Hence anyone who is to listen intelligently to lectures about what is noble and just and, generally about the subjects of political science must have brought up in good habits. For the facts are the starting-point, and if they are sufficiently plan to him, he will not need the reason as well; and the man who has been well brought up has or can easily get starting-points. (1984, 1095b3-8)

He continues, directing his words to individuals like Callicles and Thrasymachus without mentioning them by name, and biding them to "hear the words of Hesiod":

> Far best is he who knows all things himself;
> Good, he that hearkens when men counsel right;
> But he who neither knows, nor lays to heart;
> Another's wisdom, is a useless wight [i.e. living creature]. (1984, 1095b10-13)

If Plato and Aristotle are right, then there are some circumstances in which rational persuasion is a lost cause. And I do think that they are right. The appropriate questions, then, are what to make of these circumstances, and what to do when they arise.

Because there can be irreconcilable differences between two rational individuals, it is important to understand why, exactly, this is a bad thing, and whether it is actually the worst possible thing. It is a bad thing because it makes it more difficult for people to get along with one another. If everyone were truly an island, the fact that rational persuasion is sometimes impossible would be just that: a fact–brute, inert, and practically inconsequential. But human beings, for good or for worse, are communal beings, and the rational irreconcilability of their differences must be faced. And here is where the ultimate motivation for intellectualists like Socrates. It is the simple fact that it is more difficult to live with people you disagree with. If reason alone can resolve any disagreement, then reason alone can make life in community less difficult.

But even if fundamental epistemic preferences proliferate, it does not follow that all is lost. First, as Aristotle suggests that the problem of proliferating motivators is much more a potential problem for moral theory than it is an actual problem, I think the problem of proliferating epistemic preferences is much more a *potential* problem for epistemology than it is an *actual* problem. I do not think there are a great many contenders actually vying to be the most fundamental of epistemic desires. To be sure, as I have argued, there are multiple incompatible epistemic values, each of which gives different answers to questions about what one ought to believe. But the fact that epistemology must acknowledge more than one fundamental epistemic preference does not mean that it must acknowledge infinitely many of them.

Second, that there are some disagreements that cannot be resolved by reason alone does not imply that reasoning with one another, even when persuasion is unlikely, is not still a good idea. Even when reason fails to persuade, it often succeeds in revealing what a person most desires, believes, and desires to believe, thus pulling up the roots of disagreement.

The Louisiana state senate hearing is a case in point. Good faith arguments, like the one the school teacher puts to the state senator, bring our most fundamental differences into relief even when they fail to persuade.

Third, and most importantly, consensus about what is most preferable, epistemically or otherwise, is a facilitator of peaceful coexistence, not a precondition. We have to live with those with whom we fundamentally disagree, whether we like it or not. To that end the conversational irreconcilability of certain disputes must ultimately give way to procedural, political resolutions, negotiated and accepted by those who cannot come to consensus. One of Plato's many achievements in the *Republic*, an achievement elaborated and extended by Aristotle in the *Ethics*, is recognizing the importance of civil procedures and civic institutions to the peaceful coexistence of those who cannot see eye to eye on fundamental issues.

2

Reasonhood

> So don't look for here for a knockdown argument that there is something wrong with knockdown arguments, for the knockdown argument to end all knockdown arguing.
>
> Robert Nozick, *Philosophical Explanations,* 1981 p. 5

Like Nozick, I believe there is something wrong with knockdown arguments. Knockdown arguments give rationally *irresistible* reasons for their conclusions. Irresistible reasons are categorical. They protect the positions they support absolutely, making it safe to hold them no matter what your preferences. But since the safety of a position is ultimately relative to one's preferences, there are no irresistible reasons, and since there are no irresistible reasons, there are no knockdown arguments. This is my starting point, to which I shall return at the end of the paper. My chief goal here is not to defend it, but to explore its consequences.[1]

If there are no irresistible reasons, then any position whatsoever may be given up if you are willing to absorb the practical cost of doing so. If you prefer experiment and observation, then a good argument is one that gives empirical reasons for its conclusion. If you

[1] Defended in Ch. 1.

prefer conformity with scripture, then a good argument is one that gives biblical reasons for its conclusion. If you prefer conventional wisdom, then a good argument is one that gives commonsense reasons for its conclusion. And if not, then not. It is this ultimate dependence on a person's antecedent preferences that makes even the best arguments ultimately resistible. The safety of a position, as I understand it, is a function of consistency with one's preferences, and nothing else.

That said, not all conditionally, hypothetically, instrumentally, relatively, resistibly good reasons are created equal. Some are still better than others. *Compelling* reasons are good reasons, but minimally so. A compelling reason makes a position relatively safer to hold, all things being equal, than some other live option because of that option's inconsistency with one's preferences. If you prefer hazelnut gelato, the cafe's not having any is a compelling reason for holding that you should try the shop across the street. While compelling reasons are good to have, *decisive* reasons are better. They make a position safer than *every* other live option because of *every* other live option's inconsistency with one's preferences. If you prefer the testimony of your own two eyes, then–to borrow an example of Dretske's (1970)–seeing the striped animal in the zebra cage is a decisive reason for holding it to be a zebra and not an antelope or a gazelle, even if it does not rule out the animal's being a cleverly painted mule. And *conclusive* reasons are, with respect to the others, best of all. They make a position safer than every other option whatsoever because of every other option's inconsistency with one's preferences. If you prefer classical mathematics to constructivism, then Cantor's diagonal argument shows conclusively that the set of real numbers is larger than the set of naturals.[2]

Each of these reasons is like a move in an abstract adversarial game, one in which a

[2] The taxonomy of reasons as compelling, decisive, and conclusive is one I borrow from George Smith, who has talked about empirical evidence this way, though not in print. My own definitions of compelling, decisive, and conclusive reasons, however, are based on Dretske's (1971) definition of conclusive reasons. Perhaps more confusingly, Dretske's conclusive reasons are my decisive reasons. Moreover, I prefer "live option" to "relevant alternative" for its harmony with Peirce's (1957) idea of *abduction*, or *inference to the best decision*. See also Harman (1965); Thagard & Millgram (1995); Churchland (2009)

Table 2.1: Good Reasons

Class	Status	Reasons in this class make a position...	Closed under entailment?
Compelling	Good	safer than some other live option because of the option's inconsistency with one's preferences.	No
Decisive	Better	safer than every other live option because of every other live option's inconsistency with one's preferences.	No
Conclusive	Best	safer than every other option whatsoever because of every other possible option's inconsistency with one's preferences.	Yes

compelling move protects you from some nearby enemy attack, a decisive move protects you from every nearby enemy attack, and a conclusive move protects you from every attack, near or far. The analogy illustrates how objectivity comes apart from categoricity. Moves in a game can be objectively good, in that whether they protect your position from attack is not up to you. And yet their objective goodness does not entail their irresistibility, in that whether you play the game is a matter for you to decide. Were it possible for both reasons and moves in a game to be categorically good, and not merely objectively good, the preferences upon which they depend would themselves have to be irresistible.

I must stress that these suggestions and stipulations represent my own preferred way to think and talk about reasonhood. They make up a terminological framework that purposely avoids reference to beliefs, propositions, justification, and knowledge, among other tropes of epistemology, none of which I believe to be indispensable to the thoughtful treatment of the problems they were invented to solve. My purpose in adopting a framework that avoids them is to avoid their baggage.[3] With neither conclusive, nor decisive, nor even compelling reasons for holding that these terms are indispensable to a fruitful discussion of reasonhood, I shall proceed on the assumption that it is safe to do without them.

[3] For example, I aim for a discussion that is neutral with respect to *propositional* projects, which have to do with the good reasons, justification, evidence, knowledge, etc., independent of their possession, and *doxastic* projects, which have to do with the possession of good reasons, justification, evidence, knowledge, etc.

The framework here is nevertheless intended to apply to the safety and protection of various philosophical positions, reflexively including philosophical positions on reasonhood like skepticism, foundationalism, coherentism, and infinitism. It goes without saying that there are no knockdown arguments for–or against–any of these positions, and so I shall not attempt to make one. Rather, contrary to orthodoxy, I shall propose that the conditions of a position's safety preferred by Sextus Empiricus and the Pyrrhonian skeptics are indeed consistent with one another.[4] It follows that none of them need be given up on pain of irrationality, as many philosophers have held, which in turn means that there is no compelling reason for choosing between the major alternatives to skepticism represented by foundationalism, coherentism, and infinitism. Moreover, there are no decisive, let alone conclusive, reasons for any single one of them.

2.1 Skepticism

I define skepticism as the position that three epistemic preferences pertaining to the safety and protection of a given position are inconsistent, which implies that at least one of them must be given up on pain of irrationality. The first preference is for positions with the protection of safe reasons over positions without. The second is for noncircular protection over circular protection. And the third is for sequences of reasons that avoid regressing *ad infinitum* over those that do not avoid a regress. Paradoxically, claims the skeptic, the first and the second preferences appear to be inconsistent with the third. To resolve the paradox, some philosophers–Aristotle and Descartes, most famously–have challenged the first

[4] Sextus Empiricus identifies five modes or routes to the suspension of judgment. Three of these are structural modes, insofar as they characterize properties of reasons in relation to one another: "The one based on infinite regress is that in which we say that what is offered as support for believing a given proposition is itself in need of such support, and that support is in need of other support, and so on ad infinitum, so that, since we have no place from which to begin to establish anything, suspension of judgment follows.... And the one based on hypothesis comes into play when the Dogmatists, involved in an infinite regress, begin with something that they do not establish but that they deem worthy of acceptance as agreed upon without question or demonstration. And the circularity mode occurs when what ought to make the case for the matter in question has need of support from that very matter; whence, being unable to assume either in order to establish the other, we suspend judgment about both." (*Outlines of Empiricism*, 166-169, trans. Mates 1996)

preference, while some have more recently challenged the others. Richard Robinson (1971; 1981), for example, has criticized the preference for noncircularity, while Peter Klein (1998; 1999; 2003) has called for abandoning the preference for safer and safer reasons, at least if it implies that there must be a limit to how safe a position can be.

The preference for positions with rational support over those without requires that you protect a position by giving safe reasons for holding it. But since a reason is just another position, every safe reason must therefore have the protection of safe reasons of its own. The preference for noncircularity requires that you not protect a position with a position it protects directly or indirectly. Since the positions that protect a position are the reasons for it, this means that you cannot include among the reasons for a position that position itself. Finally, the preference for sequences that do not regress *ad infinitum* appears to require that no sequences of reasons continue without end. On the assumption that the aim of argument is to answer a question by terminating inquiry, and that infinite sequences do not terminate, no infinite sequence of reasons can be capable of making a position safe to hold.

If the skeptic's three preferences for rational protection, noncircularity, and finitude are indeed inconsistent, then this inconsistency is a compelling reason for each of the major alternatives to skepticism, the live options being foundationalism, coherentism, and infinitism, although neither a decisive nor conclusive reason for any one of them in particular. The first thing to do, then is to determine whether or not the preferences for rational protection, noncircularity, and finitude are indeed inconsistent. As I shall now show, however, these preferences are inconsistent if and only they are interpreted very narrowly.

The skeptic's first preference is for positions with rational protection, that is, a preference for positions protected by other protected positions. It is the one that foundationalists like Aristotle prefer to demote. Schematically, we can express it thus:

15. $(\forall x)(\exists y)Ryx$

This schema is satisfied by all *serial* relations, relations like *is-greater-than* and *is-to-the-left-*

of. The reasonhood relation satisfies it if and only if every safe position has the protection of–is made safe by–a safe position. It is satisfiable on a finite domain if and only if reasonhood is circular.

The skeptic's second preference is the one for noncircularity. It is the preference for positions that are not included among the positions that protect them, directly or indirectly.[5] Schematically, we can express it thus:

16. $(\forall x)(\forall y)(Rxy \rightarrow -Ryx)$

This formula says that no position can protect a position protecting it. It holds of all *asymmetric* relations, like *is-wiser-than* and *defeats*. If Aesop is wiser than Bert, then Bert cannot be wiser than Aesop. If paper defeats rock, rock cannot defeat paper.

Although *is-wiser-than* and *defeats* are both asymmetric relations, there is one important difference between them. It is that *is-wiser-than* is *transitive*, while *defeats* is not. A relation is transitive if and only if, for *R*, if *a* bears *R* to *b*, and *b* bears *R* to *c*, then *a* bears *R* to *c*. If reasonhood is transitive, then if *a* makes *b* safe to hold, and *b* makes *c* safe to hold, then *a* makes *c* safe to hold. If not, then *a*'s protection of *b* and *b*'s protection of *c* does not entail *a*'s protection of *b*. Schematically, we can express transitivity thus:

17. $(\forall x)(\forall y)(Rxy \& Ryz \rightarrow Rxz)$

The difference between transitive and intransitive relations is important because if reasonhood is transitive, two things follow. The first is that reasonhood must be *irreflexive*. Schematically, we can express irreflexivity thus:

18. $(\forall x) - Rxx$

Relations like *is-the-mother-of* and *is-taller-than* are irreflexive. No one can be his or her own mother, and no one can be taller than him- or herself. If reasonhood is irreflexive, then

[5] Circular reasons, and the arguments they constitute, are sometimes said to *beg the question*; for the purposes of this paper, I shall take begging the question to be one and the same thing as giving a circular argument.

no position can make itself safe to hold. In other words, if the preference for noncircularity implies that reasonhood be asymmetric and transitive, then you cannot consistently hold that some positions may stand in their own self-defense. To prefer noncircularity is to prefer that every reason for position be a *new* position altogether, one that does not already occur in a sequence of reasons and the position they are reasons for.

The second thing that follows from preferring rational protection and noncircularity is that a position is safe only if there are infinitely many safe positions protecting it. If reasonhood is serial, asymmetric, and transitive, then not only must every safe position have the protection of at least one safe position, but every safe position must have the protection of infinitely many safe positions. But this implication is problematic. It appears to be inconsistent with the skeptic's preference for finitude, which appears to imply that no infinite sequence of reasons can make a position safe to hold.

Thinking of these preferences as the ground rules of an abstract adversarial game helps draw out their apparent inconsistency more concretely than does a logical proof. Imagine that each preference defines a rule for a board game drawn up by the skeptic, against whom you are playing. The preference for rational support states that you must take a safe position on your turn, or else forfeit. The preference for noncircularity states that each position you take must be a new one, that is, unoccupied prior to your taking it. And the preference for finitude states that there are only finitely many positions on the board. If you grant the skeptic the right to pass on any given turn, then you will eventually run out of safe, new positions to take on your own, leaving you with no option but to forfeit, eventually. Unless you are the skeptic, you cannot play this game and win.

If you cannot play the skeptic's game without losing, and you do not in fact want to lose, then you have a compelling reason to play another game, though of course neither a decisive nor a conclusive reason to play any other game in particular. Decisive reasons for one of the other live options, if there are any, rule out each of the others. The other live options are foundationalism, coherentism, and infinitism. Each of these options proceeds

on the assumption that the skeptical game is defective, that is, that one cannot consistently prefer that a position have finite, noncircular rational protection to be safe, and that this itself provides a compelling reason for demoting at least one preference with respect to the others. They differ over the preference that should be demoted. Foundationalists hold that the preference for rational protection is the one that ought to go; coherentists, the preference for noncircularity; and infinitists, the preference for finitude. I shall now consider the best arguments for these positions, and attempt to say how good–compelling, decisive, or conclusive–each of them really is.

2.2 Rational Protection & Foundationalism

Foundationalism, as I define it, is the position that because one cannot consistently prefer rational protection, noncircularity, and finitude, the preference for rational protection ought to be demoted with respect to the others. Foundationalists effectively propose to play a game in which the rule requiring that you take a new, safe position on your turn is conditionalized, the condition being that your enemy is capable of weakening the positions you already hold. If your position cannot been weakened, that is to say, if no compelling reason can be given to abandon a position you hold, then you are excused from the requirement to protect your position by providing safe reasons for it. Foundationalists still prefer positions protected by safe positions, but deny that a position is unsafe unless it is protected by a safe position of its own.

Aristotle makes one explicit argument for foundationalism. It is the regress argument, from the *Posterior Analytics* (72b5-73a20). Paraphrasing, it runs as follows:

> Proofs must be finite in length because a finite mind cannot traverse an infinite series, which implies that either (a) some steps in some knowledge-yielding proofs are themselves unprovable, or (b) some circular proofs yield knowledge. Circular proofs cannot yield knowledge because proofs proceed from the more

familiar to the less. Therefore, some steps in some knowledge-yielding proofs must themselves be unprovable.

The regress argument is not an argument against skepticism, but rather an argument against any relevant alternative to skepticism but foundationalism. This is clear because the argument turns on the assumption that one cannot consistently prefer rational protection, noncircularity, and finitude, which is the skeptic's position. Like the skeptic, Aristotle takes the inconsistency of these preferences to provide a compelling argument for demoting one of them with respect to the others. Because he believes he has compelling reasons for holding both circular reasons and infinite sequences of reasons to be vicious, he takes the regress argument to be decisive.

With foundationalist ground rules in place, Aristotle shows how one might play by them in the *Metaphysics* (1006a12-28). That is to say, Aristotle there attempts to identify a position that is safe to hold without any rational protection of its own. He proposes that the law of noncontradiction, the position that nothing is both true and false (or, alternatively, that nothing both has a property and fails to have it), is a position that is safe to hold without the protection of anything safer. Aristotle recognizes that the law of noncontradiction cannot be proved directly, since there are no noncircular reasons for the law of noncontradiction that do not themselves regress *ad infinitum*. Aristotle also recognizes that this would ordinarily be a major problem for a position, but holds that it is not in this case because the law of noncontradiction is special. It is special because although no good (i.e. sequences of finitely many, noncircular compelling, decisive, or conclusive) reasons can be marshaled for it, no good reasons can be marshaled against it either. Since no good reasons against the law of noncontradiction can be marshaled, there is no need to protect it by providing good reasons in its defense. It is therefore safe to hold without any rational protection.

Aristotle's account of the law of noncontradiction exemplifies his foundationalism. The preference for rational protection is demoted with respect to those for noncircularity and

finitude on the suggestion that a position requires the security of other safe positions if and only if good reasons can be marshaled against it. If a position is anything like the law of noncontradiction, against which Aristotle believes that no good argument–compelling, decisive, or conclusive–can be made, then it is exempt from the requirement that it have rational protection. As far as the foundationalist is concerned, this conditional demotion preserves the spirit of the requirement that *some* safe position need rational protection while conceding that no one can rationally prefer that *every* safe position have rational protection without giving up either the preference for noncircularity or finitude.

Like Aristotle, Descartes is also a foundationalist, and like Aristotle, Descartes holds that the inconsistency of preferences for rational protection, noncircularity, and finitude is a compelling reason for foundationalism. But unlike Aristotle, Descartes assumes foundationalism without argument. In the *Meditations*, Descartes operates on the assumption that the preference for rational protection must be demoted just as Aristotle proposes–by exempting from it positions for which there are no other options, live or not to be ruled out–and likewise attempts to identify a position that stands safely without the protection of any other safe positions. But Descartes rejects Aristotle's position that the law of noncontradiction is unassailable, for it remains possible that God has made us psychologically incapable of detecting errors in holding even those positions like the law of noncontradiction that appear to be safe without any rational protection. And so the position that Descartes identifies as unassailable is not the law of noncontradiction. Neither is it the position that you exist as long as you are thinking, with which Descartes is commonly associated. Rather, the position Descartes identifies as unassailable is the position that God cannot be malicious enough to have made us unable to detect weaknesses and improprieties in the the positions we hold most securely. For Descartes, it is God's benevolent existence that is safe to hold without the protection of any other safe positions, safer even than the law of noncontradiction, which is safe to hold because of it.[6]

[6] Even supposing that God cannot have made us so cognitively flawed that even what we cannot doubt could

The pertinent question is whether the regress argument for foundationalism is decisive, as Aristotle holds and Descartes assumes. As it stands, there are considerations philosophers have raised since that neither Aristotle nor Descartes takes up, which means there are possibilities that neither Aristotle nor Descartes considers, and therefore fails to rule out. These possibilities constitute apparently compelling objections to the argument's decisiveness. The first objection is to the position that good reasons cannot be circular. The second is to the position that good reasons cannot regress *ad infinitum*. I shall now consider these objections.

2.3 Circularity & Coherentism

Aristotle's argument against circularity depends on the law of noncontradiction. Good arguments, for Aristotle, proceed from the more familiar to the less. If the conclusion of an argument counts among the reasons given for it, then the conclusion must be more familiar than itself. But by the law of noncontradiction, nothing can be both more familiar than itself, and so nothing can count as a good reason for itself (*Pos. An.* 72b25-34).

Richard Robinson (1971; 1981) denies that there is anything wrong with circular arguments. He holds that an argument can only go wrong in one of two ways, by way of invalidity, or by offering as reasons false premises. Since neither compelling nor decisive arguments are closed under deduction, the first of these is too strong. But Robinson's second requirement, that good reasons be true, is problematic as well. The problem is that,

---

be mistaken, it has seemed to many, Mersenne and Arnauld first, that Descartes's reasons for believing it are circular. Arnauld writes in the Fourth Objections:

> You are not yet certain of the existence of God, and you say that you are not certain of anything, and cannot know anything clearly and distinctly until you have achieved clear and certain knowledge of the existence of God. It follows from this that you do not yet clearly and distinctly know that you are a thinking thing, since, by your own admission, that knowledge depends on the clear knowledge of an existing God; and this you have not yet proved in the passage where you draw the conclusion that you clearly know what you are (1984, p. 89).

This is the *Cartesian* Circle, Descartes's apparent justification of indubitability as a criterion for judgment on the basis of the indubitability of his belief in God.

following Frege (1918), Ramsey (1991; 1999), and Tarski (1944), among others, '*P* is true' and *P* are logically equivalent.[7] Since '*P* is true' and *P* are logically equivalent, saying that '*P* is true' makes *P* safe to hold is equivalent to saying that *P* protects itself. But *P* cannot make itself safe to hold unless Robinson is right, and it cannot be safe to hold that Robinson is right without a compelling reason to make it safe to hold that he is right. Therefore truth cannot be the property that makes something worth believing unless there is nothing wrong with circular arguments. Thus Robinson's criticism of noncircularity yields an account of good reasons that is itself circular, the permissibility of which is precisely the issue.

Of course the circularity of a sequence of positions is not a problem for Robinson, and so the equivalence of '*P* is true' and *P* is not a compelling reason to reject it.[8] What is a compelling reason, however, is the one David Sanford (1972) gives for holding circular arguments to be fallacious. Echoing Aristotle's position that good arguments ought to proceed from the more familiar to the less, Sanford points out that if you prefer that good arguments increase the (subjective) certainty of a position–that the aim of arguing is to be more certain of a position–then you must hold circularity to be vicious. The same goes for any substitute for Sanford's subjective certainty you might favor–familiarity, objective knowability (Biro, 1977; 1984; 2008), warranted assertability (Dewey, 1925; 1941), empirical adequacy (Van Fraassen, 1976; 1980), conformity with scripture, commonsense plausibility and so forth. As long as you prefer that an argument give reasons for a position that possess a property unpossessed by that position, you cannot condone circular reasoning. To this epistemic account I would add only that truth cannot count among the reasons one might prefer a

[7] I take a deflationary, "semantic" theory of truth for granted, here. For a concise yet thorough defense, a good place to start is Paul Horwich's *Truth* (1998).

[8] Several philosophers (Sanford, 1972, 1977; Biro, 1977, 1984; Biro & Siegel, 2008; Sinnott-Armstrong, 1999) have recognized that truth and consistency are neither necessary nor sufficient conditions of good reasons. Though there are differences among them, all of these philosophers agree that circular reasoning is not a formal fallacy, but an epistemic fallacy. Sanford and Sinnott-Armstrong believe that reasons are circular only subjectively, that is, with respect to the reasons a person does have, as well as the reasons that person ought to have on the basis of what follows from them. Biro believes that reasons are circular only if they are objectively circular, that is, with respect to the reasons a person ought to have, according to objective standards about what is right to believe.

position possess because truth is not a property.

Still, some philosophers propose to demote the preference for noncircularity for other reasons. *Coherentism*, as I define it, is the position that because one cannot consistently prefer rational protection, noncircularity, and finitude, the preference for noncircularity ought to be demoted with respect to the others. Coherentists like Lawrence Bonjour (1976) effectively propose to play a game in which individual positions are safe only by membership in a sequence or formation of positions, and not by virtue of inheriting safety from other individually safe positions. Safety, to use Peter Klein's (2003) metaphor, is not passed from position to position like a basketball is passed from player to player. And although safety may still be asymmetric, meaning that if A protects B, then B cannot protect C, and irreflexive, meaning that no position can stand in its own self-defense, if it is not transitive, then it does not follow that every safe position must be one of infinitely many safe positions.

Coherentism is informed by the fact that probabilities less than one are intransitive. Because conditional probabilities less than one are not transitive, and because many apparently good reasons provide probabilistic support for their conclusions, many apparently good reasons are not transitive. Suppose that the probability of A given B is .7, and the probability of B given C is .7. It does not follow that the probability of A given C is .7. In fact, assuming that A, B, and C are distinct, the probability of A given C is less than .7. The thrust of the objection is that because rational protection, like resemblance, comes in degrees, reasonhood, like resemblance, must be intransitive.[9]

In fact, it is not necessary to discuss probabilistic reasonhood to see that some good rea-

[9] In a recent paper, Atkinson and Peijnenburg (2009) demonstrate that in general, the more reasons there are for some position, the less its probability depends upon the probabilities of any individual reason for it. The more conditional probabilities of a sequence of positions you know, the less your knowledge of any unconditional probabilities matters. They claim that this is significant to the extent that most arguments are inconclusive arguments, arguments that offer only compelling or decisive reasons for a given position. But Peijnenburg and Atkinson's demonstration, while elegant, is orthogonal to the question of the transitivity of reasonhood. Peijnenburg and Atkinson do not show that a position becomes more safer to hold as reasons for it accrue. They show that the probability that a position is safe to hold, whatever it is, becomes increasingly determinate with the addition of more and more reasons. It is an argument about the determinacy of the probability of a position's safety, not its actual probability.

sons are intransitive. Consider a variation of an objection to the transitivity of reasonhood proposed by John Post (2009). Start with *B*, the position that Sam is a bartender. *B* makes it safe to hold *C*, the position that Sam can make a whisky sour. Now add *A*, the position that Sam has forgotten everything he learned in bartending school.

A Sam has forgotten everything he learned in bartending school.

B Sam is a bartender.

C Sam can make a whiskey sour.[10]

This is a case in which *A* makes it safer to hold *B*, and *B* makes it safer to hold *C*, but *A* does not make it safer to hold *C*. In this case, reasonhood is intransitive because *A* is a compelling for *B*, *B* is a compelling reason for *C*, but *A* is not a compelling reason for *C*. In fact, as Post notes in his discussion, *A*, the position that John has forgotten everything he learned in bartending school, makes it *less* safe to hold *C*, the position that Sam can make a whiskey sour. Post is right because neither compelling reasons, which make a position safer by ruling out some relevant alternative, nor decisive reasons, which make a position safer by ruling out every relevant alternative, are closed under deduction, which means that neither compelling nor decisive reasons are transitive. And so at least some good reasons are intransitive, which means that not all circularity need be vicious. If circularity is conditionally permissible, the condition being that the circular reasons in question are either compelling or decisive, then at least some positions may be safe without infinitely many reasons to protect them.

And yet conclusive reasons, which are closed under entailment and therefore transitive, are, relatively speaking, the best reasons of all. They offer a position safety beyond reproach. If the coherentist, or anyone who suggests demoting the preference for noncircularity with respect to the others, is willing to grant that there are no conclusive reasons, reasons that rule

[10] Thanks to Nathaniel Gindele for this variation.

out every alternative to one's position whatsoever, then the preference for noncircularity may be demoted without cause for concern.

The chief objection to doing without conclusive reasons is that they play a role in actual arguments against the transitivity of reasonhood.[11] John Post's (2009) counterexample to transitivity, of which the above case is a variation, proceeds as follows. Start with *B,* the belief that Sam is a bartender. *B* justifies *C,* the belief that Sam can make a whisky sour. Now add *A*, the belief that Sam is a bartender who has forgotten how to make a whiskey sour.

A Sam is a bartender who has forgotten how to make a whiskey sour.

B Sam is a bartender.

C Sam can make a whiskey sour.

In this case, *A* is not only both a compelling and a decisive reason for *B*; it is a both compelling and decisive because it is a conclusive reason for *B*. As every deductive reason is conclusive, it is conclusive by virtue of its implication of *B.* And to do without conclusive reasons is to do without deduction, which I assume to be out of the question.

In short, the objection to the position that good reasons cannot be circular is compelling only if you remove conclusive reasons from the equation. But if conclusive reasons are good reasons, then their transitivity leaves intact the apparent inconsistency of preferences for rational protection, noncircularity, and finitude. You cannot consistently prefer conclusive reasons to any other sort of reason if you also prefer that a given position be made safe by a finite sequence of new, safe positions. Skepticism is not just doubt about the possibility of a position's safety, in general. It is the position that conclusive safety, in particular, is

[11] Another objection to doing without conclusive reasons is that compelling reasons and decisive reasons are defined with respect to them, as the notion of a relevant alternative only makes sense against the background of a logical alternative.

impossible. The question, then, is whether there is any way to respond to the skeptic without sacrificing conclusive reasons and the safety they offer.

2.4 Finitude & Infinitism

The preference for finitude is the preference that a sequence of reasons for a given position not regress *ad infinitum*. To object to it is to hold that sequences of reasons can make a position safe to hold despite going on forever. If the objection is compelling, then the preference for finitude ought to be demoted with respect to preferences for rational protection and noncircularity. As I have mentioned, Aristotle's rationale for the principle of finitude is that no infinite sequence of steps in a proof can be proved, which means that either there are some things that can be known without proof, or that some proofs are circular. The Pyrrhonian skeptics, who were no doubt writing with Aristotle in mind, likely intended their own principle of finitude to similarly prohibit infinite sequences of reasons. The problem, however, with a preference for finite sequences of reasons, as Aristotle and the Pyrrhonian skeptics alike recognized, is its inconsistency with preferences for both rational protection and noncircularity. It is straightforwardly provable that if every conclusively safe position is made safe by a new, safe position, then every conclusively safe position must be one of infinitely many safe positions.

Until recently, no one but Peter Klein (1998; 1999; 2003) has attempted to resolve this apparent inconsistency by demoting the preference for finitude. Klein holds that a position is safe if and only if it is a member of an "endless, non-repeating" sequence of safe positions. He calls this position *infinitism*, and holds that the best reasons against it are, contrary to tradition, not genuinely compelling.

Klein's version of infinitism is a thesis about the necessary and sufficient conditions of a conclusively safe position. But as such, it is too strong. In fact it is easy to give endless, non-repeating sequences of reasons that are not sequences of good reasons at all. David Sanford

(1972) provides a simple way to do so. If you want an endless, non-repeating sequence of reason for something, simply double-negate what you are supposed to give a reason for, and give the double negation as a reason. Apply the rule infinitely many times to generate an endless, non-repeating series of reasons that fails to make anything any safer to hold. In Lewis Carroll's "What the Tortoise Said to Achilles" (1895), Achilles does something similar. The Tortoise asks him to defend his use of *modus ponens* in a geometrical argument without appealing to it–or any other related logical principles that appear to depend on it. He quickly gets himself into a regress of premises, the first a reason for his use of *modus ponens*, the second a reason for the first, and so on. In neither case are the reasons provided good reasons by virtue of their membership in an infinite, noncircular series of reasons. They are, if anything, bad reasons.

As Aristotle was remarkably close (but not close enough) to recognizing in his discussion of the infinite divisibility of finite magnitudes (*Phys.* 206A25–B3), there are two kinds of infinite sequences. Some sequences, like $1, 2, 3, 4, \ldots$, *diverge* as the number of their terms goes to infinity. Others, like Zeno's $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$, *converge* as the number of their terms goes to infinity. Both sequences are infinite, but only the second of them approaches a finite limit. As Nicholas Rescher (2010) points out, a regress of reasons need not be vicious if it converges on a finite limit. If the preference that reasons not regress *ad infinitum* entails only that sequences of reasons converge upon some finite limit–and not that all sequences of reasons be finite in length–then one can in principle consistently prefer positions with the noncircular protection of infinitely many safe positions, each safer than the next.

The fact that some infinite sequences converge upon finite limits is a compelling reason to hold that, contrary to appearances, one can consistently prefer rational protection, non-circularity, and finitude after all. In sum, there are two ways to understand what it means to have a preference for finitude in the reasons for a position. The first is that a position is safe only if there are finitely many safe reasons making it so. Following Aristotle, this is how philosophers have traditionally understood what it means to prefer that sequences

of reasons not regress *ad infinitum*. Because the principles of rational protection and non-circularity imply that every safe position must have the protection of infinitely many safe positions, it is this way of understanding the preference for finitude that yields the skeptical paradox, which in turn motivates the selection of at least one preference to demote with respect to the others.

The second way to interpret finitude is that there are any safe positions only if they have the rational protection of a convergent sequences of reasons, reasons that get safer and safer as they approach a finite limit, the limit being defined by whatever epistemic properties characterize the nature of a good reason in a particular case. I have already mentioned some of those properties, describing them earlier as preferences like any other: familiarity, subjective certainty, objective knowability (Biro, 1977; 1984; 2008), warranted assertability (Dewey, 1925; 1941), empirical adequacy (Van Fraassen, 1976; 1980), conformity with scripture, commonsense plausibility and so forth. Interpreting finitude this way upends the position that at least one of the three other preferences must be demoted.

Interpreting the preference that the reasons for a position not regress *ad infinitum* this way preserves the integrity of rational protection, noncircularity, and finitude without demoting any one of them to secondary status. But the mere consistency of these preferences does not yet mean that any position in particular, let alone any serious, controversial, philosophical position, qualifies as conclusively safe to hold. In fact, this interpretation is problematic for just this reason. First, it is difficult to come up with even one example of a sequence of reasons that meets all three requirements. Second, and perhaps even more significantly, it is difficult to believe that any example one might come up with would itself have the support of infinitely many converging reasons for holding it to be a good example.[12] Thus a milder form of skepticism remains a live option even if there are no compelling reasons for the strongest of skepticisms, one that holds conclusive reasons and thus conclusive

[12] Andrew Cling (2004) rightly points out that the infinitist has the burden of providing at least one example of an infinite sequence of good reasons, and thinks it a decisive mark against infinitism that no example is to be found.

safety to be impossible.

And yet I hesitate to concede that this milder form of skepticism is anything epistemologists ought to worry about. Epistemology's chief normative goal is produce guidelines for holding a position, and there is all the difference in the world between guidelines, what I have described as the preferred ground rules for an epistemic game, that cannot be followed because of their inconsistency, and guidelines or ground rules that cannot be followed perfectly because of the practical difficulties of meeting them. The former cannot function as guidelines or ground rules at all, while the latter can function as guidelines or ground rules perfectly well. Even if there are no infinite sequences sequences of good reasons, there are still longer and shorter sequences. In practice, the preference for finitude requires only that of two sequences of reasons that get safer and safer with each additional reason, the longer is to be preferred.

In this respect, the fact that there are may be no infinite sequences of reasons to make a position conclusively safe to hold is like the fact that there not any perfectly aerodynamic automobiles. There are better and worse sequences of reasons, just as there are more and less aerodynamic vehicles. In neither case does the nonexistence of anything that fully realizes the ideal in question justify skepticism about relative quality, which is in any even what matters to the rationality of one's decisions, whether those decisions be about what to believe or what to drive on a windy day.

2.5 Epistemology and the Evaluation of Epistemic Preferences

I have argued that the preferences for rational protection, noncircularity, and finitude are in fact consistent with one another, which means that there is no compelling reason to hold that the conclusive safety of a given position is logically impossible. I recognize, however, that this leaves open the possibility that there may no conclusively safe positions, even if conclusively safety remains intact as a logically possibility. It may be important that a posi-

tion be a logical possibility, but its logical possibility does not make it a live option. If there really are no conclusive reasons, then decisive reasons and decisive safety may be the best that any one person can do in practice.

Recasting the preference that sequences of reasons not regress *ad infinitum* refocuses inquiry on the epistemic properties that sequences of reasons are supposed to possess in greater and greater amounts. I have assumed that whatever these properties are that one one might prefer, their rationality–how safe it is to prefer them–is entirely a matter of consistency with one's other preferences. And yet one might object that some epistemic preferences are bad to have: that for a reason in a convergent sequence of reasons to be a good reason, for a sequence of reasons to make a position safe to hold, the sequence must converge upon something *good*.[13]

Some epistemic preferences, things like certainty, provability, empirical adequacy, Cartesian indubitability, and warranted assertability, have stronger reputations than others, properties like scriptural conformity, popularity, and originality, for example. The question, then, is how to compare competing epistemic preferences if epistemic preferences ultimately determine what makes for a good argument, and what positions are safe to hold. Quine (1969a) once said that with the failure of foundationalism, epistemology falls into line as a chapter of psychology. But Quine was a pragmatist too, and as he and the pragmatists have long emphasized, epistemology, the attempt to discern which epistemic preferences are good to have and which are not, is–in the limit–an ethical inquiry, an attempt to figure out what our preferences ought to be more generally. If this is right, then epistemology falls into line as a chapter of ethics as well.

But if questions about what our preferences ought to be more generally await at the end of the epistemological endeavor, then there is another problem altogether. The problem

[13] Scott Aikin (2011) makes this same point, that that membership in an infinite sequence of reasons does not make something a good reason. But while Aikin makes the case for empirical givenness as the property upon which sequences of good reasons must converge, my own instincts are that any unified account of rational goodness is bound to fail.

is that unless there are any preferences that are categorically good to have, and not merely good by virtue of their consistency with other preferences, there can be no absolute measure of what makes something a good argument, and if there can be no absolute measure of what makes something a good argument, then there are no categorically good arguments, no irresistible reasons that make a position safe to hold no matter what.

An example illustrates the problem. Imagine a disagreement between an evolutionary biologist and a biblical literalist about which biology textbook their school district should adopt for the ninth grade. Both agree that the textbook should contain positions with non-circular rational protection that avoids regressing *ad infinitum*. But the evolutionary biologist wants a textbook the claims of which are increasingly safe to hold because of their empirical adequacy, while the biblical literalist wants a textbook the claims of which are increasingly safe to hold because of their conformity with the book of Genesis. That is, the evolutionary biologist and the biblical literalist agree about everything except the what property good reasons should possess, what sorts of positions their children ought to prefer to have. As such, they fundamentally disagree about how disputes about what positions to hold ought to be settled, and so if they are to come to some rational resolution, it must be by appeal to something they both agree to be more preferable than either empirically or biblically adequate beliefs.

The problem with relying on *de facto* practical criteria to determine which sorts of positions are safest to hold–when push comes to shove over epistemic preferences–is that *de facto* criteria provide only *de facto* guidelines. If the evolutionary biologist and the biblical literalist do not share other, more fundamental preferences, like the preference that one's child attend an elite academic institution–if the biblical literalist simply changes her mind and decides that what she wants most is for her child to assimilate into a particular religious community upon graduation from high school, the appeal of admission into an elite university ceases to function as way to determine what sorts of positions are safest to hold. This is where the categoricity of epistemic preferences comes into play. If there are some things

that ought to be preferred no matter what, then these things can function as the absolute measure of the desirability of various sorts of belief. If not, then not.

I should like to conclude with two points of reflection. The first is that, as I began the paper I am not inclined to believe in the existence of any categorical goods, epistemic or otherwise. This means that I am not inclined to believe in the existence of any rationally irresistible reasons, and thus not in the existence of any argument the conclusion of which must be accepted no matter what. The second point is that even if there were categorical goods, objects of desire capable of standing as undeniable practical reasons for some positions over others, there is no guarantee that these criteria would weigh in favor of philosophically reputable epistemic preferences like certainty, provability, empirical adequacy, Cartesian indubitability, and warranted assertability and not others like conformity with scripture, popularity, and originality. To believe so seems to reflect philosophical hope more than it does anything else, a Platonic conviction that the good and the true, in the end, must be one. But what justifies this hope, that things like certainty, provability, empirical adequacy, Cartesian indubitability, and warranted assertability–and not others like resonance in scripture, popularity, and originality–are not only the essence of truth but are categorically desirable because of it? Is it, as Joseph Raz asks, "that people who flout reason sometimes prosper[?] Is the desire for some further vindication of reasons a hope that philosophical argument can show this to be an illusion?" Raz is right: "there is no illusion there." (2011, p. 101) Even if there were categorically good reasons, there is no guarantee that they would reveal the success of fools to be anything but a fact of life.

3

The Miracle of Rational Persuasion

> Perhaps philosophers need arguments so powerful they set up reverberations in the brain: if the person refuses to accept the conclusion, he *dies*.
>
> - Robert Nozick[1]

This chapter is about why bad things happen to good arguments. Good arguments are arguments that give good reasons for their conclusions, and the bad things that happen to them are that these conclusions go ignored, rejected, or otherwise unaccepted. It is widely believed that they chiefly suffer this fate because the best any argument can do is show you what you ought to believe on the condition that you recognize the merits of the reasons it gives. But without an account of how you come to accept good reasons on their merits, rational persuasion has the unmistakable appearance of a mental miracle. I believe in the power of reason, but I cannot take it on faith alone.

Like all miracles, rational persuasion is offered as an explanation of some phenomenon without an explanation of its own. But it has other miraculous features as well. It is, even for

[1] Robert Nozick, *Philosophical Explanations*. (Belknap Press, 1981). p. 4

believers, an exceedingly rare occurrence. It conforms to the predictions of wishful thinking about human beings and their special place in the world. Evidentially, there is nothing unequivocal to say in its favor, and there is much to say against it. Thus it is not simply the lack of an explanatory account that is the problem. There are many things we have a right to believe in without understanding, but rational persuasion is not one of them. Intellectual honesty requires facing the possibility that good arguments fail to persuade because they are incapable of doing so.

I have two major aims in this paper. The first is to raise doubts about the possibility that good arguments ever persuade. The second is to remove those doubts by demystifying the process of rational persuasion. To do so, I will explain how arguments persuade with good reasons while preserving the independence of those reasons from their psychological effects. First, I argue that rational persuasion is postulated as an inference to the best explanation, but without an explanation of its own. Second, I argue that rational persuasion demands an explanation. Third, I argue that rational persuasion requires intellectual character. And finally, I explore some consequences of this view both within philosophy and without.

3.1 Rational Persuasion is Postulated without Explanation

Arguments are sets of sentences, one of which is the conclusion. The others are premises. There are deductive arguments, which, if valid, have premises that guarantee the truth of their conclusions. There are also inductive arguments, the premises of which, if true, make their conclusions more likely. Deductive or inductive, all good arguments have premises that support their conclusions. This is a logical virtue. Some logically virtuous arguments are good in nonlogical ways as well. Whether these arguments are nonlogically virtuous depends on the reasons for using them. If the purpose is justification, an argument's premises should justify its conclusion. If the purpose is explanation, an argument's premises should

explain its conclusion. If the purpose is refutation, an argument's premises should imply a contradiction. The virtues these arguments have, while nonlogical, are like logical virtues in being purely intellectual. An argument's possession of intellectual virtues does not depend on whether or not anyone believes its conclusion.

Some arguments with intellectual merits are used for other purposes. They are used to impress, or to entertain, or to persuade. These uses are psychological.[2] Their purpose is to affect or, in the case of persuasion, to change a person's mind.[3] Intellectual and psychological purposes are independent of one another. Not all good arguments are persuasive, and not all persuasion is accomplished with good arguments. But in attempts at rational persuasion, intellectual and psychological purposes overlap. Rational persuasion, as I define it, is using a good argument to change a person's beliefs by the merits of the reasons you provide. It means securing someone's acceptance of the reasons an argument gives because they are good reasons, converting reasons one ought to have into reasons of one's own. Socrates, as he appears in Plato's early dialogues, is an example of someone who makes attempts at rational persuasion. He has both intellectual and psychological purposes for arguing. He tries to persuade using good arguments. Unlike many contemporary philosophers, his intellectual purposes are in service of his psychological purposes. He is not concerned with refutation for its own sake.[4] He refutes his interlocutors, or tries to refute them, to persuade them.

A good reason is something you ought to accept, but that you ought to accept something does not imply anything about your actually doing so. This may be obvious to many. Hume,

---

[2] Frank Jackson (1984) distinguishes two purposes of arguing, a *teasing-out* function and a *convincing* function. I prefer to think of the *teasing-out* function as the logical purpose of arguing, with this logical purpose sometimes serving intellectual (e.g. justifying, explaining, and refuting) purposes, and sometimes serving psychological (impressing, entertaining, and persuading) purposes.

[3] You can change a person's mind by changing beliefs or desires, though I am concerned here only with changing beliefs.

[4] Under this classification of refutation as an intellectual activity, Socrates is more like a psychotherapist than he is a contemporary philosopher, while contemporary philosophers who are not concerned with persuasion are more like ancient sophists, who were concerned with refutation only.

after all, prohibits deducing, without explanation, *ought*-statements from *is*-statements. Max Black (1964) calls this principle "Hume's Guillotine" for its severance of what ought to be from matters of fact. But the fact that showing you what you ought to believe cannot on its own make you believe it poses a complementary problem. Just as we must explain the normative influence of how things are on how they ought to be, we must explain the psychological influence of good reasons on actual thinking.[5] In either case, without an explanation we have no right to assume that there is any influence at all. Hume's Guillotine cuts both ways.[6]

The need to explain the conversion of reasons you ought to have but do not appreciate into reasons you ought to have and do appreciate is easy to overlook. Even philosophers explicitly concerned with the persuasive futility of good arguments overlook it. For example, many philosophers have recently come to ask whether and how you should revise your beliefs in light of disagreement with an epistemic peer, someone with as much credibility as you. Adam Elga (2007) claims that you should change your mind, cutting your confidence in half, while Thomas Kelly (2010) claims that you should change your mind only if the total available evidence does not support your position. These philosophers, like others interested in the phenomenon of peer disagreement, are concerned to establish good reasons for changing your mind. But they are less concerned to say how you might come to assimilate those reasons into your own thinking, or whether it is even possible.[7]

[5] See both MacIntyre (1959) and Flanagan (1993, p. 59). MacIntyre and Flanagan each defend the thesis that Hume's point in proscribing *is-ought* deductions is to create space for his attempt to explain exactly what connects what is and what ought to be.

[6] W. V. Quine (1969b) notices the need to explain the conversion of good reasons into reasons of one's own. For Quine it is a part of *naturalizing* epistemology, which he thinks takes two steps. The first step is to define good reasons as those respected within the sciences. The second is to recast epistemology as the psychological study of how we come to believe scientific theories. But because he does not appear to justify his definition of good reasons as respected within sciences, Quine's first step appears to violate Hume's is-ought proscription. Thus the stock objection against Quine's program is that it runs into the leading edge of Hume's Guillotine (Cf. Putnam (1982) and Kim (1988)). Quine's first step is too big a stumbling block for many philosophers to take the second with him, the step of explaining how objectively good reasons come to be respected as reasons of one's own. Few ever consider how we come to believe what we ought to believe because few agree with Quine's account of what we ought to believe.

[7] Of those active within the debate on peer disagreement, only Peter van Inwagen (2006; 2010) appears to

The explanatory gap between the changes a good argument recommends and your adoption of those recommendations is to be found wherever arguments are made. This goes for both public argumentation between individuals and private reasoning alike. But without an explanation of how one comes to respect an argument's intellectual merits, one must wonder whether we are ever persuaded by the appreciation of good reasons at all.

One may object there are many things that warrant belief without requiring much in the way of an explanation from a believer. Human beings knew that water causes iron to rust, for example, long before possessing explanations of oxidation and the process of corrosion. Thus some may balk at my claim that not understanding how good reasons change minds is reason to doubt that they ever do. My reply is to grant the point. While explanations cannot stand for long without explanations of their own, the objection is correct that lacking an explanation of rational persuasion is not sufficient to undermine belief that good arguments ever change minds by way of their intellectual merits. This is because the postulation of rational persuasion is intended to be what Harman (1965) calls an "inference to the best explanation." Inferences to the best explanation are merely better than other relevant hypotheses. An explanation does not need an explanation of its own to be better than other relevant hypotheses. If rational persuasion best explains what it is supposed to explain, then a failure to understand it in full can be forgiven unless there are other grounds for doubt. Unfortunately for the faithful, there are indeed other grounds for doubt.

3.2 Rational Persuasion Demands an Explanation

The lack of an explanation is just the first problem for faith in rational persuasion. Another is that the range of human behavior that rational persuasion is itself supposed to explain is on any account very small. No one suggests that all or most of what we believe or desire is a result of reasoning or arguing, and very few, I suppose, would claim that the times we seem to change our minds because of good reasons are anything but special occasions in

worry that rational persuasion is not possible.

our mental lives. The trouble with special occasions is that they call for special evidence. One's burden of proof varies directly with the objective probability of what is to be proved, and the objective probability that any given argument persuade by way of its intellectual merits is by any estimation quite low.

Another problem for rational persuasion is how it fits all too well with a self-congratulatory image of human psychology. Within this image the glow of reason, even if it exists in unfrequented corners of the soul, sheds light on the better angels of our nature. But wishful thinking is common, and claims that conform to it must be taken with a grain of salt. This is nothing but an alert to the dangers of motivational bias. As one's burden of proof varies directly with objective probabilities, so it varies with how much there is to gain from espousing what is to be proved. And human beings, out of sheer vanity, have much to gain from belief in their own friendliness to reason. In its conformity to the predictions of wishful thinking, being persuaded by good reasons is like Kant's conception of acting out of moral duty. Both are difficult to explain and exhibited seldom enough to raise the worry that we believe in them because it is flattering to do so.

There is another way in which rational persuasion is like acting out of moral duty. There is no unequivocal evidence that it ever happens. Every allegedly dutiful act is consistent with a competing egoistic rendering. Similarly, every appearance of rational persuasion may be described without crediting the quality of an argument's reasons for a change in mind.

Suppose you have a friend who says there is a seventy percent chance that a particular candidate is going to win an upcoming election. You ask her why she is so confident in the outcome, and she tells you that everyone she knows is voting for that candidate. You remind her of the dangers of unrepresentative sampling, and point out that her friends are not collectively representative of the electorate. Your friend responds by downgrading her assessment of the election's likely outcome. She now agrees that the candidate's chances of winning are not nearly as great as she thought they were.[8] This case appears to be a

[8] Thanks to Owen Flanagan for the example.

paradigm of rational persuasion, but the problem is marshaling unequivocal evidence to describe it that way. Your argument makes clear what your friend ought to say, and she says it. But even granting that your friend really does change her mind, it does not follow that your argument's intellectual merits deserve the credit. She may have come to change her mind because of your skeptical tone, or because of the confidence you project, or because of your intelligent-sounding reference to statistical sampling. None of these have anything to do with the actual quality of the reasons you give.

An explanation like this is *Machiavellian*.[9] Machiavellian accounts of rational persuasion acknowledge that some arguments give good reasons for their conclusions, but deny that these reasons every persuade us. The mystery of their apparent psychological efficacy is solved by giving credit for any given argument's persuasiveness to objective but nonetheless intellectually impertinent features of the argument. Good arguments are persuasive, but only *extrinsically* so and never because of the quality of the reasons they give. The Machiavellian about rational persuasion is like the skeptic about Kantian morality who accepts the independence of the moral from what is in our own self-interests, but denies that we ever act out of duty. While it is difficult to name a major philosopher who thinks all apparent cases of rational persuasion are better explained by a Machiavellian account, the psychologist Jonathan Haidt's (2001) account of the persuasiveness of good moral arguments is Machiavellian. Haidt thinks that, at least as far as morality and politics are concerned, it is not the quality of an argument that persuades but the mask of intelligence and trustworthiness crafted by arguing that does the trick.

Machiavellian explanations are not the only alternative to the postulation of rational persuasion to explain an apparent change of mind. There are also *psychologistic* explanations. Psychologism denies that there are such things as good reasons apart from the reasons a person actually appreciates. The mystery of how apparently good arguments persuade is

[9] Frans de Waal (2007) calls the suite of cognitive abilities that provides social animals the ability to manipulate other group members *Machiavellian*, thus the label of *Machiavellian intelligence* for any faculty with the function of managing one's social standing.

Table 3.1: Explanatory Options

Good Arguments	Intellectual Merits → Persuasive	Persuasive → Intellectual Merits
Intrinsically Persuasive	Rational Persuasion	Psychologism
Extrinsically Persuasive	Machiavellianism	N/A (Machiavellian Psychologism)

solved by reversing the direction of explanation and simply defining good reasons as the reasons a person actually finds persuasive. Good arguments are *intrinsically* persuasive, but their persuasiveness is what makes them good arguments. Psychologistic philosophers are like ethical egoists who deny the independence of the moral from what is in our interests and hold that we ought to act selfishly. John Stuart Mill's (1882)account of mathematics is psychologistic. Mill accepts that good mathematical arguments are persuasive, but defines them as good arguments on the basis of their being compelling to us.[10]

Machiavellianism and psychologism are two complementary ways of denying that rational persuasion exists.[11] They do not explain rational persuasion so much as they explain it away. As competing, independently plausible, and equally explanatory accounts of what happens when a good argument appears to change a person's mind by the merits of the reasons it gives, they void the claim that rational persuasion can stand as an explanation of anything without an explanation of its own. At the same time, they provide clear desiderata for any true explanation of rational persuasion. To avoid Machiavellianism and psychologism, any adequate explanation must say how good arguments are capable of changing minds by way of their intellectual merits while preserving the independence of those merits from their psychological effects.

In denying that there is any unequivocal evidence for rational persuasion I intend to discount the feelings we have in the midst of reasoning and argumentation. We may un-

[10] Thus mathematical inferences are valid because of their conformity to principles of empirical thought. Quine's account of what makes a good argument is psychologistic as well, though not as baldly as Mill's.

[11] There is a fourth possibility, another version of psychologism, but it is a non-contender. Call it *Machiavellian psychologism*. (See Table 3.1 on page 51) It is a non-contender because it says that an argument is good if you are convinced by obvious intellectual irrelevancies like the color of the ink in which it is written or the gender of the person giving it. This an *idea non grata*, that rare view no philosopher ever seems to have defended. I do not intend to be the first.

equivocally feel that our minds have been changed by good arguments without our feelings unequivocally supporting the conclusion that we have experienced rational persuasion. And *vice versa*. Anyway it would be odd for changes in belief, which are qualitatively thin when compared with mental states like desire, joy, lust, anger, sadness, and pain, to manifest themselves as conversion experiences. The fleeting feeling that your mind is changing is not an unambiguous glimpse into your mind laid bare. It is barely a glimpse at all, the gestalt impression of a vague and untitled mental event.

There is no unequivocal evidence for rational persuasion, but there is a wealth of evidence against it. The most studied failure of rational persuasion results from what Peter Wason calls *confirmation bias*.[12] Confirmation bias is postulated to explain the appearance of wishful thinking. It is a tendency to favor information that supports your beliefs over information that conflicts with them. The evidence for confirmation bias in human belief formation is evidence against rational persuasion in two ways. First, it suggests that good arguments are unlikely to cause you to give up something you believe or to change your mind from one position to another. Second, it suggests that the strengthening of belief that may result from reasoning and arguing, while qualifying as persuasion, does not qualify as rational persuasion. These changes in mind are the work of recognizably bad reasons, not good ones.

Take any space of discourse and you will find good arguments met with a mix of apathy and antipathy. Even mathematics witnesses failures of rational persuasion. Pythagoras is remembered most for the theorem that bears his name. But legend has it that the theorem was not something that Pythagoras and his disciples were proud of. It implied that some numbers could not be expressed as the ratio of integers, which they thought was absurd. To keep irrationals secret, they drowned Hippasus, who discovered them. This example aside,

[12] Cf. Wason (1959, 1960); Wason & Johnson-Laird (1972)

mathematical dissent seldom meets with such resistance anymore. Maybe this is because ever since Hippasus' watery demise mathematicians are no longer expected to believe everything they prove. You can take theorems of mathematics to hold hypothetically without any real conviction behind them. Under this description it is only a professional suspension of belief, perhaps brought on by Hippasus's martyrdom, that is responsible for the mathematician's apparent openness to reason.

Maybe I am preaching to the choir. Philosophers experience failures of rational persuasion often enough to worry that wherever beliefs are in place, good arguments are powerless to change them. Along these lines Peter van Inwagen (2006) recommends against defining successful philosophical arguments by their conversion of unbelievers in part because philosophers cannot even convince one another of things like the existence of motion, or other minds, or the external world. Similarly G.E. Cohen (2000) notices that we often defend our convictions against good arguments with the full awareness that we likely hold them because of upbringing alone. We fail to be psychologically responsive to what we ourselves consider reasons for changing our minds.

In a fitting twist, the impossibility of rational persuasion would at least vindicate this experience as an inference to the best explanation of it. Arguments might still appear persuasive in mathematics and the sciences, but these appearances could be written off as illusions, false miracles explained by the mathematician's and scientist's good sense to stop arguing when they reach agreement. That consensus gathers around something that has been proved does not imply that consensus has gathered because it has been proved. Maybe, as Max Planck writes, "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it" (1949, pp. 33-34). And maybe, if arguing is just something we do while waiting around for each other's funerals, we really ought to find something better to do.

3.3 Rational Persuasion Requires Intellectual Character

The case against rational persuasion channels Humean skepticism about the existence of miracles. Rational persuasion is unexplained, improbable, and inferred on the basis of suspect testimony by individuals with a lot to gain from its postulation. There is no direct evidence for it, and much against. Therefore it would be unjustified, and anyway unwise, for the believer in the power of reason to await a savior in a fell swoop of refutation. Redemption is better sought in small increments.

Redemptive accounts of rational persuasion often try to make good arguments persuasive by adding that one must be of the right mind to be compelled by good reasons. In the *Republic*, for example, Plato says that you must have a passion for reason for good argument to change your mind. Plato proposes that an appropriate passion for reason is what individuals like Callicles and Thrasymachus lack. Sometimes this character is described as a responsiveness, and sometimes, like a taste for good coffee or fine wine, it is described as a kind of sensitivity. He postulates the existence of intellectual character to explain why a person comes to believe something for good reasons, and a lack of character to explain why not.

It is not lost on Plato that the postulation of intellectual character has all the advantages of theft over honest toil. If you make good arguments persuasive by appealing to a person's dispositions, then you have to explain how those dispositions are formed. Thus Plato devotes much of the *Republic* to the formation of various personality types. His account involves two major hypotheses. The first is that conditions outside your control–political climate, family structure, and education–determine what you have a passion for. The second is that what you have a passion for determines your character. You have intellectual character if you have a passion for changing your mind for good reasons. For Plato, belief acquisition is a function of character, and intellectual character is a function of having a passion for good reasons over bad.

Along similar lines, and eschewing *eros* for the more sober *respect*, Kant holds that you must have a respect for moral reasons, what he calls a good will, to act morally. In the Kantian vernacular, one may think of intellectual character as a respect for the reasons one ought to respect. At any rate, accounts like Plato's and Kant's reduce the problem of making good arguments persuasive to the problem of acquiring an appreciation for good reasons. But there are two ways you can appreciate something. You can appreciate it for its own sake, as something with intrinsic value, or you can appreciate it as a means to an end, as something with extrinsic value. You can appreciate good reasons for being good reasons, in which case you do have a passion or respect for an argument's intellectual merits, or you can appreciate them for the wrong reasons, in which case you are concerned with intellectual irrelevancies. It is only when you appreciate good reasons for their own sake, when you believe what you ought to believe because you ought to believe it, that rational persuasion truly occurs.

The problem of making good arguments persuasive boils down to the problem of acquiring an intrinsic appreciation for good reasons. If you have an intellectual character, then you appreciate the reasons you ought to appreciate for their own sake, and if you appreciate the reasons you ought to appreciate for their own sake, then good arguments will be persuasive to you by virtue of their intellectual merits. But as Callicles and the case against rational persuasion demonstrate, there is every reason to doubt that anyone at all appreciates good reasons for their own sake. Were there widespread passion or respect for good reasons, reasoning with one another would be a much better way to change each other's minds than it appears to be. Thus Machiavellianism and psychologism reappear as forms of skepticism about intellectual character.

The Machiavellian concludes from failures of rational persuasion that it must be impossible to appreciate good reasons for their own sake. Because the occasional appearance of a good argument's successful persuasion must be an illusion, no one can really have the character necessary to appreciate good reasons. Intellectual character must be a fantasy.

The Machiavellian encourages a more cynical interpretation of apparent demonstrations of intellectual character. Consider again the situation in which your friend revises her estimate of the election's outcome after you remind her of the dangers of unrepresentative sampling and point out that her group of friends is not representative of the electorate. The Machiavellian credits features of the argument other than the merits of the reasons it gives for her change in mind, features like the tone in which it is presented, your own projected confidence, and the sophistication of the terms you use. Machiavellianism admits that your friend has the disposition to change her mind because of your argument, but denies that this disposition reflects a passion or respect for good reasons. It is a case of social intelligence masquerading as intellect, the pretense of character and not the demonstration of it.

Unlike the Machiavellian, psychologistic philosophers lay blame for the evident weakness of good arguments on the arguments themselves. Psychologism defines good reasons by reference to the reasons individuals actually accept as good reasons. The psychologistic philosopher asks us to understand apparent demonstrations of intellectual character flaws to undermine the merit of the arguments used to demonstrate them. This is how L.J. Cohen (1981) takes on the threat of systematic human irrationality posed by many experiments on human belief acquisition. Cohen denies that a lack of intellectual character can be experimentally demonstrated because human dispositions to form beliefs determine what qualifies as a good argument in the first place. For Cohen the arguments that behavioral decision theorists employ to demonstrate human irrationality are simply bad arguments dressed up as good ones.

For skeptics, the problem of being certain of a person's appreciation of good reasons is insurmountable. It appears as if the believer in rational persuasion has a choice to make. Either try to show that good arguments are indeed persuasive by virtue of the reasons they give despite the strong evidence that they are not, or admit that intellectual character is a fantasy and concede that good arguments are incapable of changing our minds. But this is

a false choice. Good arguments can be persuasive for reasons other than those that make them good arguments without their having to stay that way, if there is a way for a person to acquire an intrinsic appreciation for something. The miracle of rational persuasion is the miracle of turning things we do not initially appreciate into things that we do.

As miracles go, cultivating appreciation for good reasons is less like exorcising demons and walking on water than it is like childbirth and human flight. It actually happens. It must happen because we are not born with an intrinsic appreciation for a great many things at all. Conservatively, pleasure is the only thing we value for its own sake at birth, but even conceived more liberally to include air, food, water, sex, and sleep, the list is short. Unless one denies that we value anything for its own sake but the things we are born valuing, such that everything we do is done to satiate what Plato called our appetites, you need to explain how it is that we come to appreciate anything in life besides those things that qualify as basic human needs. You need to explain novel incentives.

There is an explanation of novel incentives. In fact it is the only explanation there is. Thorndike (1911; 1927) calls it the *law of effect*, and Skinner (1938) identifies it as the central principle of operant conditioning. It says that the robustness of a disposition is a function of the reward value of its consequences. Intensifying or diminishing a disposition by acting upon its consequences is called *reinforcement*, and the stimuli that provide reinforcement are called *reinforcers*. All reinforcers, if they are valuable, are valuable for their own sake. Some reinforcers are *primary*, in that they are innately valuable, and some reinforcers are *secondary*, in that they come to be valuable by conditioned association with primary reinforcers. All secondary reinforcers begin life as things that we value by association. They are converted to secondary reinforcers by conditioning their association with primary reinforcers. This is the only way to cultivate new values, and it is why, as Dennett (1981) says, "the law of effect will not go away."

Reinforcement learning theory explains the acquisition of intellectual character, the possession of which explains how good arguments come to compel us by the quality of

the reasons they provide.[13] Character is a function of the things we appreciate for their own sake, and the things we appreciate for their own sake acquire their value to us by being pinned by conditioning to primary reinforcers. Acquiring an intellectual character requires cultivating an appreciation for good reasons by associating them with other things that we value for their own sake, like pleasure or success.

The test case for this account is the one where your friend changes her mind about the likely outcome of the election because of the argument you give her. If she does change her mind, it is because she accepts the reasons you have given her as good reasons. The question is whether she accepts them because they are good reasons, in which case she has the prerequisite passion or respect for the quality of your argument, or because of other things she may appreciate like the appearance of shrewdness, confidence, or sophistication, in which case she has the wrong sort of character. The answer, ultimately, is that it does not matter. If she does have intellectual character, then we have license to explain her change in mind as a case of rational persuasion. If she does not, thus lacking appropriate passion and respect for good reasons at the moment of your conversation, she can nevertheless acquire that appreciation for good reasons through conditioning, by associating good reasons with things that she does appreciate, so that she may eventually come to find arguments like yours to be persuasive in their own right.

The test case, I believe, represents what happens in mathematics and the sciences, where good arguments appear to persuade more than anywhere else. Let us suppose that life in

[13] Contemporary models of reinforcement learning hold that the mere association of reward with a given stimulus is insufficient to change your disposition to respond to that stimulus in one way or another. Rather, a stimulus must be more or less valuable than you expect for any dispositional change to occur. Deviations from expectation represent *reward prediction errors*, which generate *reinforcement signals*. Reinforcement signals are scalar quantities representing the difference between actual and expected reward values. If reward is greater than expected, the reinforcement signal is positive. If reward is less than expected, the signal is negative. If reward is as expected, there is no signal at all. Associations are strengthened if reinforcement signals are positive, weakened if negative, and unchanged if neutral. The guiding principle of contemporary reinforcement learning theory is that dispositions do not change without a reinforcement signal indicating a reward prediction error. See R. Sutton & A. Barto, *Reinforcement Learning: An Introduction* (MIT Press 1998).

either culture begins without a very good sense of what makes a good argument, and that you must acquire through training the appreciation of good reasons as they are recognized within mathematical and scientific practice. To do so, your appreciation of good reasons must be reinforced with incentives that may in fact be irrelevant to the intellectual merits of a mathematical proof or a good scientific theory. These reinforcers–funding, esteem, the approval of your peers and advisors, the desire for friends–must be paired with the changes of mind you ought to make until you come to make those changes without any extrinsic reward whatsoever.

3.4 Intellectual Character Must Be Nurtured

Rational persuasion requires intellectual character, and intellectual character must be nurtured by reinforcement for good arguments to be persuasive. This thesis, a line of thought running from a proposal of Plato's, through the Skinner boxes of behaviorism, to the imaging data of contemporary computational neuroscience, explains why good arguments ever persuade, and why they so often fail to do so. Good arguments fail to persuade because we fail to develop an appreciation for the merits of good reasons. Good arguments fail to persuade without intellectual character.

From a wider angle, good arguments fail to persuade for many reasons. At the conversational level, the level at which Socrates offers his own explanation of the phenomenon in the *Gorgias* and in book one of the *Republic*, blame lies with the insincerity of one's interlocutors. At the psychological level, the level of this discussion (and the level of book eight of the *Republic*), blame lies with a lack of appropriate passion and respect for good reasons, what I describe as a characteristic appreciation for good arguments. But there are other levels beyond these at which one can engage the problem as well. At the social level, for example (the level of book two of the *Republic*), blame lies with any system of education that fails to develop an appreciation for good arguments. Character, I do not believe it a stretch

to say, is mostly fixed at an early age and fully set by adulthood. And so it is no surprise that rational persuasion should be the miracle it is among adults, even among adults for whom argument quality is prized, given that we are not trained in youth to form our beliefs on the basis of sound reasoning.

At the historical level, then, blame perhaps lies with Plato. Plato is the first to realize the importance of character development to one's appreciation for an argument's intellectual merits, and he aims to bring up the citizens of his republic to have souls governed by reason. In this respect his insight is as relevant now as it ever was. But out of concern for the abuse of argumentation characteristic of sophistry, he forbids dialectical training until one is well beyond one's formative years. Our own approach to public education remains very much a footnote to Plato. This is unfortunate, for the solution to the problem of a citizenry for whom good arguments are unpersuasive is not philosophy later in life. It is philosophy sooner.

I do not mean this as flattery to the profession, as I am under no illusions about philosophy's immediate persuasive successes. Van Inwagen's pessimism about the possibilities of successful philosophical argumentation is justified, and G.E. Cohen is right to be alarmed by the "paradox of conviction" mentioned earlier. The paradox, that we defend our beliefs against good arguments despite full awareness that we likely have them because of upbringing alone, suggests that human beings actually lack the intellectual character we aspire to. And as Cohen notices, this goes not only for our most fervent beliefs, but for our most dispassionately held philosophical opinions as well. There are indeed grounds for suspecting that we have the philosophical opinions we do by and large because of where and from whom we receive our philosophical training and not because we have good reasons for holding them.

Cohen sees two options for the alarmed. Either accept that we lack the intellectual character we aspire to, or explain the paradox away. But this is a false choice, the very same false choice encountered before. It is a choice between (a) showing that there is in

fact a psychological passion and respect for good reasons despite the strong evidence that there is not, and (b) accepting that human beings are irrational and thereby conceding good arguments to be incapable of changing our minds. The solution to Cohen's puzzle is the same. It is to recognize that we may in fact lack the intellectual character we aspire to, but to deny that this is a fate to which we must permanently resign ourselves, or indeed one that that is unimproved by philosophical training. It is poor fortune that good arguments face so many difficulties in life, but, with the proper cultivation of intellectual virtues, it is not a fortune they must perpetually endure.

There remains the curious fact that philosophers, though they may possess just as much time, skill, and enthusiasm for good arguments as mathematicians and scientists, seldom achieve the same measure of consensus found within mathematics and the sciences. To many this suggests something wrong with philosophy, but maybe it reflects something wrong with philosophers. Perhaps, among philosophers, intellectual character is overdeveloped, producing a pathological obsession with the rational support of one's beliefs and a compulsion to produce better and better reasons for them. Or perhaps philosophers always find a way to disagree because arguing is what they do best. Either way, the quest for the perfect argument will splinter any consensus.

To restore faith in the power of reason is to restore faith in philosophy itself. But that has not been my intention. What philosophy needs is not faith that good arguments persuade but a good argument demonstrating their ability to do so. The redemptive account of rational persuasion is a recursive account of rational persuasion, one that persuades by its own intellectual merits. My intention has been to replace faith in the power of reason with a rational defense of its possibility, in light of the strongest case for skepticism I can muster. It is only when one has seen this light that rational persuasion may truly be saved.

4

Two Empirical Challenges to Character

Virtue faces two challenges from social psychology. The first is from *situationism*, the thesis that broad, cross-situationally stable character traits do not exist. Situationism threatens any notion of character whatsoever. The second is from *automatic intuitionism*, the thesis that the influence of deliberation and reasoning on belief is negligible. Automatic intuitionism threatens intellectualist approaches to moral character in particular. Patience is the answer to both. It is required on the part of the observer to see the consistency of an individual's thought and behavior from situation to situation, and on the part of the individual to reduce the influence of irrelevant situational information on one's thought and behavior. Without patience, decisions about what to think and do are less likely to result from reason and argumentation, and therefore more likely to be affected by circumstantial irrelevancies.

4.1 Background

Psychology is replete with surprising, situationally prompted exhibitions of apparent irrationality and callousness. There are striking, classic studies like Solomon Asch's independence and conformity studies (1956), Stanley Milgram's obedience studies (1963), Philip

Zimbardo's Stanford prison experiment (Haney et al., 1972), Isen and Levin's (1972) "dime study," Darley and Batson's (1973) "good samaritan" experiment, Matthews and Canon's (1975) demonstration of the effect of ambient noise on helping behavior, and Tversky and Kahneman's (1981) work on framing effects. But there are also quirky, contemporary studies as well, like Bargh et al.'s (1996) experiments showing that subjects primed with words like "rude," "bother," and "infringe" are more likely to interrupt a confederate, Simonsohn's (2007) finding that college admissions officers scrutinize applications more closely on cloudy days, Schnall et al.'s (2008) finding that subjects exposed to bad odors or messy environments judge others' actions more harshly, Danzinger et al.'s (2011) finding that the harshness of judges' sentences varies with time from morning recess or lunch, and Chandler et al.'s (2012) finding that subjects judge a book to be more important if it is made heavier with a hidden weight.

Some philosophers (Harman 1999; 2000; 2003; Vranas 2009; Doris 1998; 2002; Merritt 2000, Doris and Stich 2008; Alfano 2013) believe that these findings support a new narrative of radical human contextual variability, one at odds with both commonsense and professional psychology and the psychological commitments of virtue theory. They claim that psychological research unequivocally demonstrates the power of situational influences in determining thought and behavior, notwithstanding a human impulse to overestimate the effects of personality on behavior.[1] They argue that character traits, as they are generally conceived, are neither as predictive nor explanatory as conventional wisdom, professional psychologists, or traditional philosophers holds them to be, which, together with a preference for theoretical parsimony, warrants skepticism about their role in our lives.[2] These

[1] Harman, quoting Kunda (1999), writes, "These conclusions are uncontroversial and a similar account can be found in almost any recent textbook in the social psychology. Such conclusions are supported by a very large amount of disparate evidence." But here is David Myers, author of Amazon's top-selling social psychology textbook, writing about the person-situation debate: "The pendulum of professional opinion swings. Without discounting the undeniable power of the social forces recognized in the 1960s and 1970s, the pendulum has swung back toward an appreciation of individual personality and its genetic predispositions." (2012, p. 221)

[2] Christian Miller (2014a; 2014b) defines the principal (negative) situationist thesis as the claim that "There

philosophers conclude that normative projects promoting the cultivation of character cannot succeed because of their impossible ideals, and that such programs therefore ought to be abandoned. This is the situationist challenge.

If situationism has a founding document, it is Walter Mischel's *Personality and Assessment* (1968).[3] There Mischel challenged two fundamental assumptions of what was then contemporary personality psychology. The first was the assumption that responses on personality tests and questionnaires (Q-data) predict traits, which further predict behavior (T-data). This assumption is not unreasonable. Suppose you need to hire someone to fill a job opening, and furthermore that no one in the applicant pool has done the job before. Without information about any given candidate's actual behavior, and with only what a candidate is willing to tell you to go on, you ask the applicants questions that you hope will reveal what they are like, assuming that what they tell you will predict job performance. But as Mischel pointed out, Q-data routinely predicted what subjects thought of themselves, by predicting responses to other questions, much better than they predicted actual behavior. In fact Q-data appeared to have a predictive ceiling, accounting for no more than thirty percent of observed variation in behavior.

The second assumption Mischel challenged was that behavior is consistent across relevantly similar circumstances. Circumstances are relevantly similar if the differences between them should not effect the behavior they elicit. This assumption is not unreasonable either. If a person is aggressive or generous toward a stranger before breakfast, that person should be more or less equally aggressive or generous toward a stranger after breakfast. But by the late 1960s, this assumption had yet to be borne out by behavioral evidence, as even

is a large body of experimental evidence which is incompatible with the widespread possession of certain traits" (2014a, p.86) But if this is indeed the situationist's principal negative thesis, it is hardly distinctive of situationism. As Julie Annas(2003) points out, it is something both Plato and Aristotle take to be a starting point for their own ethical projects, and not something they would find surprising at all. Moreover, it is difficult to imagine that it violates commonsense guesswork with respect to psychological demographics either.

[3] Precursors include Hartshorne and May's *Studies on the Nature of Character* (1928; 1929; 1930) and Leon Festinger's cognitive dissonance theory (See 1962; 1959).

minor situational differences appeared to disproportionately affect behavior. Thus Mischel pointed out that T-data from individual behavior also appeared to have a predictive ceiling, again accounting for no more than thirty percent of behavioral variation from situation to situation.

Using personality tests and questionnaires as a reference point for what individuals actually think about their own behavior, Mischel argued further that the professional assumption of cross-situational behavioral consistency must be germane to commonsense psychology as well. This claim received independent support from work that Edward Jones and Victor Harris (1967) had published a year earlier. Jones and Harris provided their subjects with essays for and against Fidel Castro. Some subjects were told that a debate coach had instructed the authors of the essays to convey a certain message. Surprisingly, these subjects continued to attribute the attitudes conveyed by the essays to the authors themselves, despite information about the debate coach's instructions to convey the message contained. Writing ten years later, after the effect had been demonstrated repeatedly, Lee Ross described the phenomenon Jones and Harris first detected as the "fundamental attribution error," a "general tendency to overestimate the importance of personal or dispositional factors relative to environmental influences" (1977, p. 184). Ross claimed this general tendency to be the source of the apparent professional and lay assumption of cross-situational behavioral consistency.[4]

From psychology's deep pool of counterintuitive findings, Mischel's critique, and the existence of an apparent "dispositionalist" bias, it may be tempting to conclude that character traits do not exist at all. And by the 1980s, some psychologists did toy with such wholesale skepticism, at least as a methodological assumption (Cf. Bem, 1972; Shweder, 1975).[5] But except for Harman (1999; 2000; 2003), no situationist, from Mischel (if it is

[4] While Flanagan (1991) is cautious about the significance of the so-called fundamental attribution error, Harman (1999) thinks it has serious consequences for virtue ethics.

[5] Harman (1999) writes, "Empirical studies designed to test whether people behave differently in ways that might reflect their having different character traits have failed to find relevant differences," claiming further

even right to call Mischel a situationist; Cf. 2004, 2009) to Ross and Nisbett (Cf. Ross & Nisbett, 1991), has ever truly denied that personality variables account for some measure of individual variation in behavior. To do so would be a mistake for two good reasons.

The first reason is that the correlation coefficient (*r*) of .3 (Mischel, 1968) to .4 (Nisbett, 1980, p. 124) identified as the upper bound of any correlation between Q-data and T-data is, as both psychologists (Ross & Nisbett, 1991; Sabini & Silver, 2005; Kihlstrom, 2013) and philosophers (Flanagan, 2009) have pointed out, non-negligible. In their meta-analysis of a century of social psychological results, Richard et al., 2003 report a mean correlation coefficient .21, with only 5.28% of studies claiming a correlation coefficient above .5. (p. 336). Following Cohen (2003), no psychologist considers a correlation coefficients of thirty percent to be small. To drive the point home that a .4 correlation coefficient is nothing to dismiss, Funder et al. (2012) note that if one is above (or below) the median for a given character trait, there is a seventy percent chance of one's being above (or below) the median for correlated behavior.[6]

The second reason, asFunder & Ozer (1983) point out, is that the correlation between personal factors and behavior is virtually equivalent to the correlation between situational factors and behavior. In their meta-analysis, Richard et al. (2003) report "that situational effects are similar in magnitude to person effects (mean r = .22 in 17,631 estimates of situational effects; mean r = .19 in 16,282 estimates of person effects)," though they do concede that "situational effects may be slightly larger." At any rate the evidence against radical situationism is conclusive.

---

that

that "the 'extreme view'"–wholesale skepticism about character–"is in fact widespread among social psychologists." These claims are mistaken.

[6] Funder et al. (2012) use Rosenthal and Rubin's Binomial Effect Size Display (BESD), a conversion formula developed to adequately convey the practical effect sizes of statistical correlations.

4.2 Moderate Situationism and Moral Character

While the psychological facts undermine any radical skepticism about character, situationists argue that according to what people think about their own behavior (on the basis of Q-data), personality should nevertheless play a greater role than it does in what individuals do, and that behavior itself should be more consistent from situation to situation than it is. More realistic versions of situationism, then, claim not that there are no character traits, nor that behavior is entirely inconsistent from situation to situation, but only that character traits are not as robust, and behavior not as cross-situationally consistent, as one might expect.

But this version of situationism is also fiction. As Seymour Epstein (1979; 1983; 1985) has long pointed out, when cross-situational behavioral data are appropriately aggregated, trends emerge that (a) match the predictions of Q-data-based character traits, and (b) predict, with exceeding accuracy, other aggregated behavioral data.[7] Sabini & Silver (2005), Sreenivasan (2008), and Slingerland (2011), all writing for a philosophical audience, make the same point.[8] Behavior in a single circumstance is poor evidence of a person's character, or lack thereof. One swallow does not a summer make.

Make no mistake, a trait need not be slight for its failure to appear only in the aggregate. It is no more a problem for the personality psychologist that traits appear only over the course of many averaged trials than it is a problem for the believer in a fair coin that its balance appears only over the course of many averaged tosses. The point is simply that individual situations are themselves too noisy to warrant just about any conclusion drawn

[7] As Epstein points out, even Hartshorne and May, who interpreted their own studies to challenge the existence of honesty in children, write: "Just as one test is an insufficient and unreliable measure in the case of intelligence, so one test of deception is quite incapable of measuring a subject's tendency to deceive. That is, we cannot predict from what a pupil does on one test what he will do on another. If we use ten tests of classroom deception, however, we can safely predict what a subject will do on the average whenever ten similar situations are presented" (1928, p. 135, as quoted by Epstein 1985).

[8] Sabini & Silver (2005) and Slingerland (2011) both follow Robert Abelson (1985) in using a baseball player's batting average to exemplify dispositional traits that are not nearly as predictive in individual circumstances as they are in the aggregate.

from them at all, and that this noise can only be minimized by viewing behavior from a wider angle. From a wider angle, both (a) Q-data based character traits and (b) aggregated behavior are very good predictors of other patterns of behavior across relevantly similar trait-eliciting situations.

Some situationists acknowledge the fact that character traits predict behavioral patterns but deny its significance, at least for the existence of some alleged virtues. Mark Alfano (2013), following Doris (2002), objects that high-fidelity traits like integrity, honesty and chastity require "near perfect consistency." Unlike the possession of low-fidelity character traits, the possession of high-fidelity character traits is undermined by even one bad deed. If Alfano is right that anyone is, or ought to be, committed to the existence of high-fidelity traits, then situationists are right to focus on particular instances of behavior.

The chief difficulty with Alfano's notion of high-fidelity traits, the possession of which requires "near perfect consistency," is that it does not play a role in either traditional virtue theory or commonsense thinking about virtue. First, let us consider traditional virtue theory. To support the attribution of a commitment to high-fidelity traits to Aristotle, Doris (2002, p. 17) quotes him as saying that "the blessed man... will never do the acts that are hateful and mean" (1100b32-34) and that "the good man will never voluntarily do bad actions (1128b29)." But in both places, Aristotle is making a logical–not a descriptive–point. In the first instance, the good and blessed (i.e. fortunate) person will act well because she has, by stipulation, neither personal nor circumstantial reasons for acting badly. There is nothing external to interfere with the exercise of virtue. In the second instance, Aristotle is discussing, and rejecting, the idea that it is virtuous to feel shame for voluntarily doing something wrong. If it were virtuous to feel shame for acting viciously, then one should feel good about acting badly, which is absurd. The source of the absurdity is mistakenly thinking that the good person, qua good person, voluntarily does evil.[9] In short, the pas-

[9] Julie Annas (2003) effectively makes this criticism in her review of Doris (2002), where she highlights the aspirational quality of Aristotle's accounts of the virtuous person. Doris (1998, p. 520), Miller (2003; 2014a;

sages in question do not support a characterization of Aristotle's view as one committed to high-fidelity virtues.

Even if Aristotle is not a good example of someone who believes in the existence of high-fidelity traits, perhaps there is evidence that commonsense psychology is nevertheless committed to them. As evidence, Alfano offers the example of no-tolerance policies regarding violations of academic integrity:

> A small dose of honesty, even if aggregated over time, does not make one an honest person. If the allegations against him are true, the disgraced biologist Marc Hauser, who allegedly fabricated data for his research on primates, arguably lacks the virtues of integrity and intellectual honesty even if he is guilty of only the eight counts of scientific misconduct for which he has been sanctioned. Think about that: just eight failures over the course of a decades-long career may be enough to disqualify Hauser. (2013, p. 106)

Is this a case in which public and academic reprobation signifies the commonsense expectation that one exhibit immaculate character? I do not think it does. Rather, I suspect that this is a case in which public and academic reprobation signifies the implicit judgment that eight violations of academic integrity represent aberrant behavior. The evidence for this judgment, of course, depends on the whatever the average number of instances of gross professional misconduct over the course of a decades-long career is, exactly, and on the standard deviation from that average as well. My guess is that were these numbers available, they would reveal eight instances of gross professional misconduct to be several stan-

2014b, p. 380) and Alfano (2012, p. 231) take this point about virtue as a normative ideal that is perhaps rather rarely met to be a kind of concession, insofar as it suggests that virtue theory is psychologically unrealistic. For example, Alfano writes, "Presumably one cannot intend to do something one takes to be impossible" (2012, p. 231). But consider an analogy in which it is perfectly reasonable–and common–to intend to do something one takes to be impossible. There are very few people in the world who can hit a golf ball onto the green at three hundred yards with any kind of regularity. They invariably do so because of nearly flawless mechanics that are impossible for almost anyone else on the planet to replicate with the same regularity. It would be wrong to say that these mechanics should not be modeled because it is unrealistic to believe that trying to model them will yield the same results.

dard deviations from the norm. At any rate, what is important for our purposes is that near perfect consistency is not unreasonable to expect if it turns out to be statistically normal behavior.

Still, Doris (2002), following Mischel (Mischel & Peake, 1982; Mischel, 2004), objects that averaging behavioral data eliminates crucial information revealed by individual test trials. Doris and Alfano both approvingly quote Epstein (1983), who criticizes personality psychologists for being "obsessed with the usually hopeless task of predicting specific behavioral acts in single situations on single occasions" (1983, p. 366). Doris writes:

> Let it be granted that aggregation reveals some determinative role for personality and behavior–the situationist does not claim otherwise–but it must also be granted, given the problem of the particular, that this role is rather weaker than either characterological moral psychology or... everyday thinking on personality leads one to expect. (p. 75)[10]

But this objection misses the mark. To be sure, inappropriately aggregating data in a way that ignores stable, albeit finer, behavioral patterns invariably results in information loss. Nevertheless, as Epstein has argued, and as aggregated data show, the existence of situationally-specific tendencies does not compromise the possibility of general behavioral trends at all, not even trends against which individual behavior may appear anomalous. A company's stock may rise on some days and fall on others while nonetheless trending decisively upward (or downward). A good hitter may strike out on occasion, and even reliably, against certain pitchers. And so too a person's individual behavior may change from situation to situation while nonetheless conforming to character-based projections.

What Epstein calls hopeless is not merely the attempt to predict and explain individual behavior on the basis of character traits, but the attempt to predict and explain individual

[10] In agreement Alfano calls the aggregative response a "Pyrrhic victory," for its presumed concession that it does not solve Doris's "problem of the particular," the problem of predicting and explaining individual behavior.

behavior at all. The situationist, like the Liebnizian who demands sufficient reason for the fair coin's turning up heads on a given toss, effectively complains that statistical measurements of a phenomenon suffer unless they can tell you what is going to happen, and why, at any given moment. But this is a mistake. The mistake is believing that there is a "problem of the particular" to be solved by recruiting situation-specific causes beyond those involved in predicting and explaining what happens generally.

That situationists believe the so-called problem of the particular should bother virtue theorists is undercut by the fact that a distinctive feature of virtue theory, of virtue ethics especially, is its patience, a principled resistance to the thin-slice evaluative concerns of action-based normative theories. For virtue theorists the unit of evaluative significance stretches beyond individual circumstances. This is especially true for Aristotle, at whose moral psychology situationists take repeated aim.[11]

Let us take stock. Radical situationism is false, and so too is the claim that the existence of cross-situationally consistent character traits is unsupported by behavioral evidence. Some situationists admit this, but respond that character traits nevertheless fail to predict or explain behavior in individual circumstances, even if they do predict and explain behavior in the aggregate. But this response has no bearing on virtue theory, which anyway advocates a more patient assessment of behavior. Thus situationism poses no threat to virtue theory, ancient or contemporary.

Nor, in fact, does it threaten lay theories of mind and behavior, the "everyday thinking on personality" Doris mentions. There are two sources of evidence for claims about lay conceptions of character. The first is the body of Q-data–questionnaire responses, peer-ratings, and subjective reports–that appears to Mischel (1968) to be at odds with behavioral measures of character. But the fact that these responses, ratings, and reports do correlate with aggregated behavior challenges Ross's (1977) description of them in terms of any kind

[11] For example, for Aristotle the quality of one's character does not even come into focus until generations after one's life is over.

of systematic error. The second source of evidence about lay theories of mind and behavior is the apparent tendency to overestimate the importance of dispositions demonstrated by studies like Jones and Harris's (1967) work. But research since Ross's creation of the "fundamental attribution error" label undermines the claim that humans systematically overstate the importance of one's dispositions.[12]

Jones's student Daniel Gilbert (1998) chronicles this research. Jones and Harris's (1967) study established a paradigm for investigating the effect it revealed, the assumption of a correspondence between a person's behavior and her dispositions. In 1982 George Quattrone inverted this paradigm. Rather than give his subjects situational information and prompt them with questions about the dispositional causes of behavior, Quattrone gave his subjects dispositional information and prompted them with questions about situational influences on behavior. He found that subjects tended to overestimate the importance of situational factors, effectively demonstrating a counterpart to the lay dispositionalism postulated by Ross. Gilbert's own work (1995; 1988) has shown that the tendency to credit situational effects to a person's dispositions, what he prefers to call a correspondence bias, is better explained as a result of several factors, none of which warrant the attribution of lay dispositionalism. These factors include actual ignorance about situational influences (what Gilbert describes as the invisibility of much situational information), a general difficulty seeing things from another person's point of view, a tendency to think more highly of oneself, as well as simple prompt-framing (1995). Furthermore, subjects appear to maintain their dispositional hypotheses in the face of situational evidence to the contrary, not because of any dispositionalist bias, but because of an "anchoring" bias in favor of one's initial hypotheses.

What is left of situationism after correcting for any embellishments of the alleged human

[12] Like Harman (1999; 2000; 2003), Doris (2002; 2005), and Alfano (2013), Miller (2014a, Ch. 7) thinks that commonsense psychology is grossly inaccurate with respect to the possession of "robust" character traits, but unlike these philosophers, Miller does not think that the so-called fundamental attribution error adequately accounts for this gross inaccuracy.

impulse to overestimate the effects of personality? Christian Miller writes:

> The behavior of most individuals tends to be influenced by various situational forces which activate certain of their mental dispositions, whose functioning and degree of impact on behavior are underappreciated by both ordinary people and even trained philosophers and psychologists. (2014a, p. 103)

The problem is that if this is all that situationism comes to, then the situationist challenge has been flattened into an unremarkable thesis. It is no surprise that human beings fail to grasp all of the many influences on thought and behavior, for the simple reason human beings generally struggle to assign causal credit with any precision for the behavior of complex systems.

When one focuses on patterns of behavior instead of behavior in individual circumstances, every major claim that situationists make is false. Social psychology does not support a new narrative of radical human contextual variability, one at odds with lay or professional psychology or the commitments of virtue theory. Psychological research does not unequivocally demonstrate the power of situational influences in determining thought and behavior, notwithstanding a human impulse to overestimate the effects of personality on thought and behavior, because there is no evidence that any such impulse exists at all. The claim that character traits are neither as predictive nor as explanatory as either commonsense or professional psychology holds them to be is unjustified, and any skepticism about their existence is similarly unwarranted. Normative projects promoting the cultivation of character do not have impossible ideals, and there is no situationist reason for abandoning them.

4.3 Automatic Intuitionism

The abiding lesson of the person-situation debate is that patience is required on the part of the observer to see the consistency of one's character. Because character traits manifest

themselves in the aggregate, situationism poses no general threat to their existence. But the situationist may nevertheless object to the normative project of cultivating intellectual character, specifically, even if there is no general situationist challenge to extend.[13] In a recent paper, Olin and Doris put the challenge thus:

> Reliability is widely held to be a characteristic feature of epistemic virtue... while a wealth of evidence indicates that human cognition is highly contingent on contextual variation, making the cognitive capacities of normal people quite unreliable. To put matters a bit hyperbolically: epistemic viciousness, rather than virtuousness, may best typify the human cognitive condition. (2013, pp. 669-670)[14]

The challenge of those looking to extend the situationist challenge to moral character to epistemology can be refined further by saying exactly *why* cultivating intellectual character in particular may be misguided. Cultivating intellectual character, at the least, is supposed to make an individual more reasonable, but if the long history of behavioral economics is to be believed, individuals are unreasonable primarily because they reach decisions about what to think and do *automatically*.[15] Unlike deliberative cognition–the slow, effortful, conceptually mediated, conscious, voluntary, active, affectively neutral, and self-monitored

---

[13] My own conception of intellectual character is much thinner than either of the standard "reliabilist" or "responsibilist" conceptions of epistemic virtue. I hold the notion of intellectual character only to require a disposition to act and decide in ways that exhibit mental maturity. I agree with Olin and Doris, however, that like reliabilist and responsibilist virtue epistemology both, this conception of intellectual character is "empirically committed, depending on a range of suppositions about human cognition" (2013, p. 668).

[14] Similarly, Alfano writes: "Intellectual virtues are empirically inadequate: they neither explain nor predict a sufficient portion of epistemic conduct because people are inordinately susceptible to seemingly trivial and epistemically irrelevant situational influences" (2012, p. 232). While there are merits to replies like Duncan Pritchard's (2014), which is that virtue epistemology's focus on agential considerations should not be blind to the influences of circumstance, I am unwilling to grant for the sake of argument, as Pritchard does, that the case for situationism has been made.

[15] Tversky and Kahneman (1974; 1981) are classic pieces; for more recent, popular accounts of the varieties of irrationality germane to intuitive cognition, see Kahneman (2011) and Ariely & Jones (2008). Gerd Gigerenzer (1996; 1999; 2002) and his collaborators, on the other hand, have argued that intuitive cognition tends to be *more* reliable than deliberative cognition, specifically when confronted with multidimensional choices.

faculty of *reason*–automatic cognition is fast, effortless, immediate, unconscious, involuntary, passive, often triggered by emotional cues and above all automatic. Automatic intuitions are isolated from the influence of reason and argument and much more sensitive to irrelevant situational information than are reasoned decisions about what to think and do. Enthusiasm for normative projects promoting thoughtful deliberation ought to be tempered. Thus situationism gives way to automatic intuitionism.[16]

Jonathan Haidt's work in moral psychology is a good example of contemporary automatic intuitionism. Much of the evidence for Haidt's view comes from studies of the way disgust shapes moral attitudes. Haidt et al. (1993) asked subjects to rate five scenarios on the extent to which they were harmful, bothersome, and morally wrong.[17] Haidt's team found that in limited populations–groups of low socioeconomic status (SES)–feelings of disgust tracked attributions of moral wrongdoing. Moreover, some high-SES participants confessed to an inclination to judge the highly offensive cases as morally wrong, though they took this inclination to be one that they could override in performing their ratings.[18]

[16] Intuitionism, as a thesis about attitude acquisition and change, is not new within philosophy. Many varieties of attitude, including perceptual or *phenomenal* beliefs (Russell, 1910), theological convictions (Jonathan Edwards, 2003), and moral sentiments (Hume, 2003) have been identified with intuitive cognition. By identifying certain attitudes with intuitive cognition, intuitionists partly aim to explain why reasoning and arguing fail to change those attitudes. By hypothesis, attitudes acquired by one cognitive routine cannot be changed by another. This does not mean that attitudes identified with intuitive cognition are irrational. But it does mean (a) that attempting to change them by way of reason and argumentation may be imprudent, and (b) that they may be highly sensitive to irrelevant situational information.

[17] Subjects were asked to judge the vignettes on six scales: wrongdoing, justification, harm, extent to which the action should be stopped, and extent to which it would be permissable for another culture to have such a practice. I am condensing the first, fifth, and sixth into one, here, and postponing talk of justification until later.

[18] Conceptual versions of Haidt et al.'s (1993) study have been repeated successfully with striking consequences. Wheatley & Haidt (2005) used a posthypnotic suggestion to associate a feeling of disgust with words like 'take' and 'often' in two related experiments. In the first experiment, hypnotizable subjects were presented with two-sentence vignettes, some that were offensive, and some describing mundane infractions like bribery, hypocrisy, and stealing library books, in versions with or without the loaded word. Results showed that subjects rated cases presented in versions with a trigger word as both more disgusting, and as more morally wrong, than cases presented without a trigger word.

In Wheatley and Haidt's second experiment, subjects were presented vignettes that were either not disgusting at all, though still depictions of minor infractions, or neither disgusting *nor* depictions of infractions. The most interesting of these is a story about "Dan":

> Dan is a student council representative at his school. This semester he is in charge of scheduling

Haidt argues that moral cognition operates automatically, under the radar of deliberative cognition, on the basis of somatosensory responses to situational stimuli. Moral intuitions are a function of the activation of moral modules, each of which–Haidt identifies six–automatically generates affective responses to immediate stimuli. Reasoning and arguing functions not to change moral attitudes but to rationalize holding them. Therefore, fostering moral character, to the extent that it is successful, results in better-sounding rationalizations for one's moral intuitions, but not in more reasonable moral attitudes.

If fostering moral character does not affect the intuitions that ground moral attitudes, but instead merely dresses them up for the sake of rational appearances, perhaps the virtue theorist might try cultivating intuition itself, "tuning the settings" of one's moral modules.[19] Plato and Aristotle each take this approach. Being of sound character, for Plato and Aristotle, means possessing a disposition that does not itself require reasoning or arguing to arrive at the appropriate decision about what to think or do.[20] Moral knowledge on this view is know-how that can only be learned by doing, or–more precisely–by doing as the virtuous person does. Virtue, for Plato and Aristotle, cannot be acquired didactically, the aim of ethics not being to make individuals better off directly, but to figure out what one needs to do, and whom to emulate, that one might live a more fulfilling life.

The problem with a virtue theory that promotes fine-tuning intuitions, however, is that it is precisely when we rely on automatic cognition that our attitudes are most vulnerable to irrelevant situational information. Intuitions are cognitively cheap and situationally flimsy. They are most vulnerable to irrelevant situational information because of the very features

> discussions about academic issues. He [tries to take/often picks] topics that appeal to both professors and students in order to stimulate discussion. (p. 782)

Only when the trigger word was included was any disgust or immorality rating was assigned at all; in the absence of 'take' or 'often', the story seemed to subjects completely innocuous.

[19] Flanagan & Williams (2010) describe the education of one's intuitions as tuning modular settings, terminology I borrow here.

[20] In the *Republic*, Plato departs from Socratic intellectualism and proposes a mimetic/habituation account of moral education. For discussion of the teachability of virtue, see also *Protagoras* and *Meno*. Aristotle's account of moral education in the *Nicomachean Ethics* is similar.

that make automatic cognition indispensable to limited cognitive agents and relatively reliable in the long run. Unlike deliberative cognition, automatic cognition is fast, effortless, immediate, unconscious, involuntary, passive, often triggered by emotional cues, unmonitored and automatic.

If moral cultivation means developing a person's ability to think critically about moral matters, and yet moral attitudes are generated by automatic cognition, then cultivating moral character will fail to affect the likelihood that these attitudes give way to reason and argument. If anything, a cultivated moral intellect is more likely to result in reason's conforming to, rather than changing, one's attitudes. But if moral cultivation means developing a person's automated sensitivity to argumentative merits, then, because intuitions are especially vulnerable to irrelevant situational information, one's attitudes will nevertheless vary from situation to situation just as wildly as situationists claim. The virtue ethicist must choose between cultivating reason, which may be incapable of affecting certain moral attitudes (specifically attitudes that already seem difficult to change with reason and argumentation, i.e. the very attitudes that motivate the cultivation of character in the first place), or automatic cognition, which is, unfortunately, overly sensitive to situational influences.

4.4 A Perfect Storm

There are two points to make that resolve this dilemma. The first is that automatic cognition can be tuned appropriately, and managed by reason when not. The second is that the situational influences on automatic cognition are not nearly as worrisome as the automatic intuitionist claims.

On the first point, recall Haidt et al.'s 1993 study first demonstrating the correlation between disgust and attributions of moral wrongdoing. In that study it was not disgust, but socioeconomic status that proved to have the largest effect on moral attitudes, with high-SES subjects from elite research universities judging offensive situations to be much less morally

wrong than any other group, and disgust predicting negative moral judgments only among low-SES groups. This is significant to the extent that SES correlates with education, (a) in the narrow sense of exposure to new, sometimes challenging ideas, and (b) in the broad sense of encounters with different cultures, customs, norms, and expectations.

This is not to say that those of Haidt et al.'s low-SES subjects who judged offensive actions with moral approbation were wrong to do so. It is only to point out that the evidence of SES's effect on moral attitudes is evidence that intuitions can themselves be tutored or at least overridden, just as one's palette can be trained so that foods that seem disgusting at first can come to be appetizing with time, or at least tolerated. Perhaps just as importantly, while some of Haidt et al's high-SES subjects admitted to intuitive judgments of moral wrongdoing, these subjects were able to override their own intuitions to deny any kind of moral violation. That these subjects acknowledge their inclination to identify offensiveness with moral wrongdoing and yet explicitly decide against doing so challenges the exaggerated worry that any reasoning about intuitive cognition is bound to produce conclusions that conform to one's intuitions.

More recently, Haidt (2013) has argued that political affiliation in the United States tends to correlate with the settings of the six moral modules he identifies, claiming that both moral and political intuitions are a function of SES and, in turn, educational history.[21] While Haidt denies that the correlation between educational background and moral or political ideology suggests any idealogical superiority on the part of well-educated individuals, the point is not that the morals of the educated class are superior. It is merely that both the content of one's intuitions and how such intuitions are cross-checked and controlled under deliberation are a function of one's education, which manifests itself in the long run (if not

[21] According to Lee Drutman (2011) of the Progressive Policy Institute, and on the basis of data from Gallup: "For every one percentage point increase in state-level college graduates, the percentage of liberals also increases by 0.75 percentage points. Impressively, education level explains 66 percent of variation in state-level percentage of liberals. By contrast, for every one percentage point in college graduates, there is a 0.88 percentage point decline in the share of conservatives, and this by itself explains 58 percent of the state-level variation in the number of conservatives."

clearly under individual circumstances).

This brings us to the second point. The automatic intuitionist may reply along the following lines, effectively arguing that because automatic cognition normally goes unchecked, intuitions and the attitudes based upon them remain vulnerable to irrelevant situational influences. This is the argument Olin and Doris make, effectively echoing points first made by the psychologist John Bargh (1999). They write:

> The evidence indicates that cognition often varies with epistemically irrelevant factors— font size, ease of pronunciation, and so on—that are unlikely candidates for size, ease of pronunciation, and so on—that are unlikely candidates for evidence. Where judgments are substantially influenced by non-evidential considerations, we submit, there is a plausible presumption of increased error. (2013, pp. 671-672)

Both Olin and Doris, and Bargh before them, believe that managing the effects of irrelevant situational information requires an implausible series of events. One must notice ordinarily covert stimuli, reason about them, and actively inhibit their influence appropriately. Though perhaps we do sometimes notice such influences, take them into consideration, and manage them well, the expectation that we do so most of the time, or even just when it counts most, is implausible. As Olin and Doris write:

> This may be a hard trick. Many epistemically challenging circumstances are likely to be unobtrusive, making them difficult to identify: perhaps it doesn't take much to know one should be on their epistemic guard in situations of low ambient light or high ambient noise, but who knew that hallways were epistemically toxic? (2013, p. 673)

In other words, the objection is that catching irrelevant situational factors in the act of influencing one's decisions–not just moral decisions but decisions of all kinds–requires a perfect storm of timely prompting and good fortune.

A similar problem arises in neuroscience for the reliability of visual searches, which share many features of automatic cognition. In a 2005 *Nature* paper, Wolfe et al. (2005) report an alarming error rate when subjects were asked to detect low-frequency visual targets. For their search task, Wolfe et al. (2005) asked subjects to scan visual arrays on a computer screen for possible weapons, flagging targets by pressing a button to indicate the presence of a weapon and subsequently moving to the next array. When targets appeared at a rate of 50%, subjects missed only about 7% of them; when targets appeared at very low rate, however – at 1%, for example – subjects missed about 30% of them. The reason the high error rate for low-frequency items is so alarming is two-fold: first, the very circumstances under which visual search accuracy is most critical are circumstances in which the odds of a particular item's appearing are slim: luggage scans and cancer screenings are the examples Wolfe et al. (2005) mention; second, there is great incentive to find the target objects, suggesting subjects should be even more vigilant than they would be under normal conditions.

In response to this finding, Fleck and Mitroff (2007) produced a variation of the experiments conducted by Wolfe et al. (2005) in which subjects were afforded the opportunity to correct their initial responses before advancing to the next screen. They found the error rate to drop dramatically, such that there was little difference between high-, medium-, and low-frequency searches. As an explanation Fleck and Mitroff suggest that the high error rate, or *prevalence effect*, found by Wolfe et al. (2005), reflects not a problem with sensitivity to visual stimuli but one with habituated motor responses to low-frequency targets: that is to say, not perceptual but motor errors. Fleck and Mitroff show that subjects with the incentive and the opportunity to self-correct manage to erase the prevalence effect. Why this round of studies is relevant is that the low-frequency error effect found by Wolfe et al. disappears when subjects are prompted to confirm their responses. In each case, awareness of the need for correction seems both possible, and, if prompted, instrumental in managing tendencies that potentially undermine intended behavior.

As it stands, the case is more complicated than either of these *prima facie* vindications of deliberation and self-monitoring would suggest. The prevalence effect demonstrated by Wolfe et al. (2005) does not seem to disappear for quite the reason Fleck and Mitroff propose. Wolfe et al. (2007) and Van Wert et al. (2009) report that the effect persists despite slowing down the process by which decisions are made, implying that it cannot be a mere pre-potent motor response responsible for missing low-frequency target stimuli. Wolfe et al. (2007) hypothesize instead that Fleck and Mitroff's prompt for confirmation delivers tacit feedback, and that such feedback, when provided for high-prevalence search tasks interspersed between low-prevalence tasks without feedback, creates a stricter standard for proceeding to new visual arrays. It is this stricter standard–the requirement that subjects assess and confirm their own first impressions–they argue, that cancels the prevalence effect for low-frequency stimuli.

If Wolf et al.'s (2007) explanation is right, it poses a version of the perfect storm objection raised by Bargh and Olin and Doris. In both cases, successful performance requires of one's environment that it be specifically engineered for one's success. In the visual search case, this means ensuring that target stimuli are delivered at varying rates with varying feedback in order to create a stricter standard for judgment. More generally, it means ensuring that irrelevant situational information is eliminated from a person's environment. But as the critic of virtue would surely agree, it is hard to see how opting to rig a person's environment for success qualifies as anything but a retreat from the cultivation of virtue.

Many contemporary writers emphasize the importance of "choice architecture," that is, engineering environments for people so that they are less likely to make mistakes (Cf. Thaler & Sunstein, 2008; Merritt, 2000; Doris, 2002; Alfano, 2013).[22] This is part and parcel of what Hagop Sarkissian (2010) calls the "seek/avoid" strategy to improve one's decisions. And of course, as a matter of social epistemology, it is good practice. But there are severe limits to any approach that puts too much of the weight of personal decisions on crutches,

[22] The connection to embodied or situated cognition (Cf. Clark, 2008) is not far off.

despite its being helpful under some circumstances. Besides requiring the continued curation of circumstance, it is no aid whatsoever in stubborn, intellectually hostile decision environments, those that cannot be made artificially friendlier through careful manipulations. And anyway, as Sarkissian points out, one may have no choice but to face difficult circumstances head-on.

If the perfect storm objection–the claim that a person's environment must be painstakingly engineered for success to stave off the influence of irrelevant situational factors–presents the most serious challenge to the cultivation of moral character, then limits on the efficacy of choice architecture only make matters worse. The fact is, however, that it is not true that a person's environment must be painstakingly engineered for success to stave off the influence of irrelevant situational factors. On the contrary there are two reasons for thinking that the perfect storm objection is misguided.

First, as the prevalence effect disappears when individuals are prompted to confirm their assessments of visual arrays, many situational effects simply disappear as soon as individuals become aware of their possible influence. The name-letter effect, for example–the phenomenon whereby individuals prefer things that have names with letters common to their own names (See Nuttin, 1985)–seems to go away when subjects are asked to think of *reasons* for their evaluative judgments (Koole et al., 2001). Likewise, and as Alfano recognizes (2012, p. 239), the effects of mood elevators on thought and behavior dissipate after a short time.[23] Furthermore, disproportionately large effects caused by things like failing to get enough sleep, getting cut off by another vehicle on the morning drive, or having one too many drinks are the factors least likely to remain hidden from conscious awareness over the long run, and, because they are disproportionately large, exactly the sort of situational factor we are already used to accounting for in the course of our daily lives.

Second, situational stimuli that are too small or fleeting to ever make it to conscious

[23] Alfano cites Isen et al. (1976), who were writing specifically about the duration of effects of mood elevators on helping behavior.

awareness–precisely the sorts of environmental variables that situationists characterize as disproportionately small or trivial–are in fact extremely limited in their potential effects on thought and behavior.[24] For such stimuli to play a role in thought or behavior, responses to them must be appropriately, and externally, prompted. Stanislas Dehaene (2009) argues convincingly that this is because nonconscious stimuli fail to achieve cross-modal neural integration, which leaves them unavailable as a source of internally generated–that is to say, voluntary–behavior. So while a person's judgments and actions may by affected by unconsciously processed information, she herself must be externally prompted to make and perform such judgments and actions; she will not be compelled to do so on her own. Moreover, she must also be prompted to produce these judgments and actions very soon after exposure to a given subliminal prime, somewhere in the range of 500ms (Dehaene cites Greenwald et al., 1996 and Mattler, 2005.), meaning that one is only vulnerable to circumstantial irrelevancies for very short periods of time after exposure to them.

Dehaene's discussion of blindsight helps illustrate the point. Some people with severe visual impairments can make better-than-chance guesses about the shape, size, and location of objects in their surroundings. But as Dehaene points out "a fundamental difference with normal subjects, however, is that blindsight patients never spontaneously initiate any visually guided behavior in their impaired field. Good performance can be elicited only by forcing them to respond to stimulation" (2009, p. 102). The ease with which these conditions are met in experimental settings masks the unlikelihood of their being met otherwise. Unconscious stimuli do not play a major role in thought or behavior unless one's environment has been engineered with external prompts and triggers for them to do so. This is unlikely to happen outside the artificial confines of experimental settings. In short, major situational stimuli are unlikely to remain below the radar of conscious awareness, which

[24] Alfano (2012) writes: "Both in their behavior and in their thought, feeling, and deliberation, people are astonishingly susceptible to seemingly trivial and normatively irrelevant features of their situations, such as mood elevators, mood depressors, ambient sounds, ambient smells, social distance cues, and even the weather. If few people are virtuous, then explaining human conduct in terms of the virtues would seem to be a hopeless endeavor." See also Doris 2002; Olin & Doris 2013 for similar remarks.

makes them easier to manage, and minor situational stimuli, those that situationists characterize as trivial or disproportionately small, are unlikely to influence us at all unless a lot of things go right (or wrong), which mitigates the need to manage them.

The automatic intuitionist claims that both cultivating reason and cultivating automatic cognition are bad ideas on grounds that (a) reason conforms to intuitive attitudes but does not upset them, and (b) automatic cognition is highly vulnerable to situational influences. But neither of these claims withstands scrutiny. Contrary to the first, automatic cognition can be tuned appropriately, and managed when not. Contrary to the second, the influences of irrelevant situational information on decisions about what to think or do is exaggerated by the very experiments that reveal them. Neutralizing irrelevant situational information requires the patience to stretch decisions across various situations, far enough for major situational effects to reveal themselves and minor situational effects to fizzle out.

4.5 Conclusion

Patience is a theoretical virtue required on the part of the observer to see the consistency of an individual's thought and behavior from situation to situation. It is also an intellectual virtue, one required on the part of the individual to reduce the influence of irrelevant situational information on thought and behavior. It is true that automatic cognition is especially vulnerable to situational influences, but we can and do correct for this vulnerability by forming attitudes gradually, under different circumstances and various conditions. Patience is all the more important for difficult decisions about what to think and do.

Sometimes situations present decisions that must be made immediately. To be sure we should be prepared to make them, and we are no less responsible for how we do so. But many decisions, often difficult decisions, need not be made with the suddenness with which they arise. As Churchland and Suhler (2011) remark, this is the difference between the decision to pick up a pencil, on one hand, and the decisions like those to quit law school

or switch auto insurance, on the other. By delaying these decisions appropriately, we create safeguards against momentary lapses of reason brought on by irrelevant circumstantial variables. As Adam Morton (2012) suggests, we must cultivate the ability to grapple with a problem long enough to make an appropriate decision about what to think or do. The more we stretch a decision out, the greater control we exercise over it, relative to the conditions in which it is made, and the less vulnerable we are to the irrelevancies of circumstance.[25]

[25] Of course, the possibility of patience and the prudence of its cultivation require defenses of their own, neither of which I can provide here.

Bibliography

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, 97(1), 129.

Aikin, S. (2011). *Epistemology and the regress problem*. Routledge.

Alfano, M. (2012). Expanding the situationist challenge to responsibilist virtue epistemology. *The Philosophical Quarterly*, 62(247), 223–249.

Alfano, M. (2013). *Character as Moral Fiction*. Cambridge University Press.

Annas, J. (2003). Virtue ethics and social psychology.

Ariely, D. & Jones, S. (2008). *Predictably irrational*. HarperCollins New York.

Aristotle (1984). Nicomachean ethics. In J. Barnes (Ed.), *The complete works of Aristotle : the revised Oxford translation*. Princeton, N.J. : Princeton University Press.

Asch, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70(9), 1–70.

Atkinson, D. & Peijnenburg, J. (2009). Justification by an infinity of conditional probabilities. *Notre Dame Journal of Formal Logic*, 50(2), 183–193.

Azzouni, J. (2000). *Knowledge and reference in empirical science*, volume 2. Psychology Press.

Bargh, J. A. (1999). The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects. Dual-Process Theories in Social Psychology. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford Press.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology*, 71(2).

Bem, D. J. (1972). Constructing cross-situational consistencies in behavior: Some thoughts on alker's critique of mischel. *Journal of Personality*, 40(1), 17–26.

Biro, J. & Siegel, H. (2008). In defense of the objective epistemic approach to argumentation. *Informal Logic*, 26(1), 91–101.

Biro, J. I. (1977). Rescuing 'begging the question'*. *Metaphilosophy*, 8(4), 257–271.

Biro, J. I. (1984). Knowability, believability and begging the question: A reply to Sanford. *Metaphilosophy*, 15(3-4), 239–247.

Black, M. (1964). The gap between "is" and "should". *The Philosophical Review*, 73(2), 165–181.

Bonjour, L. (1976). The coherence theory of empirical knowledge. *Philosophical Studies*, 30(5), 281–312.

Burge, T. (1993). Content preservation. *The Philosophical Review*, 102(4), 457–488.

Carroll, L. (1895). What the tortoise said to Achilles. *Mind*, 4(14), 278–280.

Chandler, J. J., Reinhard, D., & Schwarz, N. (2012). To judge a book by its weight you need to know its content: Knowledge moderates the use of embodied cues. *Journal of Experimental Social Psychology*, 48(4), 948–952.

Churchland, P. S. (2009). Inference to the best decision1.

Churchland, P. S. & Suhler, C. (2011). Me and my amazing oldfangled reward system. From "Free Will and Responsibility: Perspectives From Neuroscience, Psychology and Philosophy".

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension: Embodiment, Action, and Cognitive Extension*. Oxford University Press.

Cling, A. D. (2004). The trouble with infinitism. *Synthese*, 138(1), 101–123.

Cohen, G. (2000). Paradoxes of conviction. In *If you're an egalitarian, how come you're so rich?* Cambridge, Mass. : Harvard University Press.

Cohen, J. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, N.J: L. Erlbaum Associates.

Cohen, L. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4(03), 317–331.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17), 6889–6892.

Darley, J. M. & Batson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27(1), 100.

De Waal, F. (2007). *Chimpanzee politics: Power and sex among apes*. Johns Hopkins University Press.

Dehaene, S. (2009). Conscious and nonconscious processes: Distinct forms of evidence accumulation? *Seminaire Poincare*, XII, 89–114.

Dennett, D. (1981). Why the law of effect will not go away. In *Brainstorms: Philosophical essays on mind and psychology*. MIT Press.

Descartes, R. (1984). *The Philosophical Writings of Descartes Volume II*. Cambridge University Press.

Dewey, J. (1925). Logic: The theory of inquiry (1938). *The later works*, 1953, 1–549.

Dewey, J. (1941). Propositions, warranted assertibility, and truth. *The Journal of Philosophy*, (pp. 169–186).

Doris, J. (2002). *Lack of character: Personality and moral behavior*. Cambridge University Press.

Doris, J. & Stich, S. (2005). *As a matter of fact: Empirical perspectives on ethics*. Oxford: Oxford.

Doris, J. & Stich, S. (2008). Moral psychology: empirical approaches. *Stanford encyclopedia of philosophy*.

Doris, J. M. (1998). Persons, situations, and virtue ethics. *Noûs*, 32(4), 504–530.

Dretske, F. (1971). Conclusive reasons. *Australasian Journal of Philosophy*, 49(1), 1–22.

Dretske, F. I. (1970). Epistemic operators. *The journal of philosophy*, (pp. 1007–1023).

Drutman, L. (2011). More college graduates, more democratic voters?

Edwards, J. (1741/2003). *Sinners in the Hands of an Angry God*. Bridge Logos Foundation.

Elga, A. (2007). Reflection and disagreement. *Noûs*, 41(3), 478–502.

Epstein, S. (1979). The stability of behavior: I. on predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37(7), 1097–1126.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, 51(3), 360–392.

Epstein, S. & O'Brien, E. J. (1985). The person–situation debate in historical and current perspective. *Psychological Bulletin*, 98(3), 513.

Festinger, L. (1962). *A theory of cognitive dissonance*, volume 2. Stanford university press.

Festinger, L. & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, 58(2), 203.

Flanagan, O. (1991). *Varieties of moral personality: Ethics and psychological realism*. Harvard University Press.

Flanagan, O. (1993). *Varieties of moral personality: Ethics and psychological realism*. Harvard University Press.

Flanagan, O. (2009). Moral science? Still metaphysical after all these years. *Personality, identity, and character*, (pp.5[illegible]).

Flanagan, O. & Williams, R. A. (2010). What does the modularity of morals have to do with ethics? Four moral sprouts plus or minus a few. *Topics in Cognitive Science*, 2(3), 430–453.

Fleck, M. & Mitroff, S. (2007). Rare targets are rarely missed in correctable search. *Psychological Science*, 18(11), 943.

Frege, G. (1918). Thoughts.

Funder, D. C., Guillaume, E., Kumagai, S., Kawamoto, S., & Sato, T. (2012). The person-situation debate and the assessment of situations. *Japanese Journal of Personality*, 21, 1–11.

Funder, D. C. & Ozer, D. J. (1983). Behavior as a function of the situation. *Journal of Personality and Social Psychology*, 44(1), 107.

Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4), 650.

Gigerenzer, G. & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. Mit Press.

Gigerenzer, G., Todd, P. M., & Group, A. R. (1999). Simple heuristics that make us smart.

Gilbert, D. T. & Malone, P. S. (1995). The correspondence bias. *Psychological bulletin*, 117(1), 21.

Gilbert, D. T., Miller, A. G., & Ross, L. (1998). Speeding with ned: A personal view of the correspondence bias. *Attribution and social interaction: The legacy of Edward E. Jones*, (pp. 5–66).

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and social Psychology*, 54(5), 733.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science*, 273(5282).

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–34.

Haidt, J. (2013). *The righteous mind: Why good people are divided by politics and religion*. Random House Digital, Inc.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *J Pers Soc Psychol*, 65(4), 613–28.

Haney, C., Banks, C., & Zimbardo, P. (1972). *Interpersonal dynamics in a simulated prison*. Defense Technical Information Center.

Harman, G. (1965). The inference to the best explanation. *The Philosophical Review*, 74(1), 88–95.

Harman, G. (1999). : (pp. 315–331).: JSTOR.

Harman, G. (2000). The nonexistence of character traits. volume 100 (pp. 223–226).: Wiley Online Library.

Harman, G. (2003). No character or personality. *Business Ethics Quarterly*, (pp. 87–94).

Hartshorne, H. & May, M. A. (1928). *Studies in deceit*. Macmillan.

Hartshorne, H., May, M. A., & Maller, J. B. (1929). Studies in service and self-control.

Hartshorne, H., May, M. A., & Shuttleworth, F. K. (1930). Studies in the organization of character.

Horwich, P. (1998). *Truth*. Oxford University Press.

Hume, D. (1739/2003). *A treatise of human nature*. Courier Dover Publications.

Isen, A. M., Clark, M., & Schwartz, M. F. (1976). Duration of the effect of good mood on helping:" footprints on the sands of time.". *Journal of Personality and Social Psychology*, 34(3), 385.

Isen, A. M. & Levin, P. F. (1972). Effect of feeling good on helping: cookies and kindness. *Journal of personality and social psychology*, 21(3), 384.

Jackson, F. (1984). Petitio and the purpose of arguing. *Pacific Philosophical Quarterly*, 65(1), 26–36.

Jones, E. E. & Harris, V. A. (1967). The attribution of attitudes. *Journal of experimental social psychology*, 3(1), 1–24.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research*, 66(3), 612–640.

Kelly, T. (2010). Peer disagreement and higher order evidence. *Social Epistemology: Essential Readings*, (pp. 183–217).

Kihlstrom, J. F. (2013). The person-situation interaction. In D. Carlston (Ed.), *Oxford Handbook of Social Cognition*. New York: Oxford University Press.

Kim, J. (1988). What is "naturalized epistemology"? *Philosophical perspectives*, 2, 381–405.

Klein, P. (1998). Foundationalism and the infinite regress of reasons.

Klein, P. (1999). Human knowledge and the infinite regress of reasons. *Noûs*, 33(s13), 297–325.

Klein, P. (2003). When infinite regresses are not vicious*. *Philosophy and Phenomenological Research*, 66(3), 718–729.

Koole, S., Dijksterhuis, A., & Van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, 80(4), 669.

Kunda, Z. (1999). *Social cognition: making sense of people*. MIT press.

MacIntyre, A. (1959). Hume on "is" and "ought". *The Philosophical Review*, 68(4), 451–468.

Mates, B. (1996). The skeptic way: Sextus Empiricus' Outlines of Pyrrhonism.

Mathews, K. E. & Canon, L. K. (1975). Environmental noise level as a determinant of helping behavior. *Journal of Personality and Social Psychology*, 32(4), 571.

Mattler, U. (2005). Inhibition and decay of motor and nonmotor priming. *Attention, Perception, & Psychophysics*, 67(2), 285–300.

Merritt, M. (2000). Virtue ethics and situationist personality psychology. *Ethical theory and moral practice*, 3(4), 365–383.

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), 371.

Mill, J. S. (1882). *A system of logic, ratiocinative and inductive*. Project Gutenberg: A Project Gutenberg Ebook, 8th edition.

Miller, C. (2003). Social psychology and virtue ethics. *The Journal of Ethics*, 7(4), 365–392.

Miller, C. (2014a). *Character and Moral Psychology*. Oxford University Press.

Miller, C. (2014b). Moral virtues, epistemic virtues, and the big five. In A. Fairweather & O. Flanagan (Eds.), *Naturalizing Epistemic Virtue*. Cambridge University Press.

Mischel, W. (1968). *Personality and assessment*. Wiley.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, 55(1), 1–22.

Mischel, W. (2009). From personality and assessment (1968) to personality science, 2009. *Journal of Research in Personality*, 43(2), 282–290.

Mischel, W. & Peake, P. K. (1982). Beyond déjàvu in the search for cross-situational consistency. *Psychological Review*, 89(6), 730.

Morton, A. (2012). *Bounded thinking: intellectual virtues for limited agents*. Oxford University Press.

Myers, D. G. (2012). *Psychology*. Macmillan Higher Education.

Nisbett, R. E. (1980). The trait construct in lay and professional psychology. *Retrospections on social psychology*, (pp. 109–130).

Nuttin, J. M. (1985). Narcissism beyond gestalt and awareness: The name letter effect. *European Journal of Social Psychology*, 15(3), 353–361.

Olin, L. & Doris, J. M. (2013). Vicious minds. *Philosophical Studies*, (pp. 1–28).

Peirce, C. S. (1957). The logic of abduction. *Peirce's essays in the philosophy of science*.

Planck, M. (1949). *Scientific autobiography and other papers*. New York: Philosophical library.

Post, J. (2009). Infinite regress arguments. In J. Dancy, E. Sosa, & M. Steup (Eds.), *A companion to epistemology*. John Wiley & Sons.

Pritchard, D. (2014). Re-evalutating the situationist challenge to virtue epistemology. In A. Fairweather & O. Flanagan (Eds.), *Naturalizing Epistemic Virtue*. Cambridge University Press.

Putnam, H. (1978). Reference and understanding. In *Meaning and the Moral Sciences* (pp. 97–119). Routledge & Kegan Paul.

Putnam, H. (1982). Why reason can't be naturalized. *Synthese*, 52(1), 3–23.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of personality and social psychology*, 42(4), 593.

Quine, W. (1960). *Word and object*. The MIT Press.

Quine, W. (1969a). Epistemology naturalized. In *Ontological relativity and other essays*. Columbia Univ Press.

Quine, W. V. (1969b). Epistemology naturalized. In *Ontological relativity and other essays*, volume 1. Columbia Univ Pr.

Ramsey, F. (1999). On facts and propositions.

Ramsey, F. P., Rescher, N., & Majer, U. (1991). *The nature of truth*, (pp. 6–24). Springer.

Raz, J. (2011). Reason, rationality, and normativity. In *From Normativity to Responsibility*. Oxford University Press.

Rescher, N. (2010). *Infinite regress: The theory and history of varieties of change*. Transaction Publishers.

Richard, F., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7(4), 331.

Robinson, R. (1971). Begging the question, 1971. *Analysis*, 31(4), 113–117.

Robinson, R. (1981). Begging the question 1981. *Analysis*, 41(2), 65.

Rorty, R. (1986). Pragmatism, davidson and truth.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in experimental social psychology*, 10, 173–220.

Ross, L. & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Mcgraw-Hill Book Company.

Russell, B. (1910). Knowledge by acquaintance and knowledge by description. volume 11 (pp. 108–128).: JSTOR.

Sabini, J. & Silver, M. (2005). Lack of character? situationism critiqued*. *Ethics*, 115(3), 535–562.

Sanford, D. H. (1972). Begging the question. *Analysis*, 32(6), 197–199.

Sanford, D. H. (1977). The fallacy of begging the question: a reply to Barker. *Dialogue*, 16(03), 485–498.

Sarkissian, H. (2010). Minor tweaks, major payoffs: The problems and promise of situationism in moral philosophy. *Info: August 2010*, 10(9), 1–15.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Pers Soc Psychol Bull*, 34(8), 1096–109.

Shweder, R. A. (1975). How relevant is an individual difference theory of personality?1. *Journal of Personality*, 43(3), 455–484.

Simonsohn, U. (2007). Clouds make nerds look good: Field evidence of the impact of incidental factors on decision making. *Journal of Behavioral Decision Making*, 20(2), 143–152.

Sinnott-Armstrong, W. (1999). Begging the question. *Australasian Journal of Philosophy*, 77(2), 174–191.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century-Crofts New York.

Slingerland, E. (2011). The situationist critique and early confucian virtue ethics*. *Ethics*, 121(2), 390–419.

Sreenivasan, G. (2008). Character and consistency: Still more errors. *Mind*, 117(467), 603–612.

Tarski, A. (1944). The semantic conception of truth: and the foundations of semantics. *Philosophy and phenomenological research*, 4(3), 341–376.

Thagard, P. & Millgram, E. (1995). 19 inference to the best plan: A coherence theory of decision. *Goal-driven learning*, (pp. 439).

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. Macmillan.

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology*, 39(1/4), 212–222.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157), 1124–1131.

Tversky, A., Kahneman, D., & Choice, R. (1981). The framing of decisions. *Science*, 211, 453–458.

Van Fraassen, B. C. (1976). To save the phenomena. *The Journal of Philosophy*, (pp. 623–632).

Van Fraassen, B. C. (1980). The scientific image.

Van Inwagen, P. (2006). *The problem of evil*. New York: Oxford.

Van Inwagen, P. (2010). We're right. they're wrong.

Van Wert, M., Horowitz, T., & Wolfe, J. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, & Psychophysics*, 71(3), 541–553.

Vlastos, G. (1982). The Socratic elenchus. *The Journal of Philosophy*, (pp. 711–714).

Vranas, P. B. (2009). Against moral character evaluations: The undetectability of virtue and vice. *The Journal of Ethics*, 13(2-3), 213–233.

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, 11(2), 92–107.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology*, 12(3), 129–140.

Wason, P. C. & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*, volume 86. Harvard University Press.

Wheatley, T. & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychol Sci*, 16(10), 780–4.

Williams, B. (2006). *Ethics and the Limits of Philosophy*. Taylor & Francis.

Wolfe, J., Horowitz, T., & Kenner, N. (2005). Rare items often missed in visual searches. *Nature*, 435(7041), 439–440.

Wolfe, J., Horowitz, T., Van Wert, M., Kenner, N., et al. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, 136(4), 623.

Biography

Stephen J. Martin, IV was born and raised in Philadelphia, where he attended La Salle University. He received a master's degree in philosophy from Tufts University before receiving his doctorate in philosophy from Duke University in 2014. He is currently a lecturer in the department of philosophy at the University of Tennessee, Knoxville.