Cognitive *Akrasia* in Moral Psychology and Normative Motivation

By

Brandon Scott Gillette

Submitted to the graduate degree program in Philosophy and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson Dr. Dale Dorsey

_____

Dr. John Bricke

_____

Dr. Ben Eggleston

_____

Dr. Mabel Rice

_____

Dr. Sarah Robins

_____

Dr. Tom Tuozzo

Date Defended: 20 November 2013

The Dissertation Committee for Brandon Scott Gillette
certifies that this is the approved version of the following dissertation:

Cognitive *Akrasia* in Moral Psychology and Normative Motivation

_____
Chairperson Dr. Dale Dorsey

Date approved: 20 November 2013

ABSTRACT

A number of persistent questions surround *akrasia*. Is *akrasia* (acting intentionally against one's own better judgment) possible? If it is, how best to explain *akrasia* in a way consistent with acceptable theories of normative motivation? I argue that *akrasia* is possible—in fact, *akrasia* is actual. Research in psychology and information science, suitably interpreted, contains an empirically informed account of *akrasia* that is consistent with the traditional philosophical concept of *akrasia* as notably explored by Aristotle, Augustine, Aquinas, Hume, Hare, and Davidson.

My account of *akrasia* appeals to our best current research in order to develop an account of how someone could have knowledge of the good without attending to that knowledge, or could make normative judgments that motivate, but that do not include all of the factors at play in a more complete normative judgment (i.e. better judgment) that would motivate the agent differently.

Adopting this empirically informed account of *akrasia* requires abandoning positions that are incompatible with its existence. One such view is the view that normative judgments are necessarily connected to motivation (often called normative judgment internalism, or NJI). Drawing on works by Sarah Stroud and Ralph Wedgwood, I demonstrate that NJI can be amended to allow *akrasia*, long thought to be a straightforward counterexample to NJI, while preserving what is plausible about NJI.

My account of *akrasia* is termed 'cognitive *akrasia*' because I appeal to cognitive states as playing a central role in identifying and understanding *akrasia*. Preserving an amended NJI by means of a strongly cognitive understanding of *akrasia* means arguing against an opponent of NJI, which is normative judgment externalism (NJE). The most common form of NJE is

Humean in character, and explains *akrasia* in terms of desiderative or other affective states. That is, one is *akratic* when one judges that A is better than B but has less desire to do A than B.

My response to NJE as a view that explains *akrasia* is also empirically informed. I make use of clinical research into addiction and addiction treatment, because addiction has long been a fruitful source of examples of *akrasia*. Many addicts judge it better not to be addicts and yet occasionally or repeatedly fail to reform their addictive behavior. In this analysis, I provide a plausible family of everyday accounts of persons changing their behavior without changing their desires. I also point out that recent research indicates that specifically cognitive bias modification provides better clinical outcomes among addicts than approaches that attempt to change the addicts' desires. One important consequence of cognitive *akrasia*, then, is that it represents support for theories that hold that motivation can be a product of cognitive and not only affective states.

ACKNOWLEDGEMENTS

No dissertation represents the work of only a single individual, and this one is no exception. I would here like to recognize the contributions of a number of individuals.

First, I owe special thanks to my dissertation advisor, Dale Dorsey, who I pounced on for this project when he had only recently arrived at the University of Kansas. I'm sure he didn't quite know that he was preparing to advise a student who wished to defend, among other crazy assertions, that cognitive states alone could provide motivation. Dale was flexible, patient, and above all committed to making sure that my project was the best that it could be. He was always had concrete and useful suggestions for my writing, and saved me a great deal of time in suggesting the most efficient way of communicating my ideas.

I would further like to thank the other members of my committee. Ben Eggleston, for all kinds of advice while department chair, and for being an outstanding example of what a prepared, professional professor should be. Jack Bricke, for an extremely productive independent study on Davidson, a brilliant Hume seminar, and very thoughtful commentary on my project over the years. Tom Tuozzo, for a great Aristotle seminar during which I first entertained notions of working on *akrasia*. Sarah Robins, for jumping on board only weeks after arriving at the University of Kansas, and providing excellent feedback on further research directions. Finally, I would like to thank Mabel Rice for taking time out of a very busy schedule to serve on both my Comprehensive Oral Exam board and also my defense committee. Also, I want to thank Dr. Rice for providing a home for me outside of the philosophy department in the Language Acquisition Studies Lab, where I had a great time and learned a great deal.

My thanks also to our administrative staff: to Cindi Hodges, with my apologies for all of the bizarre paperwork that my work in the LAS Lab and the College of Electrical Engineering and Computer Science necessitated, and to Morgan Swartzlander, who was extremely helpful in

guiding me through all of the administrative necessities of the dissertation and graduation process.

My thanks also to my fellow grad students, for being willing to have ideas bounced off of them and for always being encouraging and supportive. Thanks especially to long-time officemates Clark Sexton and Doug Fishel.

I'd like to name every teacher and professor I've ever had, as they have all had a hand in my academic career, but there are too many to list. Any of them that still remember me over the years know who they are.

I would like to thank my family, especially my mother and father, for never even questioning why I wanted to get a PhD in philosophy. My wife, Joanna, deserves the most heartfelt thanks of all, as nothing would have been possible for me without her constant support and confidence.

BSG

CONTENTS

"At dawn of day, when you dislike being called, have this thought ready: "I am called to man's labour; why then do I make a difficulty if I am going out to do what I was born to do and what I was brought into the world for? Is it for this that I am fashioned, to lie in bedclothes and keep myself warm?"

"But this is more pleasant."

"Were you born then to please yourself; in fact for feeling, not for action? Can't you see the plants, the birds, the ants, the spiders, the bees each doing his own work, helping for their part to adjust a world? And then you refuse to do a man's office and don't make haste to do what is according to your own nature."

"But a man needs rest as well."

"I agree, he does, yet Nature assigns limits to rest, as well as to eating and drinking, and you nevertheless go beyond her limits, beyond what is sufficient; in your actions only this is no longer so, there you keep inside what is in your power. The explanation is that you do not love your own self, else surely you would love both your nature and her purpose.""[1]

[1] Marcus Aurelius, *Meditations*, Book V

INTRODUCTION

Akrasia, a Greek term often translated 'weakness of will' or less often 'incontinence', means, in a nutshell, acting intentionally against your own better judgment. *Akrasia* represents a philosophical puzzle because it pits two very plausible statements against one another. It seems obviously true that when we view one option as the better or best option, we are motivated to pursue that option instead of the inferior one. We do not often choose lesser value over greater, nor do we choose the worse to the better if we can help it. This would seem to make it impossible to act intentionally against our own better judgment. It seems just as obviously true that we often fail to live up to our own standards for ourselves. We do not always find the courage to be honest despite its seeming to us the right thing, we do not always stick to our diets or exercise regimens despite forming such regimens for the best of reasons, and we sometimes put off important projects in favor of things that we ourselves know to be less important. What is the difference between cases in which we act in accord with our better judgment and cases in which we do not? What can explain this common part of human experience?

Akrasia posesses a long and distinguished history in Western philosophy. Indeed discussion of *akrasia* as a philosophical puzzle goes all the way back to its Socratic origins. Socrates denies the possibility of *akrasia* while Aristotle devotes a substantial part of the Nicomachean Ethics (the first book in the Western tradition treating ethics as its own subject) to defending the possibility of *akrasia*. *Akrasia* is mentioned as a difficulty in the writings of the Apostle Paul, and it is puzzled over by Augustine and Aquinas. Disputes concerning the possibility and/or intelligibility of *akrasia* continue through the 20th century in exchanges between Hare and Davidson, and contemporary writers like Sarah Stroud, Michael Smith, and

Ralph Wedgwood continue to examine whether *akrasia* is possible, and if it is, how to explain *akrasia* in a way consistent with acceptable theories of normative motivation.

In the first chapter of this work I supply a conceptual account of *akrasia* that has appeared reasonably consistently through the history of philosophical discussion of *akrasia*. I also supply a set of common examples of *akrasia* that have had historical currency, and demonstrate that they share a common set of criteria. The main purpose of this effort is to show that the concept of *akrasia* in which I am interested is the same concept as written and thought about by the aforementioned luminaries of Western philosophy from Socrates all the way through to the present. A broad reading of ancient, medieval, and recent accounts of *akrasia* reveals that in all the basics, this is the case. 'The basics' as I put it, are a set of features that would apply to any alleged instance of *akrasia*. The features are these:

(i) *Akrasia* is action against a belief about what is better or best to do.

(ii) *Akrasia* is irrational.

(iii) *Akrasia* is voluntary.

(iv) *Akrasia* is blameworthy.

(v) *Akrasia* is episodic.

In the second half of the first chapter, I discuss common examples of *akrasia*, matching them up to said features, and demonstrate the long-standing philosophical interest that *akrasia* has generated.

Though it is important to have a clear view of *akrasia* as a concept, my primary goal in this work is to bring together long and productive philosophical discussion of *akrasia* with modern empirical psychology.

In Chapter 2, I propose an answer to both of the most persistent questions concerning *akrasia.* Those questions are, recall, is it possible and if it is how do we explain it? As my goal is to empirically inform the philosophical discussion (as well as to philosophically inform the empirical discussion) I address these two questions more or less at the same time. If I can identify in empirical research a plausible psychological explanation for what philosophers have termed '*akrasia*' then I have at once provided an explanation for *akrasia* and a very good reason to believe that *akrasia* is possible—that it is actual.

Very little empirical literature deals with *akrasia* by that name, so I have had to piece together a variety of different research areas in order to provide a plausible explanation of *akrasia*, which I will briefly summarize here.

My empirically informed account of *akrasia* (as far as I know, the first of its kind) relies on three avenues of empirical research: heuristic decision-making, cognitive bias, and Global Workspace Theory. Aside from being an empirically informed account of *akrasia*, my approach deviates from tradition in that in some ways, I do not view *akrasia* as a defect or a failing, or at least not entirely such. While people have good reason to avoid *akrasia* and should wish always to act in accord with their own best judgment, *akrasia* is a by-product of cognitive properties that we should not wish to be without.

Briefly, we live in an information-rich external environment and it would be hopeless to expect a finite system like the human mind to deal with all of that information. So it is useful to us to have mental systems that help us to focus on some things rather than others, to notice some things rather than others, and to react certain ways to certain parts of our environments while ignoring other parts. Gerd Gigerenzer and his colleagues have pioneered research into the "fast and frugal" heuristics (informational shortcuts) that the human mind uses to accomplish complex

tasks quickly with a minimum of mental resources necessary. These heuristics are immensely beneficial to us, but they don't always get it right. In circumstances in which our heuristics tend to lead us to ignore what we ought to attend to or focus on the wrong aspect of a situation or misestimate a probability, we refer to the heuristic as a cognitive bias—an instance in which we have a tendency to be irrational. Tversky and Kahneman are pioneers in the field of cognitive bias, and their work provides the second leg of an empirically informed account of *akrasia*.

Not only is our external environment more information-rich than we can realistically handle, but our internal mental environments (all of our thoughts, feelings, and mental processes) are similarly information laden, and just as we must have mental systems that allow us to deal with the volume of information bombarding us from without, we must similarly have systems that perform the same tasks for our internal environments. We are not always aware of all of our mental states at all times, and some things seem to happen automatically while others require attention and concentration. The idea that we can only be aware of so much information, even about our own mental states, at any given time is the central insight of Bernard Baars' Global Workspace Theory.

In combination, these insights provide an explanation for acting intentionally on the basis of one normative criterion (utilizing a one-reason heuristic in Gigerenzer's terminology) while a more careful analysis (one in which more normative criteria are present in the global workspace of the mind) would regard a different normative criterion as the more important criterion, allowing the bias in favor of attending to the first criterion to be overcome.

For example, the sweetness of a candy might commend itself to a dieter much more quickly than the dieter could focus on the caloric content of the candy and his judgment that, though the candy is indeed sweet, it is better to refrain. In such a case, the dieter breaks his diet

while judging that it is better to maintain the diet. This is of course a different case from the dieter consciously rationalizing his decision. Most often, one-reason heuristics do their work below the threshold of consciousness, or outside the global workspace. As a result the dieter is in the position of breaking his diet without paying enough attention to what he is doing. This does not mean the dieter does not know he is breaking his diet, rather he simply pays it too little mind as it is happening. The fact that these judgments are *often* not subject to conscious scrutiny does not mean that they *cannot* be subjected to conscious evaluation. Indeed, it is the person who is more often able to pay attention to what they are doing who is most resistant to *akrasia*.

Adopting this empirically informed account of *akrasia*, what I term 'cognitive *akrasia*', requires abandoning positions that are incompatible with its existence. Earlier, I mentioned that the possibility of *akrasia* tends to fly in the face of the fact that we most often do what we think we ought to do, and when we do not, it doesn't make sense. This is the view that normative judgments are necessarily connected to motivation (a view called 'normative judgment internalism'). In Chapter 3, drawing on some works by Sarah Stroud and Ralph Wedgwood, I demonstrate that normative judgment internalism can be amended to allow *akrasia*, long thought to be a straightforward counterexample to normative judgment internalism, while preserving what is plausible about the view. In the process of reconciling the possibility of *akrasia* with the notion that when we judge that A is better than B, we pursue A if we can, I must describe precisely what kind of irrationality is involved in instances of *akrasia*.

Of course, modifying normative judgment internalism is only one approach to accommodating *akrasia*. One might just as well deny the truth of normative judgment internalism altogether. An approach that denies the truth of normative judgment internalism is referred to as normative judgment externalism. Externalist views generally have no problem

accommodating *akrasia*. A common externalist position subscribes to the Humean Theory of Motivation, which draws a line between beliefs and desires, and holds that only desires can motivate. *Akrasia*, on this externalist view, is not such a puzzle. If normative judgments are beliefs, and beliefs do not motivate, then it is no great surprise when one acts against their normative judgments. It is a matter of the *akrates* thinking that one action is better than another without desiring to do that better action.

There are many subtle distinctions among varieties of normative judgment internalism and externalism. Chapter 3 first discusses a variety of methods for preserving what is plausible about normative judgment internalism while still allowing for *akrasia*. This effort is followed by an examination of several plausible forms of normative judgment externalism, and the problems that attend to each.

By the end of the third chapter, I have two competing theories of motivation that can each accommodate *akrasia*. The internalist theory posits a defeasible connection between cognitive rationality and motivation, while the externalist theory posits a defeasible connection between desires (which are by definition motivational) and normative judgments. These theories are in opposition to one another, because the internalist theory holds that the normative judgment is prior to the motivation while the externalist holds that the motivation is prior to the judgment.

In the fourth chapter, I provide an argument for preferring the internalist, cognitive view over the externalist, non-cognitive view of normative motivation. If the externalist position is correct, then the way to combat *akrasia* is to work at changing a person's desires so that they want to act as they think they ought. If the internalist position and the cognitive bias account of *akrasia* is correct, then cognitive bias modification ought to have the best outcome in combatting *akrasia*. This is an empirical question.

The trouble is that *akrasia* is an episodic phenomenon, and so it is difficult to specify when *akrasia* has been overcome and when it has not. This is why I appeal to research on addiction. Addiction, aside from being very well-researched, provides an excellent test case for my argument because it involves recurrences of the same kind of *akrasia* in an individual over time. The *akratic* addict is one who judges that they ought not to continue being an addict, but who relapses into the addictive behavior. If the cognitive bias modification treatment model is more effective than desire modification in reforming the *akratic* addict, then we have reason to treat the internalist view as the correct view of normative motivation.

The research does indicate that cognitive bias modification treatment is more effective than desire modification, and the reasons for this are very interesting. As it turns out, cognitive bias plays a large role in addiction of any kind. Specifically, addicts are less likely to recall the negative consequences of past addictive behavior and are also more likely to notice or fasten onto addiction-related stimuli in their environments. These biases are known as recall bias and attention bias, respectively. These biases play a prominent role in all addictions, whether they involve an addictive substance or not.

Common thinking about addiction gives too much thought to chemicals involved in addiction, and too little thought to what a gambling addiction and a cigarette addiction have in common. People who try to break addictions must all come to grips with the fact that it's not about just a chemical. People who break a caffeine habit must make other behavioral changes, like switching to decaf coffee, because it is more difficult to give up coffee drinking than it is to give up caffeine. Smokers who try to quit will tell you that nicotine patches or nicotine gum are helpful but not ultimately effective because it is the activity of smoking that they miss, not necessarily the chemical.

My grandfather's case is typical of those who have successfully quit smoking. He smoked for 40 years, and when the health concerns of smoking were becoming more public and accepted, he decided to quit. Patches or gum were not then available. He soon found that he had to replace smoking with other things to successfully resist relapse. After meals, when he used to smoke, he chewed gum, to keep his mouth occupied (always Wrigley's Doublemint, until the end of his life nearly 40 years later), and took up handicrafts to keep his hands and mind occupied until the craving passed. The necessity for these measures persisted well after physical withdrawal symptoms had ceased.

The cognitive bias account of *akrasia* provides an empirically informed explanation for the above anecdotal evidence. Once cognitive biases can be implicated in *akratic* addiction, the research shows not only *that* cognitive bias modification yields better results for *akratic* addicts than desire modification, but *why* cognitive bias modification yields better results for *akratic* addicts. I believe that the cognitive bias account of *akrasia* provides good reason to prefer the internalist account of normative judgment to the externalist account.

In this work as a whole I set out to address the major questions surrounding *akrasia*. First, I develop a common notion of *akrasia*, and then furnish an empirically informed explanation of *akrasia* that demonstrates that *akrasia* is indeed intelligible and actually a part of common human experience. In addition to that, I develop an account of normative judgments that can accommodate *akrasia* in a way consistent with the empirically informed account of *akrasia*. *Akrasia* is a long-standing philosophical puzzle, and to my knowledge, there has been no attempt to articulate an empirical explanation for *akrasia*. In addition to providing such an account, I also demonstrate that the account is fully compatible with an acceptable philosophical notion of normative motivation.

CHAPTER 1: WHAT COUNTS AS *AKRASIA*?

The above passage finds the Roman Emperor Marcus Aurelius struggling with himself to get up in the morning. This sort of experience is not unique to emperors, and his words speak to us across the better part of two millennia to remind us that in many ways, the more things change, the more they stay the same. The passage is a dialog with only one speaker, Aurelius, and we recognize this internal dialog as a familiar element of our inner lives. As familiar and commonplace as is this description of dragging (persuading, even) oneself out of bed each morning, it subtly reveals much about human motivation. The voice in favor of staying in bed utilizes multiple approaches. The first, "But this is more pleasant" is a base appeal to visceral pleasure, while the second statement, "But a man needs rest as well" is a subtle rationalization designed to subvert the argument that it is part of the proper and natural function of humankind to get up each morning and work. More than anything, this passage illustrates a tension between immediate inclinations and reflectively considered judgments in a context in which it is intelligible for either to prevail.

Aurelius here has a choice to make. Aurelius' beliefs about which choice to make seem clear; however it is far from certain that Aurelius will actually get out of bed rather than lying in bed longer. Should Aurelius act against his judgment of what is best, his action is an example of what the Greeks called *akrasia*. Discussion of *akrasia* first appears in Western philosophy as a disagreement between Aristotle, expressed in Book VII of his Nicomachean Ethics (Aristotle 99-113) and an opinion that Socrates was said by Plato to have held in the dialog with Protagoras (Plato 45-117). The disagreement, simply put, is whether anyone intentionally does that which they know to be wrong. The disagreement put into those terms is in no way exclusive to philosophers. Nearly everyone has at some point stopped to consider whether anyone, even history's most notorious evildoers, willingly does what they believe to be wrong. Any attempt to

provide accounts of how or why people make decisions and act as they do is incomplete without at least some consideration of *akrasia*.

Discussion of *akrasia* in Western Philosophy extends back to nearly the beginning of it. In all that time, it has somehow resisted becoming a dead issue, or an issue of interest only in the history of philosophy. There is a remarkable degree of historical agreement on what *akrasia* is. Most disagreements concern whether it happens, and how to explain it. In this chapter I will first describe the central features of *akrasia* that run throughout the philosophical tradition, and second, discuss a host of examples that have been historically prominent in discussions of *akrasia*. When these examples occur in philosophical writing, those who cite them generally do so in order to set up specific disagreements. It will be useful to separate these examples from their contentious contexts so that we may develop the clearest picture of what *akrasia* is, apart from questions of whether they accurately describe a genuine part of human experience, or whether they properly fit with theories of motivation and action.

Traits of *Akratic* Action

The following traits are not intended to be an entirely strict set of necessary conditions (though I think that in combination, they are sufficient) for akrasia. However, the following traits have traditionally formed the most notable talking points. Much writing has concerned the extent of one or more of the below traits. Those who have questioned the necessity of one of the below traits have typically acknowledged that they responded to a traditional view that regarded the trait as necessary. For my own part, I think there are at least some interesting cases of *akrasia* that arguably do not possess one or more of the below traits. In any case, the most common and interesting cases of *akrasia*, those cases that are so often used as a starting point, tend to share the following traits:

(i) *Akrasia* is action contrary to one's own normative judgment (i.e. judgment of what is best).

This is the overwhelmingly common phrase that most philosophers use to define *akrasia*, though examples and discussions of *akrasia* reveal other traits paired with the mere fact of action against one's own normative judgment.[2] This is the place to start when looking for a case of *akrasia*. Other features, discussed below, are key in distinguishing *akrasia* from other actions contrary to one's own normative judgment.

There are several key components to this characteristic of *akrasia*. First of all, an instance of *akrasia* is an action. Traditionally, having thoughts that one thinks one ought not have does not count as *akrasia*, though I can imagine an account of having thoughts that one judges that one ought not have being very similar to an account of *akrasia*. For the purposes of this analysis, we will restrict our inquiry to the things that people do.[3] This is quite diverse enough, as it includes violation of both positive and negative normative judgments, that is, doing things that one judges one ought not do, and failing to do things that one judges one ought to do, respectively.

Further, actions do not include mere physical movements, but can include the making of noises, speech acts, and other feats of more and less ambiguous non-verbal communications. Additionally, we ought to be generous in a sense and count attempts to contravene normative judgment as instances of *akrasia*. For example, consider a man who judges that it would be best not to open the door on the right. If he tries to open the door on the right and finds it locked, I

[2] For examples: "*Akrasia* is exhibited in behavior which is contrary to one's better judgment." (Mele, Self-Control, Action, and Belief 169), "There is a long tradition which views the problem of weakness of will as the problem of how agents can intentionally do what they consider wrong." (Jackson 1), "Weakness of will is possible…because it is possible to act freely against one's own best judgment…" (Buss 13)

[3] In (Incoherence and Irrationality 193)Davidson supplies an example of *akrasia* as an example of irrational action, as opposed to irrational belief or irrational emotion.

think it is still fair to say that his action counts as an instance of *akrasia*. He intentionally *tries* to open the door, and the impediment to his action is external to him. He initiated an action that, if successful, would have contravened normative judgment. He gets no credit for having been thwarted by the locked door.

In general, a good criterion for the kinds of actions with which we shall concern ourselves are those actions that are in principle observable to others. This criterion generally limits the discussion to actions as distinct from thoughts and to attempted actions as distinct from intentions.

That leaves the phrase 'contrary to a normative judgment' next in line for some clarification. 'Contrary' is reasonably straightforward, indicating that one logically cannot perform the action while following the normative judgment. For example, if Frieda judges that it would be better to give money to Oxfam than to buy a piece of luxury consumer electronics, but then buys the piece of luxury consumer electronics, this would count uncontroversially as an example of *akrasia*. Frieda could not have followed her judgment and bought the piece of consumer electronics at the same time.

The sort of judgment to which Frieda acts contrary is also important. Frieda might judge that some painting is larger than another. This is a judgment, but not a normative judgment. If she judges that the painting is a good painting, then she has made a normative judgment, but not the sort of judgment to which an action (or attempted action) could be contrary. So to be an instance of *akrasia*, some action must be contrary to a judgment about what one *ought or ought not to do*. There are of course many ways of phrasing action contrary to a normative judgment. The *akrates* could fail to act as she judges that she ought, she could act as she judges she ought

not, she could choose the inferior of two or more options, or she could fail to choose the better or best of a number of options. All of these count as action against normative judgment.

Finally, throughout this work I shall take normative judgments to be beliefs. I will not take this stance without some argument, but I will not argue extensively for this stance at this time. At this point, suffice it to say that I regard normative judgments as primarily cognitive. It may or may not be the case that a person cannot form a judgment that x is better than y without having some feeling, desire, or other non-cognitive state toward x and y. I think that this is not the case, but my account of *akrasia* does not turn on that issue.

(ii) *Akrasia* is irrational.

This trait is an evaluation that is nearly universally applied to cases of *akrasia.* This is because accounts of *akrasia* don't sound altogether sensible to any party (the *akrates* or anyone else) when recounted. Something like, "I knew I shouldn't have done that, but I did it anyway" cry out for explanation.

'Irrational' here is ambiguous. For the moment, it is sufficient to say that any action against one's judgment of what is better or best constitutes action against one's own understanding of the reasons that one has for acting in a particular way, and is hence irrational action in that sense, though there are other senses. In Chapter 3, I specify the particular dimensions of irrationality involved in instances of *akrasia*. For now, it is enough to recognize that there is something odd about a case like Frieda's above, and any other case of *akrasia*. It is not that she acts for no reason whatsoever; in some sense she has a pro-attitude toward both available actions (and presumably many others). So the act of buying the piece of consumer electronics is not in itself unintelligible, but when coupled with a genuine judgment that it is the inferior option, Frieda should acknowledge something irrational in her behavior. Certainly, she

often acts in accord with her normative judgments and thinks nothing of it. After all one is *supposed* to do what one judges is best.[4]

(iii) *Akrasia* is voluntary.[5]

This trait separates instances of compulsion from instances of *akrasia*. There is little that is philosophically interesting about defying our normative judgments in cases in which we are physically, psychologically, or otherwise unable to follow our normative judgments. In such cases the normative judgment is irrelevant to the action. It is only when one can but does not act in accord with normative judgment that one is *akratic*.

Suppose that Aurelius is called in the morning, feels like staying in bed, but judges that it is best to get up and take up his work, and as he begins to get up, realizes that he has been tied firmly to the bed and so fails to get up when called. This is not an example of *akrasia*, as the act of staying in bed was involuntary. Only when Aurelius is able to get up but does not is his decision an example of *akrasia*.

There are, of course, many gradations of meaning inherent in 'voluntary'. I shall be unable to avoid committing myself to the view that at least some actions are voluntary in some sense of the word, but I hope to largely avoid narrow disputes about voluntary versus involuntary actions. I may avoid such disputes because I contend that the preponderance of cases of *akrasia*

[4] Nomy Arpaly supplies an excellent examination of the requirement that *akrasia* be irrational. I address this work more fully in Chapter 3, but consider: "A rational agent's manual cannot instruct the agent to act against her best judgment… A rational agent's manual is a deliberator's manual, and acting against one's best judgment is not the sort of thing one settles on doing as a result of good deliberation." (490)

[5] For some, this is a point of emphasis in the characterization of *akrasia.* See for examples: "An agent, succumbing to last ditch akrasia, freely, knowingly, and intentionally performs an action A against his better judgment that an incompatible action B is the better thing to do" (Walker 653), "In a case of weak-willed action the agent acts—freely, deliberately, and for a reason—in a way contrary to his best judgment, even though he thinks he could act in accord with his best judgment." (Bratman 153), "Weakness of will occurs only if one knowingly does something contrary to one's better judgment. We will see that this condition does not distinguish between weakness and compulsion." (Watson, Skepticism About Weakness of Will 316) See also (Mele, Is Akratic Action Unfree?).

are uncontroversially voluntary for most viewpoints that hold that there are voluntary actions at all.

The cases of *akrasia* can of course be stretched further and further until we are faced with an action that is not clearly voluntary, or even that is clearly involuntary. Consider Aurelius' case again. It may be that he is fatigued from a particularly strenuous day as Rome's philosopher/king. In common experience, bodily and mental fatigue can be overcome, to a point. Given the way I initially describe the case of Aurelius, I take it for granted that Aurelius is not so fatigued that it is not physically possible for him to accede to his better judgment and arise. I think such a stipulation is reasonably uncontroversial, and in such a case, if Aurelius remains in bed, it is a clear example of *akrasia*.

I can, however, imagine a kind of sorites paradox in which the details of the case are stipulated differently. One could make Aurelius just a bit more fatigued and yet still hold that his staying in bed contrary to better judgment was voluntary. One could increase the increments of tiredness (however such increments are conceptualized) until one reaches a point at which it is not physically possible for Aurelius to arise, so great is his fatigue. At this point we have a set of minutely different cases of *akrasia* culminating in a case that is involuntary.

Such a line of cases does not show either that there is no robust concept of voluntary action, or that cases of *akrasia* cannot be clearly demarcated from cases of "I judge that it is better to do A than not to do A, but I cannot do A." A sufficient number of minute changes can lead to a significant difference. I contend that there are sufficiently numerous cases of *akrasia* that are examples of voluntary actions if anything is voluntary.

(iv) *Akrasia* is blameworthy.

This is another common evaluation of *akratic* action, and the trait falls out of *akrasia* being voluntary, against normative judgment, and irrational. There is a difference between acting against a normative judgment that others hold and acting against one's own normative judgments. While others may blame one who acts against their view of what ought to be done, both the *akrates* and some other party can justifiably disapprove of one who does not even live up to their own standards. If Aurelius decides to stay in bed despite his judgment that it is best to rise when called and see to his natural purpose, Aurelius himself regrets such a decision. Not only does Aurelius express regret, but anyone informed of his situation would reasonably express disapprobation of such a case of one who "knew better, but did it anyway".

The blameworthiness of *akrasia* is in proportion to the extent to which *akrasia* is intentional. We do not generally condemn people for acting in ways that they cannot control, though we sometimes blame those people for forming bad enough and strong enough habits that they cannot any longer act as they judge best.

(v) *Akrasia* is episodic.

Nowhere in the history of philosophy is there widespread discussion of persons who *always* act against their better judgment (Stroud, Weakness of Will and Practical Judgment 143-144). Rather, *akrasia* is a behavior that may beset us all at some point, and may be a more frequent problem for some than for others.

It is true that Aristotle discussed *akrasia* as a kind of character trait, a "weakness of will" with which some are more strongly afflicted than others. If this is the case, then *akrasia* is a more or less permanent condition for at least some. The contemporary writer uses the term *akrasia* to refer to a type of action, while the ancient writer used the term to denote a character

trait. However, I think this is not a particularly important distinction because many virtues or vices are dispositional.

Consider, for example, the virtue of generosity. A person can be generous without giving at every single opportunity, so to talk of generosity as a dispositional character trait (those who are generous give in many circumstances, and always give in the best of circumstances) comes to much the same as discussing any given action as either being an instance of generosity or not being an instance of generosity. The instances in which the generous person gives, and the actions that we consider tokens of the type 'generous' substantially (perhaps completely) overlap.

Mutatis mutandis, we can consider *akrasia* as a dispositional character trait or as an action type without losing anything critically important about *akrasia* itself. In any case, discussions of universal *akrasia* are not in the traditional or contemporary discussions of *akrasia*, leaving actual cases of *akrasia* to be discrete puzzles which share the criteria of being episodic as well as the criteria previously mentioned.[6]

Considerations of actions that fit the above criteria would be considerations of a robust and interesting concept of *akrasia*. However, there is a long philosophical tradition concerning discussion of *akrasia*, and if my treatment of *akrasia* is to be most useful, then it must interact with this tradition. The claim that at least the major treatments of *akrasia* in the philosophical tradition are employing the same kind of concept as one another when they discuss *akrasia* is not a trivial claim. Philosophical careers are made tracing the differences between contemporary and

[6] The kinds of everyday examples of *akrasia* are generally episodic things, for examples: "The motivation of weak behavior is generally familiar and intelligible enough: the desire to remain in bed, or the desire for another drink." (Watson, Skepticism About Weakness of Will 316), "…yielding to temptation, procrastinating, eating beyond healthy limits, avoiding pain, and even abandoning the greater good for the sake of pleasure. This phenomenon is the phenomenon of weakness of the will or *akrasia*." (Tenenbaum, The Judgment of a Weak Will 876)

historical ideas about many concepts. What follows should establish that *akrasia* as a concept has changed little, even if the reasons for being interested in *akrasia* have been variable. Additionally, it is instructive to encounter the difficulties faced by historical philosophers in fitting *akrasia* into larger theories about motives and actions.

Before beginning this historical exegesis, a brief word on terminology is in order. The beginning researcher should not expect to get very far past the Greek philosophical discussions of *akrasia* if '*akrasia*' is the primary search term. The Greek term *akrasia* is most often translated into English in one of two ways. One way in which *akrasia* is translated is as 'incontinence'. This term in English suggests a lack of self-control, but (to the contemporary ear) too specifically refers to a lack of control of the bowels, and as a result, this translation has become the less common one in philosophical writing. The more common translation of the term *akrasia* is 'weakness of will'. Certainly the *akrates* (the person who acts intentionally against better judgment) is criticized for it, but whether or not *akrasia* necessarily involves some kind of *weakness* is a matter of dispute, and a matter of dispute that users of the term 'weakness of will' do not always enter into; often they use it as a term of art that means something different than the meanings of the individual words in the phrase would suggest. Additionally, the term 'weakness of will' is quite often used by philosophers who do not otherwise speak of The Will[7] as a human faculty, and who do not even mean to suggest anything of the kind in the use of the term. In philosophy, then, the terms 'weakness of will' and 'incontinence' have become terms of art that simply stand in the place that the word *akrasia* would stand. Because of this, I will

[7] Donald Davidson, for example titled his very influential paper about *akrasia* "How is Weakness of the Will Possible?" but in the paper "Intending" explicitly stated a desire to avoid talk of The Will (87-88).

simply use the Greek term.[8] '*Akrasia*' is the name for the concept of acting intentionally against one's better judgment; an *akrates* is one who acts *akratic*ally.

Examples of A*kratic* Action

The episode from Aurelius' Meditations that begins this chapter is a look at just one person's experience of *akratic* action. As mentioned previously, there has been a more or less continuous discussion of *akrasia* and closely related issues for the last two and a half thousand years of Western philosophy. As evidence for the claims I make above concerning the commonly held features of *akrasia*, I submit a brief description of the examples of *akratic* action that philosophers have commonly discussed. This section is primarily expository. Following presentation of these examples, I shall advance some arguments concerning what we are to learn about investigating *akrasia* from these historical approaches.

The Dieter:

Aristotle may surprise any who believe that obsession with body image, health, and "eating right" are issues unique to the recent world. One of our earliest examples of *akratic* action is related to dieting. Aristotle famously describes an *akratic* state as similar to madness, sleep, or drunkenness, but when he requires a concrete example of *akrasia* to discuss, he considers a man on a diet. The dieter makes his first appearance as Aristotle distinguishes the vice of intemperance from *akrasia*. "[The intemperate person] thinks it is right to pursue the pleasant thing at hand; the incontinent person, however, thinks it is wrong to pursue this pleasant thing, and yet pursues it."[9] Aristotle also asserts that intemperate persons feel no regret about their actions because they simply do what they want without considering things further. In contrast, the *akrates* is always prone to regret. So the intemperate one will probably not diet at

[8] William Charlton does this in his book, and says that it is becoming standard for philosophers to do so, but still he entitles his book on *akrasia* "Weakness of Will".
[9] Nicomachean Ethics: Book 7, Chapter 3, Section 2 (102)

all, while it is the dieter's commitment that sets him up as a potential *akrates*. We modern readers imagine the dieter breaking his diet and becoming upset with himself, and we imagine something very familiar to us.

By way of further distinguishing *akrasia* from intemperance, Aristotle asserts that the *akrates* is better than the intemperate person because "the best thing, in principle, is preserved in him."[10] This again is in accord with considered moral judgment. We tend to praise people who attempt self-betterment (like dieting) even when they occasionally slip up. We praise those who slip up less often more than we praise those who slip up more often, but even those who slip up often are at the very least trying, and so they are better than those who make no attempt to better themselves (the intemperate).

Aristotle supplies us with a metaphor and its interpretation to further emphasize the point of the previous paragraph. Aristotle quotes a man named Demodocus as saying "The Milesians are not stupid, but they do what stupid people would do."[11] Likewise, we are to read, the *akrates* is not simply intemperate, but sometimes does as the intemperate would do. The difference is that the *akrates* may be persuaded out of it. The metaphor is fairly clear. A person who is not stupid should be able to be persuaded out of doing stupid things while someone who is just stupid is not amenable to this persuasion. Now imagine an interloper into our dieting example above. They bring up what the dieter has previously said, and perhaps discuss reasons in favor of avoiding the sweets. Now, given this interloper, if the dieter still chooses to eat the sweets, then we might not believe that he ever had a hindering belief, and this looks like a case of mere intemperance. If the dieter avoids the sweets as a result of further reflection and rational

[10] NE, Book 7, Chapter 3, Section 5 (103)
[11] NE, Book 7, Chapter 3, Sections 3-4 (102-103)

persuasion, then the interloper has saved the dieter from an *akratic* episode. The potential to act in accord with reason is with the *akrates* the whole time.[12]

The alert reader will already have noticed that Aristotle's account of the difference between the *akrates* and the intemperate preserves each of the features of *akrasia* previously listed. The alert reader will, then, hopefully excuse me for being explicit on this point. If (i) *Akrasia* is action against a belief about what is better or best to do, then Aristotle's gesturing at the lapsing dieter fits nicely. Committing to a diet in the first place implies a belief that it would be a good thing to diet, and failing to follow the diet (ii) is irrational in the relevant sense. The dieter could have been persuaded into action consistent with his commitments, and was presumably not irresistibly compelled (iii) to eat the sweet which broke his diet. While the dieter is to be praised for his commitment, he can be justifiably chastised (iv) for those occasions (v) on which he breaks his diet.

The Apostle:

Thomas Aquinas wrote about *akrasia* for two ostensible reasons. The first of these is that Aristotle wrote about *akrasia*. Aquinas wrote a comprehensive commentary on the Nicomachean Ethics, and in this effort performed some considerable interpretive work. Secondly, *akrasia* was of interest to Aquinas because of a passage in the Apostle Paul's letter to the Romans:

> "For we know that the law is spiritual; but I am of the flesh, sold into slavery under sin. I do not understand my own actions. For I do not do what I want, but I do the very thing I hate. Now if I do what I do not want, I agree that the law is

[12] The phrase '*akratic* episode' may serve to make sense of a strange statement that Aristotle makes in Book 7, Chapter 8, Section 1: "[I]ncontinence is more like epilepsy" (110). Certainly it is not like epilepsy in the sense that it is a disease and people get it whenever they get it, regardless of their behavior or mental states, but rather, *akrasia* is like epilepsy in that is not a constantly apparent condition, but that it is episodic in nature.

good. But in fact it is no longer I that do it, but sin that dwells within me. For I know that nothing good dwells within me, that is, in my flesh. I can will what is right, but I cannot do it. For I do not do the good I want, but the evil I do not want is what I do. Now if I do what I do not want, it is no longer I that do it, but sin that dwells within me. So I find it to be a law that when I want to do what is good, evil lies close at hand." (New Revised Standard Version, Romans 7.14-25)

Though Aquinas appeals to a different set of theoretical commitments in accounting for *akrasia*, it is noteworthy that *akrasia* itself is the same sort of thing as it is in Aristotle. Consider again the elements of *akrasia* identified above.

(i) *Akrasia* is action against a belief about what is better or best to do.

(ii) *Akrasia* is irrational.

(iii) *Akrasia* is voluntary.

(iv) *Akrasia* is blameworthy.

(v) *Akrasia* is episodic.

The blameworthy aspect (iv) of *akrasia* has not departed the discussion. For Augustine, Aquinas and philosophers for quite a long time afterward, blameworthiness is expressed in terms of sin. To sin when one knows better is to do something irrational (which is property ii of *akrasia*), as indicated by St. Paul's exasperated "I do not understand my own actions…"[13]. If the whole of moral law is divinely commanded, and if one is aware of a part of divine law and yet acts contrary to it, we have sinned (thus property i of *akrasia*, that it is an action against better judgment). The charge of weakness involved in *akrasia* is only intelligible in those who are committed to certain normative standards, in this case a commitment to following divine

[13] This kind of introspective realization of irrationality is akin to "Recognizing in himself something essentially surd" (Davidson, How is Weakness of the Will Possible? 42).

commandment. So *akrasia*'s blameworthiness (property iv) takes the form of persons failing to act according to their own normative commitments. Anyone with an understanding of sin and sinfulness understands that sins are unworthy and undesirable behaviors (this is built into the definition of 'sin'). For example, when one lusts for another while understanding that such behavior is sinful, one has a choice as to what they end up doing (which fits property iii of *akrasia*, that *akrasia* is voluntary). If they had no such choice, then they are not culpable for their sin, but instead God (who made them the way they are) is. Despite having these passions, people do not always succumb to them or allow them to lead to sin. In that sense, *akrasia* is episodic (property v), and is not some kind of compulsive pathology.

As an historical note, it seems that the perception that *akrasia* is and has been traditionally a *moral* problem is due to Aquinas representing *akrasia* in the particular way that he does. For example, Donald Davidson's landmark paper "How is Weakness of the Will Possible?" begins with a discussion of the "traditional" representation of *akrasia* that is in the vein of Aquinas, depicting it as a struggle between the sinful passions of the body and the wisdom of divine command. While it is true that *akrasia* for Aquinas, especially as represented by the apostle Paul's account, invariably involved transgression of divine command, Aristotle was unencumbered by any such theology. Aristotle's account of *akrasia*, while it occurs as part of a book on ethics, treats *akrasia* as a moral problem only in the sense that it is an undesirable part of character with typically undesirable consequences for the *akrates*' quality of life. Nor does Aristotle typically consider *akrasia* as an intrinsically "body versus soul" phenomenon.

I must here point out a very important distinction between ways of *explaining akrasia* and ways of *identifying akrasia*. Explanations for *akrasia* abound, from Aristotle's notion of having but not attending to knowledge, to the Judeo-Christian notion of our fallen natures, to

Davidson's distinction between all-out and prima facie judgments (all subjects to be dealt with in later chapters). The identifying features of *akrasia,* however, remain startlingly consistent over time and school of thought. When someone acts contrary to their better judgment, and could have acted in concert with their best judgment, they are *akratic*, and subject to disapprobation. It is to identification of *akrasia* that I now return.

The Lover:

Few things spark discussion of irrational and/or self-destructive behavior more than mention of people in love and in circumstances hostile to the flourishing of said love. One of many potential examples of *akrasia* which springs from this fertile ground is identified by R. M. Hare. Hare states early in "Backsliding", Chapter 5 of Freedom and Reason, that he "approaches these questions in a way that goes back to Aristotle and beyond, but has been associated especially (how justly, I do not know) with the name of Aquinas." (Hare 69) The examples of *akrasia* or weakness of will that Hare points to a page later (and describes as "extremely well-worn") are first, a passage from Ovid's Metamorphosis describing Medea trying to resist falling in love with Jason. The last two lines of this passage read: "Urged this way—that way—on Love or Reason's course, I see and praise the better but do the worse." The second example Hare points to is the same passage reproduced above from chapter 7 of the Apostle Paul's letter to the Romans that motivated Augustine and Aquinas to consider *akrasia*.

For Hare, following Aquinas, *akrasia* is a problem specifically for moral philosophy (he most often calls it "moral weakness" rather than *akrasia*, incontinence, or weakness of the will). It is clear, however, that Hare is addressing the same bundle of properties outlined above when

he addresses moral weakness, but means to argue that a proper understanding of what a moral judgment is makes it unintelligible for all of those properties to apply to a single action.[14]

It should be no surprise that many examples of *akrasia* involve sex, because sex is at the same time an activity much sought after, and also an activity fraught with strong normative convictions concerning its practice. At the risk of being redundant, consider the case of someone involved in what they themselves take to be a case of sexual misconduct. That it is *blameworthy* is implied in the term 'misconduct'.

That it is *episodic* (i.e. that the person in question does not succumb to each and every opportunity to transgress her sexual mores) is assumed given an average person with no clinically recognized mental or behavioral disorders. Of course, compulsions can be episodic as well, but generally speaking, we view our previous adherence to our own normative judgments as evidence that we are able to continue to do so in similar circumstances. We may be wrong about this on occasion and misidentify a compulsion as case of *akrasia*, but I think it reasonable to suppose on the basis of previously followed normative judgments that episodic failures to do so are candidates for *akrasia* pending further evidence of compulsion.

The regret that follows such misconduct is typically accompanied by consternation at a *violation of commitments to a standard of behavior that one was committed to before the act, during the act, and to which one is committed still* (else, what cause for regret?). This is recognition of the *irrationality* of failing to follow one's normative judgments. Also, at no point do those in such circumstances believe themselves to have been *compelled* to act as they did (again, what cause for regret exists otherwise, and what justification for disapprobation of the *akrates*?). Also, the fact that *akrasia* of this kind is episodic and not universal indicates that it is

[14] Again, more is said about Hare's specific arguments for the unintelligibility of *akrasia* in Chapter 2. For an excellent discussion of Hare's views of weakness, see (Frankena).

indeed within the general capacities of the agent in question to act in accord with their relevant normative judgments.

Identifying *akrasia*

How to tell what cases are cases of *akrasia* has not changed in two and a half thousand years. There is a set of symptoms that correspond with the presence of the above criteria. If a person chooses an action that he himself believes to be worse than another option available to him (it is the ultimate sensibility of such sentences that gives many pause), he is *akratic*. The evidence for distinguishing the *akratic* from the intemperate is as it always was. From the first-person perspective, there can be little doubt as to whether one is intemperate or *akratic*. I know my own better judgment.

It is much more difficult for a third party to distinguish the *akrates* from the intemperate. Consider an example. Austin[15] is a scholar who does research and writes on issues of distributive justice, and as such a scholar Austin has firm commitments about issues in distributive justice, even in the smallest circumstances. Austin is at a formal dinner with a group of scholars. An appetizer plate of various bite-sized pieces passes him, and as it does, he takes two pieces. As it happens, there is one appetizer per person at the table on the serving plate, and in taking two, Austin has necessarily deprived another of an appetizer. This conflicts with Austin's normative judgments concerning distributive justice. His action is voluntary; he could have taken all, none, or some of the appetizers. His action is contrary to his normative commitments, is something he would not necessarily have done every time, and opens him up to one or another kind of disapprobation from himself and from those around him. One kind of disapprobation would be the kind reserved for the intemperate person. Another kind is the

[15] The name of the scholar in this example is a nod to J. L. Austin, who provides a similar example.

disapprobation that *akrasia* begets. According to Aristotle (and I think most would agree), the latter ought to be less severe than the former, but still present. The difference between first and third-person identification of *akrasia* is exemplified in this way: Austin knows immediately that he is *akratic* rather than intemperate, for when he comes to realize that in taking two pieces he deprived someone of a piece, he feels regret, and perhaps even apologizes aloud. The observer does not have the same kind of access to the relevant information as does Austin. The observer may note visible signs of regret, might hear an apology from Austin, or might know Austin's views on distributive justice from prior acquaintance. These sorts of evidence, taken together or separately, indicate to the observer that Austin has done something that he himself would take to be the worse of several options open to him. Consider also the following exchange between Austin and his neighbor at table:

> Neighbor: Why did you take that second piece of appetizer, thus depriving Bob of his share?
>
> Austin: Oh! I apologize; I neglected to count the number of pieces and match them up to people, though I could easily have done so, and ought to have.
>
> Neighbor: So you don't think it's important to be fair even in these relatively insignificant situations?
>
> Austin: But of course I do! You know my commitments to fairness in all things as well as anyone!
>
> Neighbor: I've always heard that actions speak louder than words
>
> Austin: But the great majority of the time I believe that I do act fairly. This episode is but a momentary lapse.

An exchange of this kind should be sufficient to convince the interrogator of one of two things. Either Austin has just revealed himself as an *akrates*, or Austin is intemperate and is lying about it, perhaps to lessen the disapprobation that accompanies the action. There are not available to the third party any foolproof methods of distinguishing truth from lies in such a context. Typically, we trust that others are telling the truth, especially about their own internal states, unless there are compelling reasons to expect falsehood. In any case, people themselves know if they act intemperately or *akratic*ally and their reports provide the most ready means of distinguishing the one from the other, though this method is not without its unavoidable disadvantages. A study of *akrasia*, then, requires that to a large extent we must take seriously what people say about their own reasons for action, and the normative judgments that they hold or do not hold.

I should mention that I do not here mean to defend a general position that people are incorrigible with respect to all of their internal states, nor even to all of their normative judgments. I grant that it may be that someone might be mistaken concerning what is their own best judgment, but as long as people are sometimes or (as I think is the case) most often correct concerning their own normative judgments, we must be prepared to take peoples' reports of their own normative judgments seriously.

The Passions:

Traditionally, *akrasia* is thought to be associated with strong emotions or strong desires, or in other words, "The Passions". I must here make use of the distinction I mentioned earlier between identifying and explaining *akrasia*. I mean to argue in this section that looking for evidence of the passions at work is not a necessary part of identifying *akrasia*.

Consider how Davidson introduces *akrasia* in his landmark essay "How is Weakness of the Will Possible?":

> "An agent's will is weak if he acts, and acts intentionally, counter to his own best judgment; in such cases we sometimes say he lacks the willpower to do what he knows, or at any rate believes, would, everything considered, be better. In using this terminology I depart from tradition, at least in making the class of incontinent actions larger than usual." (21)

Again, what I hope that the reader will note is that Davidson's idea of *akrasia* looks very much like everybody else's, including the account in this chapter. In what sense then does Davidson think he is departing from tradition?

I read Davidson as separating himself from what he perceives to be two traditions in discussing *akrasia*. One of these is that he wishes to separate himself from those who represent *akrasia* as a conflict between reason and the passions. The other of these is that he wishes to separate his discussion of *akrasia* from discussions of specifically moral weakness as discussed by Hare and others. The second is one that I shall not discuss except to say that his success in that venture depends entirely upon the degree of overlap between the normative and the moral—a controversy that I shall not address here. As to the first separation between Davidson and the "tradition", I disagree that the philosophical tradition dealing with *akrasia* has been as generally concerned with strong emotions or strong desires (that is, passions) as has been commonly supposed.

Davidson summarizes the views of Aquinas, Aristotle, and Hare as being like "a battle or struggle between two contestants", each with a single argument, and one of them, "reason or passion" wins (in the case of *akrasia*, the wrong one, i.e. passion) (How is Weakness of the Will Possible? 35). Davidson is, to my mind, guilty of some oversimplification in this characterization. Aristotle, for example, never portrays an internal battle or struggle in

discussing *akrasia*, neither does he identify passions, nor strong emotional nor desiderative states as the sole culprits in cases of *akrasia*.[16] Davidson's characterization of Aquinas is closer to the mark, but still overemphasizes the role of strong emotional or intrapersonal conflict. The common and abbreviated understanding of Aquinas's position is as a straightforward account of physical passions (lust is the typical example) overcoming an understanding of what is moral (say, for example, the commandment not to commit adultery).[17] While it is true that this is one way that Aquinas would describe *akrasia*, his account includes more than this. At issue is partially that the word 'passions' has a connotation to the modern ear that does not match the way that we apply the term to what Aquinas wrote. A passion is an emotional state, and the modern hearer of the word likely associates it with an *intense* emotional state as in '*passionately* angry'. This is not necessarily the case in Aquinas's context. Aquinas follows Aristotle closely, particularly in Aristotle's assertion that the *akrates* acts as if mad or drunk or asleep. Aquinas concurs with Aristotle that intense emotional states, or intense passions, can render a person temporarily mad. But if those passions are less intense, a person may only be acting *as if* drunk or asleep, or if even less intense, might be acting as someone who has memorized something by rote, and so is speaking it, but is not attending to the meaning of what she is saying.[18]

To be mad is certainly to be in an altered cognitive state, but there are many ways to interpret what kind of alteration is madness. Contemporary folk psychology (as well as contemporary law) has a concept of "temporary insanity" usually used as a plea in a court of law. The idea of temporary insanity is taken more seriously by some than by others, but might it

[16] Aristotle does speak in a number of places of strong feelings contributing to some instances of *Akrasia*, but regards this as only a part of the phenomenon.
[17] Socrates, in "Protagoras" discusses the idea of someone being overcome by pleasure to do what they know to be evil and dismisses it (106-112).
[18] This is Aquinas's understanding of Aristotle's Book 7, Chapter 3. Aristotle's repeated example is of the drunk reciting Empedocles without attending to the meaning of his utterances.

provide a parallel for what Aristotle describes as the state that the *akrates* finds herself in? I think not. Temporary insanity in the modern sense is too strong to preserve the core features of *akrasia*. Most notably, a temporary insanity plea in court is first and foremost an attempt to mitigate responsibility for and thus blameworthiness of a given action. The *akrates* finds herself in the position of acting contrary to her better judgment, but is acting intentionally, and may not be attending to her judgment, but still possesses it. One who is mad is certainly irrational, and this is one of the core features of *akrasia*. Mere irrationality is enough to behave "as if mad" although again, Aristotle is silent as to the cause of this state of something-like-madness. Clearly, the passions may be very intense or very mild, and may cause anything ranging from near insanity to mere distraction, all of which can contribute to *akrasia*, but none of which is the whole story for either Aquinas or Aristotle.

Davidson is by no means the only perpetuator of this understanding of *akrasia*. Justin Gosling, in his book "Weakness of the Will" makes a great deal of a distinction between passionate *akrasia* and passionless *akrasia*. Gosling considers Aristotle's account as well as Augustine's account to be passionate *akrasia*, and considers that what Aquinas qua Aristotle interpreter wrote is exemplary of an account of passionate *akrasia*. According to Gosling, what puts passionless *akrasia* on the map is Aquinas's discussion of the fall of Lucifer. Aquinas was very concerned with how Lucifer, an archangel, could passionlessly choose the evil course (Gosling). His full account is interesting, but not germane to this analysis.[19] In any case, Gosling overstates Aquinas's description of human *akrasia* as passionate, where 'passionate' indicates an intense emotional or desiderative state, and not just some emotional or desiderative state. I must point out here that though such accounts are most common, *akrasia* need not necessarily involve any emotional or desiderative state whatsoever. In any event, most who

[19] See Especially Gosling's Chapter 10

focus on the role of emotion in *akrasia* tend to draw the reader's attention to a distinction between *akrasia* involving strong emotional states and *akrasia* involving mild (or no) emotional states. Gosling finds passionless *akrasia* to be the more interesting sort.

As one good example, an often cited passage from J. L. Austin serves to point out that *akrasia* can be quite calm, emotionally. Austin writes of a fondness for ice cream, and a situation in which dishes of ice cream are set at high table, one for each person present. Austin is tempted to take two portions, against his principles, and does so. He writes: "But do I lose control of myself? Do I raven, do I snatch the morsels from the dish and wolf them down, impervious to the consternation of my colleagues? Not a bit of it. We often succumb to temptation with calm and even with finesse." (Austin 24)

Christine Tappolet, contra Gosling, finds *akrasia* that involves intense emotional states to be the more interesting sort. In "Emotions and the Intelligibility of *Akratic* Action" Tappolet favors an interpretation of *akrasia* that involves emotional states, but only for what she calls "hot" as opposed to "cool" *akrasia*. Hot *akrasia* tends to involve intense emotional states, while cool *akrasia* is much like the account from Aurelius that begins this chapter. Implicit in the distinction between hot and cool *akrasia* is a recognition of the position that *akrasia* does not necessarily involve passions in the sense of strong emotions or desires. There may sometimes be a great deal of dispute between adherents to one explanation of *akrasia* or the other, but the point I wish again to emphasize is that *akrasia* is still *akrasia*, independent of how it is explained and either made consistent with or inconsistent with a broader theory of decision and action.

Perhaps I have overly belabored this point, but it is important to recognize that in looking for examples of *akrasia* in our lives and in the lives of others, we need not look for people in the grasp of strong (or any) emotional or desiderative states. We need not look for ravening lunatics,

or even for people who look conflicted. Instances of *akrasia* can range from the mundane to the life-changingly significant, but in neither case is passion a necessary symptom, nor, on reflection, has it ever been widely considered such.

Akrasia Zoo

Despite the startling agreement on a general conception of *akrasia* through a vast span of philosophical history, there have been and continue to be a variety of distinctions within *akrasia* which usually take the form of phrases prefixed to '*akrasia*', 'incontinence', or 'weakness'. I will address a few of these distinctions, but first I will make a general remark. It can certainly be said that there are strong desires and weak desires, good desires and bad desires (relative to some purpose), general desires and specific desires, etc. Philosophers do work on specific aspects of various theories and views that have to do only with some specific kind of desire versus another. These distinctions are useful for philosophical work, but do not generally imply that it is under dispute what, in general, desires are. Though desires have different uses in different views on psychology, the general concept 'desire' is reasonably coherent from one context to another.

I submit that it is the same with *akrasia*. There are many phrases prefixed to a word for *akrasia*, but this does not call into question that the word '*akrasia*' has a general and reasonably consistent use. For example, one of these distinctions is drawn by a multitude of Aristotle commentators. This distinction is between the weak *akrates* and the impetuous *akrates*. This is at heart a dispute only about what Aristotle really says about the way that *akrasia* affects practical reason. I will not strenuously take a side on this technical issue, as every position on it has already been contested. The weak *akrates* is the one who acts intentionally against her better judgment in that she knows full well that the sweet thing is both pleasant and is unhealthful, but sides with the pleasant in spite of a belief (the belief that hinders) in the superiority of the choice not to eat the sweet. Pears calls this account of *akrasia* "last-ditch" *akrasia*, because at this

stage, the agent understands everything they need to understand in order to make the choice that accords with better judgment. This has also been called "open-eyed" or "clear-eyed" *akrasia* for the same reason. The impetuous *akrates*, on the other hand, does not fully take the time to consider whether the sweet is also unhealthful, though he judges that it would be better not to eat that which is unhealthful. The weak *akrates* seems to be personified in the example with the sweets, while the impetuous *akrates* seems more like the person in Aristotle's example about the beans.[20] It may be that Aristotle really had in mind two different ways that someone might be *akratic*, or it might be that one of these sides represents the "correct" interpretation of Aristotle. Such disputes are of philosophical value, but they are not crucial to an overall account of *akrasia*. No single version of *akrasia* is any more likely to be THE proper description of *akrasia* any more than any particular chair is THE proper example of a chair.

My account of the nature of *akrasia* contains enough substance to cover a significant family of cases. I submit that all of the prefixed versions of *akrasia* share at least the features outlined at the beginning of this chapter. Most exhibits in the *akrasia* zoo are specifically constructed to explore other disputes about decisions, actions, and motivations, some of which I address in future chapters.

At the close of this chapter it is important to see *akrasia* as a generally stable concept over time in the philosophical discourse. Examples of *akrasia* all share a set of criteria, namely that they are examples of episodic rather than habitual action against a better judgment. Also, instances of *akrasia* are intentional, in some sense irrational, and are blameworthy. In identifying cases of *akrasia*, we must take seriously what people report about their own

[20] Nicomachean Ethics, Book 7 Chapter 3 Section 6 "Perhaps…someone knows that dry things benefit every human being, and that he himself is a human being, or that this sort of thing is dry; but he either does not have or does not activate the knowledge that this particular thing is of this sort." (103) In this example, beans are often the discussed example of a dry food such that a man might know the general principles but not identify beans as dry or might not identify some particular food as beans.

normative judgments, and there need not be any role for intense emotional or desiderative states. Subsets of, or specific kinds of *akrasia* have historically been of interest, and this fact may have obscured the extent to which '*akrasia*' refers to a coherent and stable concept. The examples and analysis provided heretofore should be useful in identifying *akrasia*.

In the next chapter, I shall turn to the goal of explaining *akrasia*. If *akrasia* is genuinely a part of human experience, it should be amenable to explanation consistent with and informed by our best empirical data concerning human psychology. The next chapter constitutes such an empirically informed explanation of *akrasia*.

CHAPTER 2: AN EMPIRICALLY INFORMED ACCOUNT OF *AKRASIA*

In Plato's "Protagoras", Socrates is the first recorded philosopher in a long line of philosophers who deny the possibility of *akrasia,* claiming "For no wise man, as I believe, will allow that any human being errs voluntarily, or voluntarily does evil…but they are very well aware that all who do evil and dishonorable things do them against their will" (94) and "No man voluntarily pursues evil or that which he thinks to be evil." (112) In this chapter, I intend to join an equally long line of philosophers claiming that *akrasia* is indeed possible. My account of *akrasia* will differ from those preceding mine[21] chiefly in that mine is an empirically informed account.

The first thing I must point out is that, strictly speaking, discussions of the possibility or impossibility of *akrasia* are really discussions about the intelligibility or unintelligibility of *akrasia*. For example, we say that faster-than-light travel is impossible because it is not compatible with other commitments of physics. To a physicist, or anyone sufficiently well-versed in physics, talk of getting to light speed and then just turning on a few more rocket engines to exceed light speed is unintelligible babble the proper response to which is "Look, you don't understand the way things work. That might sound plausible to you, but we have empirical evidence to the contrary."

Science fiction authors have long proposed various means of faster-than-light travel that have been variously scientifically intelligible, and so in a sense one might claim that these writers have demonstrated that FTL is possible. What I have in mind is a stronger sense of 'possible' than this. There may be modes of transport that rely on hitherto undiscovered principles or technologies or that, while not specifically prohibited by our current understanding

[21] Most prominent are: (Bratman), (Buss), (Davidson, How is Weakness of the Will Possible?), (Hare, Freedom and Reason), (Jackson), (Mele, Is Akratic Action Unfree?), (Pears), (Smith, Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion)

of physics, are also not witnessed in nature nor are created in our laboratories. Since the philosophical discussion of *akrasia* is current and has been an issue for about two and a half millennia, a claim that *akrasia* is empirically intelligible should come with some indication that *akrasia* is actual, and not merely possible.

Given the general prevalence and successes of empirical methods of inquiry, I think it is an important feature of at least some philosophical discussions that empirical data be consulted, and that philosophical concepts should be intelligible by the lights of reasonably well-established empirical work. For the *akrasia* debate, merely imagining and describing a plausible-sounding situation in which a person acts contrary to their own acknowledged better judgment is to my mind insufficient for claiming that *akrasia* is possible.

A truly rigorous defense of the possibility of *akrasia* requires some demonstration that *akrasia* comports with theories derived from empirical study of human behavior and decision-making.

I understand that not everyone shares the above requirement for a defense of the possibility of *akrasia*, and in discussing *akrasia*'s possibility, various philosophers have had various goals. I would like to answer some common charges of the unintelligibility of *akrasia*. Following that discussion, and before presenting an empirically informed account of *akrasia*, I would like to explain the ways in which I find existing accounts of *akrasia* to be insufficient in establishing the kind of empirical intelligibility in which I am interested.

One way of claiming that *akrasia* is unintelligible is claiming that it is *logically* unintelligible. That is the assertion that something about the meanings of the words in the phrase 'acting intentionally against one's better judgment' are mutually incompatible and therefore, contradictory and thus logically impossible. If this is the objection, then it is an objection that

accuses of unintelligibility any possible explanation for intentional action against better judgment. If someone presents an explanation of an event that is genuinely logically contradictory, or a set of events that is genuinely logically inconsistent, then they are speaking unintelligibly in the logical sense of the word.

Some have seemed to favor just that analysis of *akrasia*. However, some others have opted for a narrower sense of unintelligibility. This might be termed psychological unintelligibility. For these objectors, the claim is that there is nothing outright contradictory in holding that a person could willingly defy their better judgment, but rather, the claim is that those in the psychological state of judging one action better than another are psychologically incapable of acting contrary to that judgment. A being whose psychology functioned differently from our own (or perhaps a person with a significantly damaged or abnormal psychology) would perhaps be able to be an *akrates*, but not a psychologically ordinary adult human being.

Any legitimate category of human behavior, if actual, should be amenable to empirical identification and analysis. Moreover, it would be odd for empirical science to have entirely missed any common human experience. I contend that a proper and empirically informed account of *akrasia* ought to settle the longstanding philosophical question of whether *akrasia* is indeed possible.

Two Prominent Challenges to the Intelligibility of *Akrasia*

I would here like to address two of the most prominent (of which most other objections are a sub-type) challenges to the intelligibility of *akrasia*. First, I will address the objection (owed chiefly to R.M. Hare) that descriptions of *akrasia* are logically unintelligibile. Second, I will address the objection (originally forwarded by Socrates) that *akrasia* is psychologically unintelligible. Both of these famous and well-trod positions have elicited answers at least as famous and well-trod. Davidson responds to Hare's position, and Aristotle to Socrates'. These

responses are, to my mind, sufficient as far as they go, but both stop short of providing an adequately comprehensive account of the intelligibility of *akrasia*.

The Logical Unintelligibility of *Akrasia*

Hare argues that descriptions of *akratic* episodes like those in the previous chapter are unintelligible if presented as cases of *akrasia* rather than some other phenomenon. More specifically, Hare considers descriptions of *akrasia* to be semantically inconsistent, and therefore logically impossible, as talk of square circles would be logically impossible. Hare makes three claims that, taken together, serve to represent his position on this issue. The first is "The test, whether someone is using the judgment "I ought to do X" as a value-judgment or not is "Does he or does he not recognize that if he assents to the judgment, he must also assent to the command "Let me do X"?". (Hare, The Language of Morals 168) The second claim concludes the fifth chapter of Freedom and Reason. "If a man is faced with a difficult moral choice, and asks a friend or advisor 'What do you think I ought to do?', is it not sometimes the case that if he says 'You ought to do *A*', and if the man then proceeds not to do *A*, he will be said to have rejected the advice?" (85) Thirdly, Hare makes even more explicit his notion of *akrasia* as logically impossible, saying, "It is a tautology to say that we cannot sincerely assent to a command addressed to ourselves, and at the same time not perform it, if now is the occasion for performing it and it is in our (physical and psychological) power to do so." (Freedom and Reason 79) In the face of examples of *akrasia* proposed by other philosophers, (Hare cites some examples described in the previous chapter) Hare makes a second claim, that the better way of explaining the cases that other philosophers have historically identified as cases of *akrasia*[22] is to describe them as either cases of "want to but can't" or "Don't want to but can't resist" where 'can't' is a

[22] In Hare's case, he focuses on Medea, in Ovid's *Metamorphosis* and Chapter 7 of the apostle Paul's letter to the Romans.

form of psychological (or even physical) impossibility.[23] I will deal with issues of psychological impossibility in responding to the Socratic account below.

Here, I wish to respond to Hare's claims about the logical impossibility of *akrasia* in two ways. My first response is that Hare's account, taken on its own merits, ignores an important fact about identifying *akrasia* and distinguishing it from intemperance. My second response is that what Hare claims is tautologous is not so. Davidson provides what is to my mind a decisive blow to Hare's account of the logical impossibility of *akrasia*, but Davidson's account leaves out an explanation for what he claims about judgments. My account preserves Davidson's view that *akrasia* is logically intelligible and furthermore provides a model of psychology that Davidson's account could fit into.

Distinguishing *akrasia* from intemperance

My first response to Hare and to those who hold relevantly similar positions is that such positions fail to distinguish *akrasia* from intemperance. It should be no surprise that a particular description of a case should not fit the description of *akrasia* if the case as described is not an example of *akrasia*. Consider what Hare says above about the man who seeks moral advice. Hare regards it as clear from the fact that the man does not follow the advice, he has rejected the advice. Assuming that the advice is actually good advice, the man who rejects it is intemperate, but the man who accepts it and yet acts against it is *akratic*.[24] Hare apparently appreciates no such distinction, inferring immediately from action alone a person's normative commitments.

A person's actions alone are not sufficient evidence to make a determination of *akrasia*

[23] I will here reiterate that this kind of account does not distinguish cases in which, to follow the example, Paul sins and cases in which he does not. If some psychological impossibility to follow divine command is at play, the account is silent as to why Paul sometimes does and sometimes does not follow divine command, while regarding it the whole time as divine command.

[24] For a more detailed discussion of the relationship between *akrasia* and advice, see (Wiland)

or intemperance.[25] Some report from the alleged *akrates* or intemperate is required to make any determination at all. It is a defect of Hare's account, and accounts like it, that they do not allow any such first-person report to bear on the distinction between the *akrates* and the intemperate. Any account that, like Hare's, does not allow for the reports of the alleged *akrates* or intemperate is, by the above reasoning (mutatis mutandis), an insufficiently fine-grained account.

Positively demonstrating the logical intelligibility of *akrasia*

Davidson presents the following triad to set up what he construes to be Hare's reasoning concerning the logical unintelligibility of *akrasia*.

> "(**P1**) If some agent ***a*** wants to do x more that she wants to do y and she believes herself free to do either x or y, then she will intentionally do x if she does either x or y intentionally.
>
> (**P2**) If ***a*** judges that it would be better to do x than to do y, then she wants to do x more than she wants to do y.
>
> (**P3**) There are incontinent actions." (How is Weakness of the Will Possible? 23)

It is clear that, given the previous citations of Hare, that Hare would agree with P1 and P2 and would reject P3 on the ground that P1 and P2 are true and that an "incontinent action" is an action in which agent ***a*** judges that x is better than y, can do either one, and intentionally does y.

Because I consider Davidson's account of the possibility of *akrasia* to be reasonably clear in its own right, and because there are a number of quality summaries of his account[26], I shall be brief with regard to exposition of Davidson's reply to Hare. In presenting a précis of

[25] The degree of plausibility of a distinction between intemperance and *akrasia* rests on the degree of plausibility of cases like those described in the first chapter, most notable Austin's case, in which it is necessary to consider his own view of his internal states in order to properly judge, by considering a distinction between intemperance and *akrasia*, the blameworthiness of his action.

[26] The introduction to Stroud and Tappolet, Weakness of Will and Practical Irrationality (1-9) contains a very good summary.

Davidson's reply, I want to point out that I take two things to be important about Davidson's account. The first is that as a pure demonstration that the triad above is not logically inconsistent, I regard Davidson's account to be entirely successful. Second, Davidson's account must be supplemented with additional empirical data in order to provide some positive explanation of how *akrasia* can come to be. To be fair to Davidson, he did not intend to prove more than that the apparently inconsistent triad above is not, contra Hare, actually inconsistent.

To explain: Davidson renders the above triad consistent by appealing to a distinction between unconditional judgments and prima facie (or pro tanto) judgments. Unconditional judgments are of the form 'action *a* is better than action *b*'. Prima facie judgments are of the form '*pf*(*a* is better than *b*, *r*)' Which is to be read, "Prima facie: *a* is better than *b*, given *r*", where r is a reason or set of reasons for the judgment. (How is Weakness of the Will Possible? 37-39) Davidson explains:

> "But now there is no (logical) difficulty in the fact of incontinence [*akrasia*], for the *akrates* is characterized as holding that, all things considered, it would be better to do *b* than to do *a*, even though he does *a* rather than *b* and with a reason. The logical difficulty has vanished because a judgment that *a* is better than *b*, all things considered, is a relational, or *pf*, judgment and so cannot conflict logically with any unconditional judgment." (How is Weakness of the Will Possible? 39)

An action against one sort of judgment but in accord with another is an example in which each of the propositions in the above triad can be true at the same time. The question becomes, is this what is really going on when someone is *akratic*? I must say that I cannot be a very harsh critic of Davidson on this point, simply because Davidson acknowledges that these points are beyond the scope of his limited purpose, which was to demonstrate the logical consistency of the

above triad. There remains the task of explaining *akrasia* while acknowledging its irrationality. The sense in which Davison views *akrasia* as irrational is as follows:

> "But if the question is read what is the agent's reason for doing *a* when he believes it would be better, all things considered, to do another thing, then the answer must be: for this, the agent has no reason [Of course he has a reason for doing *a*; what he lacks is a reason for not letting his better reason for not doing *a* prevail.]" (How is Weakness of the Will Possible? 42)[27]

Davidson concludes: "What is special in incontinence is that the actor cannot understand himself: he recognizes, in his own intentional behavior, something essentially surd." (42)

In "Paradoxes of Irrationality", Davidson proposes that the appropriate way to explain *akrasia* while preserving its irrationality is to identify a cause of action that is not a reason for action. (176-178) For example, consider someone who buys themselves some new clothes with money that she had previously decided to give to charity, and still, on reflection thought it best to give the money to charity. Whatever desire or other pro-attitude that caused her to buy the clothes was not a reason to buy the clothes; she really judged that it would be best to give the money to charity, meaning that she had every reason to give the money to charity. This action is explainable in the sense that a cause can be identified but irrational in the sense that the cause is not a reason.

Davidson continues to outline a framework for the sort of explanation that can account for *akrasia*. We generally have no trouble acknowledging that some mental events may cause others without the first being a reason for the second. To paraphrase Davidson's example, Bob wants Amelia to visit his garden, so he plants in it a beautiful flower. Amelia desires to see the flower, so visits Bob's garden. Bob's want is the ultimate (though not proximate) cause of

[27] I have here placed Davidson's footnote in brackets.

Amelia's desire and action, but is not the reason for it. (180-181)

Note that the example is in an interpersonal rather than intrapersonal context. Davidson's purpose in appealing to the interpersonal context is to propose that one must see some sort of functional divisions, or "quasi-independent structures" in the mind to adequately explain some kinds of irrationality (like *akrasia*) so that his interpersonal context example above can translate into the intrapersonal context. (181)

Davidson's proposal for explaining irrational action is a good one, and reconciles two projects that are, at first glance, incompatible. It seems like explaining something renders it no longer irrational while something irrational is by definition inexplicable. If there can be causes that are not reasons, then actions can be irrational but still explainable, whenever one's own mental state is the cause of one's action, but not a reason for it. This is a nice conceptual outline that is missing only a description of which quasi-independent structures of the mind are responsible for *akrasia*. My full account is found below. Note that my account will rely on heuristics, cognition, metacognition, and Global Workspace as functionally quasi-independent features of the mind. I will describe how it comes to be that one mental state can be a cause of an action without being a reason for that action.

My empirically informed account of *akrasia* fills in Davidson's framework. Davidson is correct that to the average person, their failing to act in accord with what they themselves take to be preferable can be subjectively inexplicable. However, to the philosopher or the psychologist, an appropriate account of *akrasia* can provide such answers. Hare's account does not allow for a sufficiently complex picture of human psychology and decision-making, and should be displaced by an account that includes knowledge that scholars and researchers have gained by careful observation. Davidson's account anticipates, but stops short of fully describing, such an account

as I shall develop.

Concerns like Hare's with respect to the intelligibility of *akrasia* are one motivation for denying that *akrasia* happens. I think that Davidson's response has successfully made room for the possibility of *akrasia*, but I am more interested in the actuality of *akrasia*. Before presenting the empirically informed account of *akrasia*, it is necessary to present the other major challenge to the intelligibility of *akrasia* so that it is clear what kind of opposition the empirically informed account is intended to overcome.

Psychological Unintelligibility of *Akrasia*

Certainly, there is some connection between what a person thinks they ought to do and what they actually end up doing. The Socratic position with respect to *akrasia* takes this connection as inviolable. The Socratic position, and views like it, take the stance that it is a feature of human psychology that no other motivational force can overpower knowledge of the good (since all actions are aimed at the good). In this section, I shall summarize the Socratic position and then I shall present an historical response to it that is in need of some completion and modification.

The Socratic Position

In his dialog with Protagoras, in order to clarify the position that he will be critiquing, Socrates states (and Protagoras assents to) the following:

> "Now the rest of the world are of the opinion that knowledge is a principle not of strength, or rule, or of command: Their notion is that a man may have knowledge, and yet that the knowledge which is in him may be overmastered by anger, or pleasure, or pain, or love, or perhaps by fear,--just as if knowledge were

a slave, and might be dragged about anyhow… [M]en are commonly supposed to know the things which are best, and not to do them when they might." (103-104)

It is clear that what the people who hold the general opinion about knowledge are talking about is (at least in part) *akrasia*. Knowledge is (at least) belief, and in the above quotation, a belief about what is best in some cases. Yet in some of these cases, people are "overcome" by something else, and act against their belief about what is best.

I have profound doubts about any explanation of *akrasia* that relies simply on the strength of one motivational force overpowering some other motivational force[28], for the reason that such explanations strain the requirement that *akrasia* be intentional. I shall resume discussion of this issue in Chapter 4.

What is clear from Socrates' statement above is that he accuses his opponents of misunderstanding the role that knowledge plays in decision and action. He reacts with incredulity at the suggestion that knowledge could be dragged around by the passions. He accepts, but does not argue at length for (because Protagoras also accepts), a view of human psychology that prevents anyone from knowing that one thing is better than another and yet choosing the acknowledged inferior choice. Socrates and Protagoras both accept this as an intuitively obvious truth (in spite of the opinions of the general population). No doubt Socrates is correct about something here; it is indeed plausible that knowledge has something to do with why we act, but it is not so plausible to hold that such a straightforward relationship exists between knowledge and action.

Pre-theoretically, something akin to what Socrates suggests has great explanatory force. In fact, assuming that knowledge of the good necessarily motivates actions or intentions can

[28] Davidson also was skeptical of such a formulation, writing, "…it is not clear how we can ever blame the agent for what he does: his action merely reflects the outcome of a struggle within him. What could he do about it?" (How is Weakness of the Will Possible? 35)

provide a kind of sufficiency in terms of explanation of why some actions are taken or intentions formed. Imagine asking someone why she decided to stay home sick from work rather than go in on some particular occasion. If her response is that she judged that it would be better for her to stay in, we could view this as a sufficient, non-enthymematic explanation of her action in that case.

These concerns seem to effectively motivate the Socratic position of a strong and immediate psychological connection between a person's knowledge of the good and their decision to pursue it. If this connection holds, then it follows that *akrasia*, a straightforward counterexample to this connection, is unintelligible in any normal psychology. *Akrasia* has never been entertained as a form of aberrant, deviant, or abnormal psychology, but instead always as a part, undesirable though it is, of a more or less ordinary psychology. If *akrasia* were simply a form of mental illness, then it would not be a philosophical puzzle, but a psychiatric one. Philosophers have not spent any time arguing about why one's nose runs when one has a cold. The force of the Socratic objection is that if *akrasia* cannot be part of a normal psychology, then the claim that *akrasia* is possible or even actual is a claim without much impact on our understanding of the general human condition. I, of course, wish to dispute the understanding of psychology that excludes *akrasia* from the ordinary range of human experience, behavior, and mental function.

One might go along with Socrates' position for a long distance, holding the connection between knowledge of the good and action pursuing it to be very strong, yet not indefeasible. However, anyone who would claim that knowing what is good (or better or best) does not always cause us to pursue it must deal with the apparent explanatory power of Socrates' account. In other words, Socrates' opponent owes a story of how having knowledge of the good does indeed

sometimes provide us with sufficient reason to pursue it and how yet we sometimes fail to pursue it. Such an opposing account requires a more complex psychology than Socrates provides.

Just as a plausible response to positions like Hare's require a more fine-grained approach, I offer a more sophisticated psychological explanation for what is going on when a person judges that A is better than B and does B. The Socratic position treats all reasons for action as pursuing a single, overarching normative criterion—the good. This does not fit well with a more careful inspection of reasons for action. (Davidson, Actions, Reasons, and Causes)

Interestingly, Aristotle's response to the Socratic position concerning *akrasia* contains a kernel of the empirically informed account of *akrasia* that I shall soon develop. What is worth noting is that Aristotle's approach satisfies the requirement for a more complex moral psychology than that found in the Socratic position, but still contains some gaps, some deficiencies, and can be improved considerably by the application of concepts in contemporary psychology. I shall discuss briefly some of the salient features of Aristotle's reply to the Socratic position because I regard my account as a refinement of rather than a radical departure from the traditional philosophical discussion of *akrasia*.

Aristotle's Reply to the Socratic Position:

Aristotle begins with a distinction between attending to knowledge and having knowledge (without attending to it). He notes that it would be extraordinary for someone to have and attend to their knowledge and still act contrary to it (and in this way adheres to the Socratic position), but that it would not be so implausible for someone to have knowledge and act against it because he is not "attending to" that knowledge. The Socratic position makes no such distinction.

The relevant passage in Aristotle's account is important enough (and awkward enough to summarize) to quote in full:

> "Suppose, then that someone has the universal belief hindering him from tasting; he has the second belief, that everything sweet is pleasant and this is sweet, and this belief is active; but it turns out that appetite is present in him. The belief then, [that is formed from the previous two beliefs][29] tells him to avoid this, but appetite leads him on, since it is capable of moving each of the bodily parts. The result, then, is that in a way reason and belief make him act incontinently. The [second] belief is contrary to the correct reason, but only coincidentally, not in its own right. For the appetite, not the belief, is contrary [in its own right to correct reason]." (104)[30]

One of the most difficult things about dieting is sticking to a diet in the face of previous dietary habits and assumptions which are usually correct, but can sometimes contribute to leading one astray. In Aristotle's above example, the very same object (the sweet) is an object judged to be valuable based on the criterion of its pleasant taste, and judged to be of negative value on the criterion of its healthfulness (to supply a plausible reason for the admonition to avoid sweets). A dieter (by virtue of being a dieter, and not merely intemperate) regards the healthfulness of the food as the more important criterion of value. Aristotle contends that in the cases that the agent chooses to eat the sweet, they are doing so for a reason, and in accord with a judgment of theirs, and are not under any compulsion. They are not, however, acting from their

[29] This is a difficult passage. I take the meaning of the bracketed phrase as indicating that the "universal belief hindering him from tasting" and the belief "everything sweet is pleasant and this is sweet", may operate together in an agent's mind and favor whatever normative concern implied in the first belief over the pursuit of pleasantness identified in the second belief to yield an overall belief telling the agent to "avoid this" in spite of its pleasantness.

[30] Book 7, Chapter 3, Sections 10 and 11 (Irwin 1999).

better judgment. The more patiently the decision is made, the more likely that both criteria get compared, but we have all experienced moments of weakness that are quickly resolved, and that we quickly regret.

Many sweets present themselves to us in a colorful and appealing way, which draws attention to their desirability more than to their healthfulness, so it should be of little surprise that information concerning the sweet's desirability on a taste criterion should be brought to our attention first. It is important to consider that many of our mental operations are features that were/are of critical importance to our survival as a species, so a quick decision-making tool that identifies pleasantly tasting things as highly desirable is a heuristic that separates potential food sources into more and less highly caloric categories, with a strong preference for the more caloric. This one-reason heuristic is quite beneficial when food sources are scarce, but detrimental in the circumstances of extreme plenty and even profligacy that are increasingly common in the developed world. It is a common experience to act on this one-reason heuristic while believing that it is better to keep one's diet. It is classically known as *akrasia*.

In supplementing Aristotle's reply to the Socratic position with some of the language and conceptual framework of modern evolutionary psychology, I am foreshadowing the content of my empirically informed account, and demonstrating Aristotle's essential compatibility with such an account. Just as Davidson demonstrated that a more finely grained understanding of what a judgment is allowed for the logical intelligibility of *akrasia*, Aristotle similarly argues that a more finely grained understanding of the psychology of judgment and motivation makes room for *akrasia* that is not present under the Socratic position. Both Davidson's position and Aristotle's are plausible, but the question remains whether either is in any way supported by any substantial observations of modern empirical science. It is to this question that I now turn.

An Empirically Informed Account of *Akrasia*

The effort of providing an empirically informed account of *akrasia* is complicated by the fact that the terms familiar to philosophy (*akrasia*, incontinence, weakness of will) are not terms of currency in the various empirical fields of study that are relevant to this analysis. *Akrasia* necessarily involves peoples' judgments of what is better or best, i.e. normative judgments. Since psychologists are scientists, they speak descriptively, and their habits of language generally assist them in maintaining both objectivity and the appearance of objectivity, both of which are important to scientific credibility. So instead of studying a resoluteness or irresoluteness, described in terms of personal virtue or vice, a psychologist might avoid the appearance of morally criticizing his or her study subjects and examine 'ego-depletion'. (Baumeister, Bratslavsky and Muraven) Since *akrasia*, or its translated cousins 'incontinence' or 'weakness of will' all describe a personal vice, this all means that empirical accounts of what philosophers would term *akrasia* are likely to be termed something else (or will be mixed among several terms) even if psychologists were in general familiar with the traditional philosophical discussion of *akrasia*, which they probably are not.

However, this should not stop a determined philosopher (or anyone else) from crossing disciplinary borders and looking for commonality in the concepts studied, however they are termed. If my position is correct and *akrasia* is not merely possible but actual and reasonably common, then it should have some manifestation in empirical observation of human psychology. The understood mechanisms and features of observed human psychology should also suggest an intelligible explanation for *akrasia*. What would result is an empirically informed account of *akrasia* that ought to settle the question of whether *akrasia* is indeed possible by arguing that it is actual (hence possible) and intelligible on our best theories.

My presentation of an empirically informed account of *akrasia* contains three parts. The first is a review of other literature specifically intending to provide an empirically informed account of *akrasia*. The second part is my proposal that research in cognitive bias forms the most compelling way to empirically study *akrasia*. The third part presents a model of cognition (the Global Workspace Theory) that fits and explains cognitive bias as well as *akrasia*.

A comprehensive survey of empirically informed accounts of *akrasia*:

I know of only one work of empirical science focused on *akrasia* as traditionally discussed by philosophers. It is called "Weakness of will, *akrasia*, and the neuropsychiatry of decision making: An interdisciplinary perspective". (Kalis, Mojzisch and Schweizer) It is not a formal study, but rather, as the title indicates, is an article intended to bridge empirical research about defects in decision making and philosophical analysis of the same issue. This is at heart my project in this chapter. I propose to examine some of the conclusions of Kalis et al., but ultimately, I shall argue that there is too much that their account of *akrasia* leaves out.

To begin, Kalis et al. identify a "sequential model of decision making and action" and put forth three proposals for where to locate *akrasia* in the model: 1) option generation, 2) option selection, or 3) option initiation. They term *Akrasia* in option generation, "*Akrasia* as accidie"[31]. This is an explanation of *akrasia* such that the *akrates* fails to recognize the option that she would have regarded as the best option among her perceived options. To incorporate the full characterization of *akrasia*, including, notably, the stricture that *akrasia* be intentional action, the putative *akrates* must have been able to identify/consider the missing option.

Akrasia in option selection is called "decisional *akrasia"*. In decisional *akrasia*, the putative *akrates* recognizes several options and selects an action other than one that they have

[31] 'Accidie' is an alternate spelling of 'acedia', an archaic English word that suggests spiritual sloth or indifference.

some reason to prefer more than the one they actually selected. Kalis et al. leave open an explanation for decisional *akrasia*, invoking many studies that suggest that different kinds of risk/reward evaluations are carried out by different cognitive apparati in the brain. I think it is fair to assimilate this variety of possible explanations for decisional *akrasia* into the broader claim that the decisional *akrates* selects one option (including inaction) above the option that is judged to be best (leaving the sense of 'best' to encompass a variety of normative concerns).

The third proposal for the location of *akrasia* in an appropriate model is in option initiation. They term this "last-ditch *akrasia*". What Kalis et al. refer to as "last-ditch *akrasia"* is quite different from what Pears, who uses the same term, deals with at length (Pears). For Kalis et al. last-ditch *akrasia* is taken to involve neurophysical disorders like apathetic motor inertia, in which the afflicted wishes to exert motor control to no effect, or motor impulsivity (e.g. alien hand syndrome), in which the afflicted experience bodily movements over which they have no volition. There are good reasons to regard behavior resultant from most psychiatric disorders as something the agent "cannot help," i.e. as unfree action. Both in society and in courts of law, we regard those acting under some form of psychiatric disorder as of diminished culpability. The pathologies that Kalis et al. identify as candidates for *akrasia* generate action that seems to me sufficiently *un*intentional to qualify as "acting intentionally against one's better judgment". That Kalis et al. primarily try to identify *akrasia* in terms of one sort or another of psychiatric disorder is a pervasive concern of mine throughout the article, and this difficulty is at one point anticipated by the authors:

> "…[I]n the philosophical literature, *akrasia* is often contrasted with symptoms such as addiction and compulsion. Addiction and compulsion are then presented as…unfree actions or actions caused by irresistible desires. The philosophical understanding of compulsion and addiction differs greatly from the way these

> phenomena are understood in psychiatry and clinical psychology. An important difference is that in psychiatry, the concept of freedom does not play a dominant role in demarcating diagnostic categories; psychiatrists do not generally assume that psychopathological behavior is unfree." (47)

It occurs to me that it may be more appropriate to rephrase the last clause as 'psychiatrists generally take no explicit position at all with regard to philosophical theories of freedom.' There is a great deal of room for dispute about what kinds of actions (if any) are free, and I shall not at this point offer a comprehensive analysis. In any event, historical and contemporary discussion of *akrasia* has considered *akrasia* a problem, but not some kind of psychiatric disorder.[32] For example, consider a person, call him Durst, who is arrested for shoplifting. Assume further that it were found out that he had some kind of psychopathology that either prevented the option of *not* shoplifting from even occurring to him, or that caused involuntary motor impulses that had the effect of causing Durst to grab an item on his way out of a shop in which he decided not to make a purchase. In both of these cases, Durst would not be held responsible owing to his lack of ability to act otherwise. In neither of these cases would Durst be reasonably accused of *akrasia*, in the first place because if he is incapable of entertaining a non-shoplifting option, he cannot have been said to have judged that shoplifting was the inferior course, and yet chose it. In the second place, he cannot be said to have acted intentionally against his better judgment if he did not act intentionally at all.

I think these concerns constitute an adequate reason to reject the first and the third of the possibilities that Kalis et al. propose for the appropriate place of *akrasia* in a psychological model. *Akrasia* in the option selection phase, the second proposal, is much more promising as a starting point for identifying philosophical *akrasia* in psychological study and terminology. Kalis et al. do not identify any specific mechanism for explaining *akrasia*, but instead propose

[32] Even Aristotle describes the *akrates* as being *as if* mad.

that one of a variety of situations that involve different brain structures being active for one decision could be the culprit in *akrasia*. One example is the psychological propensity of humans to value smaller, immediate rewards over larger but delayed rewards. Kalis et al. cite a study (McClure, Laibson and Loewenstein) that indicates that fundamentally different brain structures are employed for each evaluation.

I think that Kalis et al. are on the right track with this proposal, and the account of *akrasia* that I will present is not incompatible with their suggestion that *akrasia* is an artifact of multiple evaluative processes running concurrently and independently in different parts of the brain. Of course, Kalis et al. are proposing only a sketch of a possible approach to explaining *akrasia*. I wish to offer a more complete account.

In the interest of brevity, I want to conclude discussion of the Kalis et al. paper with the following remark. I think that their general approach is useful to the collaboration between psychologists and philosophers and ought to be emulated. There is room for a fruitful dialog between philosophers and psychologists concerning the proper model for *akrasia*. It strikes me (and Kalis, et al.) that the sequential model of decision making and action that they propose is not the only and may not be the best way to outline a research project to study *akrasia*, though it is certainly a good start. At minimum, Kalis et al. have demonstrated that the conceptual analysis that philosophy is known for and the empirical method of testing precise hypotheses that characterizes all of the sciences including psychology and psychiatry work better when working together. The implicit suggestion of Kalis et al. is that some mutual translation of vocabulary and jargon is all that may be required to facilitate such cooperation. This is a good suggestion. I depart from the Kalis et al. analysis in that I think there is a broad category of research in

psychology that maps onto the characteristics of *akrasia* as a concept, namely research in cognitive bias. It is to cognitive bias that I now turn.

Cognitive Bias

In the past several decades, empirical studies of human behavior and decision making have revealed many interesting and surprising results. This research is not limited to psychology or neuroscience; nearly every area of the social sciences is making use of research in cognitive bias for various purposes.[33] By and large, a model of human beings as consistently rational decision makers has had to be substantially discarded. When people face decisions, they often do not pause to subject their perceived options to a comprehensive analysis; they tend to act by habit or employ heuristics to make their decisions. Various heuristics even affect which options are perceived as options at all. I think we can supplant the Kalis et al. sequential model for describing *akrasia* by pointing out that cognitive bias affects both option generation *and* option selection rather than only one *or* the other, and thus is a part of vocabulary in psychology that more neatly maps onto philosophical conceptual analysis of *akrasia*.

The Concept of Cognitive Bias

The conclusion often drawn by a superficial understanding of research in cognitive bias is that human beings are hopelessly irrational and unsystematic decision makers who stubbornly resist any rational, scientific understanding of their decision making. To begin to argue against this unwarranted conclusion, consider how difficult all decisions would be if every time one were faced with a decision, even a trivial one, that one would have to sift through all of the information available, actively separate relevant from irrelevant information, and then perform a rational calculation based on all and only the remaining information, actively arranged and

[33] Advertisers often exploit cognitive biases like the availability bias. Economists rely on research in cognitive bias to understand how people perceive value and act on those perceptions. Public safety officials design warning signs to capture attention (i.e. to initiate a certain sort of explicit cognition).

evaluated…all in real time. Given a deeper and more reflective look at heuristic decision making, it becomes clear that, by and large, our decision heuristics are more beneficial to us than detrimental, though they are unquestionably detrimental at times.[34] Put in the terms of commercial products, heuristic decision-making is both a feature and a bug.

It is in those circumstances in which our heuristics tend to lead to error, that the term "cognitive bias" is applied. Here is one famous example. Tversky and Kahneman, credited as pioneers in investigating cognitive bias, studied what they termed the "representativeness bias" (Judgment Under Uncertainty: Heuristics and Biases). In one study, (Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment) these two researchers noticed that their subjects evaluated coin flip results (H for 'Heads', T for 'Tails') like TTTHHH, HHHHTH, and TTTTTT as less probable than results like HTTHTH. Of course, any specific result of six tosses is just as probable as any other specific result of six tosses of any fair coin. It seems that the subjects confused a number of other concerns with the actual task of evaluating probability. The HHHHTH and TTTTTT results do violate the expected balance of H and T results, but the task is not to evaluate how well the results conform to expected balance of H and T results. So the representativeness bias selected responses that represented deviation from an expected outcome rather than selecting mathematical fact.

Though Tversky and Kahneman focus on coin toss results, it is easier to explain why what they observe is an example of the representativeness bias using poker hands as an example. Hand A is (A♠, K♠, Q♠, J♠, 10♠). Hand B is (3♠, J♣, 2♣, 7♦, 9♥). Hand A and hand B are equally probable, though most people would rate hand A as more improbable. Why is that? The answer seems to be that Hand B is more representative of an average hand in poker, that is, a

[34] See especially (Haselton, Nettle and Andrews) for an account of the benefits of most of our decision heuristics, despite their role in occasional instances of harmful cognitive bias.

hand that would not be likely to be a winning hand, while Hand A, the royal flush, is the strongest possible hand, and thus is strongly representative of an unusual hand in terms of its expected result (an automatic win, barring the astoundingly improbable outcome of another royal flush). Most poker players will play for years without ever seeing a royal flush occur (at least without manipulation of the deck).

Another example of the representativeness bias from Tversky and Kahneman serves to fill out an explanation of representativeness bias. Subjects of the study read a personality description of Linda: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and participated in anti-nuclear demonstrations." Subjects are then asked to compare the probability of two statements about Linda. Statement A is "Linda is a bank teller" and statement B is "Linda is a bank teller and is active in the feminist movement". Despite the fact that a conjunction cannot be any more likely than either of its conjuncts, 80% to 90% of subjects rated statement B as more likely to be true. The explanation for this mistake seems to be that the second conjunct in statement B is more representative of a reasonable conclusion given the evidence than the first conjunct, which is also the entirety of statement A.

It seems that the representativeness bias has its roots in a human ability to quickly separate signal from noise and focus on information relevant to the accomplishment of goals. So when shown poker hands or coin tosses, the information is prioritized and evaluated in terms of what is representative *of success or failure* at poker games or coin tosses, rather than what is probabilistically expected or unexpected. When people read a personality description, facts more relevant or closely connected to the provided description are likely to stand out as signal as opposed to noise. In carefully crafted situations, this particular cognitive bias can be

systematically exploited to entice subjects to commit formal fallacies like the conjunction fallacy, or versions of the gambler's fallacy.

What seems to be happening in these cases is that a one-reason heuristic selects the option that fits a commonly valued normative criterion, but fails with respect to another more highly valued normative criterion in special circumstances. For example, in the case of representativeness bias, a carefully constructed experimental setup exposes a one-reason heuristic that allows rapid identification of common goal-relevant concerns at the expense of ignoring much more cognition-intensive logical or mathematical truths. A generally useful one-reason heuristic, then, can be a factor in getting the "wrong answer".[35]

In a significant sense, cognitive bias research is research into systematic distraction. A case of cognitive bias usefully correlates to a case of *akrasia* when a normally useful heuristic (if we even have any other kind) is a factor in actions that are contrary to the agent's own beliefs about what is better or best. Introducing the notion of cognitive bias has several important contributions to make toward our understanding of *akrasia*. First, cognitive bias is a reasonably well-explored area of empirical research, and so if cognitive bias as explored by empirical psychology is a significant feature of what philosophers call *akrasia*, then those who study *akrasia* can make productive use of this wealth of research and insight. Second, if the account of cognitive bias as a key factor in *akrasia* is successful, then any story about the psychological unintelligibility of *akrasia* must be swept aside, along with any story of the logical impossibility of *akrasia* (if something is actual, then it's possible).

Akrasia and Attentional Bias

[35] Of course, this "wrong" answer is the right answer, but on a less valued normative criterion. Still, we should not wish to be without such effective cognitive architecture. "…[L]ogic, mathematics, probability theory…are computationally weak: incapable of solving the natural adaptive problems our ancestors had to solve reliably in order to reproduce." (Cosmides and Tooby 329)

Some philosophical work concerning *akrasia* that is similar to what I suggest has been done by Christine Tappolet. A subset of cognitive bias is attentional bias. Attentional biases result from heuristics that have as their purpose to direct our attention to important things in our environments, both external and internal (mental). Sometimes these heuristics are misapplied and instead direct our attentions to unimportant features of our environments. For example, when we pay attention to what someone else is doing, we often look at what they are looking at. This is very often the proper place to look, but this is an attentional bias that magicians have long used to misdirect an audience's attention.

Tappolet approaches analysis of *akrasia* as attentional bias in the limited case of *akrasia* that involves emotions. In her paper, "Emotions and the Intelligibility of *Akratic* Action", Tappolet proposes a distinction between "hot" and "cool" *akrasia*. Hot *akrasia* is characterized by the presence of strong emotional states that serve to focus the attention of the person who is deliberating onto certain courses of action and opinions. Cool *akrasia* is absent any significant emotional influence. Tappolet looks at emotions in a way highly informed by recent psychology. Rather than regarding emotions as blind forces at war with reason, which has been one traditional way of looking at emotion's role in *akrasia*,[36] Tappolet regards emotions in the way that has become accepted in recent years. Many emotional states can be understood as perceptions of value. For example, fear as an emotional state often serves to identify and direct one's attention to danger in one's surroundings. One is sometimes afraid when one has no need to be afraid, and judging that the object of the fear is not really something to be afraid of does not always assuage the fear. In the cases in which the fear is not assuaged, one may be less disposed

[36] Most notably in Aquinas, Summa Theologica, Part II, Q. 77, Article 2. See also (Davidson, How is Weakness of the Will Possible?) for discussion of this point. This might also be the general sort of thing that Hume has in mind in denying that there can be any real conflict between reason and the passions (See especially the Treatise of Human Nature 2.3.3)

to notice other features of one's environment. In this way, deliberating in the presence of strong emotions is described as an attentional bias. Tappolet puts it this way:

> "[C]ases of *akrasia* caused by emotions involve a conflict between a value perception and an evaluative judgment that can be compared to perceptual illusions such as the Muller-Lyer illusion, in which one sees the lines as being of a different length even though one judges or knows that they are of the same length. This is particularly easy to see in cases in which one judges that one's emotion is not appropriate, such as when I fear something while I judge that there is no danger and nonetheless act on my fear" (111)

Tappolet is convinced that this approach makes hot *akrasia* intelligible, but leaves cool *akrasia* (deliberation not accompanied by strong emotion) with no such factor that makes it intelligible. Since Tappolet is concerned only with the intelligibility of hot *akrasia*, her account ends there, claiming that the intelligibility of cool *akrasia* is the greater puzzle. I agree that strong emotions (as attention-directing heuristics) account for a cognitive or attentional bias that could play a significant role in explanations for some *akratic* actions.

For cases of cool *akrasia*, I think a similar story can apply, though the bias involved is one that affects cognition without any appeal to strong emotions. The cognitive bias account of *akrasia* would collapse Tappolet's distinction between hot and cool *akrasia* from a distinction between a clearly intelligible form of *akrasia* and a puzzle to a distinction between *akrasia* involving one form of cognitive or attentional bias and another. As a result, I will at this point cease to apply the distinction.

Consider as only one example of the kind of thing that makes *akrasia* intelligible a bias of temporal proximity. In general, people have good reasons to prefer benefits to themselves that

are temporally nearby rather than far away (i.e. a bird in the hand is worth two in the bush). As a result, people often pursue strategies based on this preference without reflecting on the preference (availability is the "one reason" for this one-reason heuristic). However, this generally effective one-reason heuristic can sometimes lead us astray from what we know to be the best course of action.

For example, consider Buck, who has some mild tooth problems. Say that Buck judges that it would be best for him to see a dentist, though he knows the experience will be unpleasant. Many times the thought occurs to Buck that it would be best for him to make a dental appointment (say even that the telephone and the dentist's phone number are near to hand) and still he does not. When finally after many *akratic* episodes, Buck makes a dentist's appointment, he fails to keep the appointment, though he remembers it and judges that it would be better for him to keep his appointment than to miss it. Buck then follows this *akratic* action with *akratic* episodes similar to the first ones when judging that it would be best to reschedule his missed appointment.

Buck is an example of one for whom *akrasia* is more frequent, and not an especially isolated incident. His action is intelligible in the sense that we know why Buck should not want to go to the dentist, but certainly Buck *could have* made and kept an appointment promptly. We have no reason to doubt that he sincerely judged a dental visit as better than no dental visit. We also blame Buck for his failure, and look upon this defiance of his better judgment as irrational. At every turn we might imagine Buck believing that he will make and keep a dentist's appointment before his dental problems escalate, and the reason that he believes this is because he judges that such an action would be best. His desire to avoid pain is served both by avoiding the appointment and making the appointment, so this is not a case of some desire overwhelming

reason, as *akrasia* is often portrayed. What is going on in Buck's case is an example of cognitive bias. A generally useful preference for immediate and certain value over longer term and less certain value (or in this case an aversion to the immediate and certain disvalue encountered in the dentist's office) is a factor contributing to Buck's actions against his own judgment of what is better. Temporal proximity in this case is acting as a one-reason heuristic.

Say that Buck (after perhaps someone's urging him, or perhaps not) says to himself something to the effect of "I knew then, and I know now, that I need a dentist and that it will be unpleasant, but if I continue to wait, it will become far more unpleasant. I don't feel that bad yet, but I really must see the dentist." In doing this, he is focusing his attention against the one-reason heuristic that prefers temporally near to temporally far benefits. Perhaps he writes down a note or sets a reminder to the same effect. Then Buck finally reschedules and keeps his appointment, because his better judgment (i.e. the criterion that is more important to Buck than temporal nearness) is finally substituted for the previous one-reason heuristic. Should any aspect of this behavior baffle anyone? Temporal bias may be something that some have more trouble countering on occasion than others, but it appears that all of the details of Buck's case are intelligible and can be given an empirically informed explanation.

Global Workspace

To complete the work of giving *akrasia* a suitably empirical characterization without losing the concept that has been traditionally explored in philosophy, cognitive bias itself needs something of a deeper explanation. Explaining *akrasia* in terms of cognitive bias leaves open the question, "Why do we have cognitive biases?" A deeper explanation for cognitive bias makes *akrasia* less a thing that, just as a matter of fact, happens to people, but a thing that practically

must beset us at times. Again, my goal here is to establish *akrasia* as an intelligible part of normally functioning human psychology—an undesirable part, to be sure, but normal.

It is obvious that human beings are incapable of taking explicit notice of everything in their surrounding environment. We exist in such an information-rich setting that our greatest challenge is usually in separating the useful information from everything else. Our abilities to do so are not perfect; we cannot notice everything of importance to us in the world around us. Failure to notice certain features of our environment can often have dire consequences for us, but even so, it is a necessary consequence of our ability to focus on some information to the exclusion of other information that sometimes the more important information is missed. Nobody disputes this.

Similarly, our internal environment, accessible via introspection, is more information-rich than our ability to manage it all. We do many things intentionally without being fully aware of them at the time that we do them. For example, we walk and talk and manipulate objects without being aware of moving each muscle and putting down each foot and transferring weight and we notice where we are going without noticing that we notice where we are going. We often have feelings without knowing why we have them, or what they are even directed toward (ask any therapist). It is obvious that our awareness is limited and that our ability to focus on certain internal states to the exclusion of others carries with it the necessary consequence that sometimes we will focus on something that is not as important to us as what is excluded.

When researchers in cognitive science and AI research need to model and simulate human cognition, they increasingly rely on one or another development of Global Workspace Theory. Bernard Baars is credited as the pioneer of GWT, and summarizes it this way:

"The idea that consciousness has an integrative function has a long history. The

> late Francisco Varela and colleagues called it the "brainweb" (2002).[37] Global Workspace theory suggests a fleeting memory capacity that enables access between brain functions that are otherwise separate. This makes sense in a brain that is a brainweb, viewed as a massive parallel distributed system of highly specialized processors. In such a system coordination and control may take place by way of a central information exchange, allowing some specialized processors -- such as sensory systems in the brain -- to distribute information to the system as a whole. This solution works in large-scale computer architectures, which show typical "limited capacity" behavior when information flows by way of a global workspace. A sizable body of evidence suggests that consciousness is the primary agent of such a global access function in humans and other mammals (Baars, 1983, 1988, 1997, 1998)[38]." (Baars)

GWT is generally well regarded not only in cognitive science research, but also in AI research.[39] The central insight behind GWT is that attention (a workspace shared by a multitude of cognitive apparati) is limited, and necessarily so given the amount of information that we encounter and must deal with in some reasonably effective way. Much of this information can be dealt with without ever entering the global workspace, and when the global workspace imports some new information, some old information is crowded out as a necessary consequence of the limited size of the global workspace.

Cognitive bias then, and *akrasia* by extension, has a deeper explanation in terms of the finite capacity that human beings have to pay attention to information both from the external

[37] Baars Cites: Varela, F., J.-P. Lachaux, E. Rodriguez, J. Martinerie, The brainweb: Phase synchronization and large-scale integration. *Nature Reviews - Neuroscience*, 2, April 2001, p. 237.
[38] Baars Cites: (1983): Conscious contents provide the nervous system with coherent, global information." In R.J. Davidson, G.E. Schwartz & D. Shapiro (Eds.), *Consciousness & Self- regulation*. p. 41 NY: Plenum Press. (1988): A cognitive theory of consciousness. New York: Cambridge University Press. (1997): *In the theater of consciousness: The workspace of the mind*. New York: Oxford University Press. (2002): The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Science* Vol. 6 (1), p. 47-52.
[39] For an impressive though not exhaustive list of developments utilizing GWT, see (Wallach and Allen) Chapter 11.

world and to their own internal states. Various cognitive abilities to triage information quickly will on occasion exclude information from our attention that we would prefer not to be excluded from our attention, but in general we would not wish to be without such cognitive abilities.[40] I contend that *akrasia* is not simply a fluke of human psychology, but quite the opposite; vulnerability to *akrasia* is a necessary side-effect of the ability to quickly separate important information—even about our own internal states—from everything else. The case of Aurelius arguing his way out of bed that begins the first chapter is an excellent example of someone attempting to force a dominant consideration into the global workspace, to the exclusion of competing considerations. Methods of self control like the one Aurelius exemplifies are more extensively addressed in Chapter 4.

Some philosophical treatments of *akrasia* are more compatible with the above insights than others, and that compatibility can be a criterion for whether those philosophical treatments of *akrasia* are good treatments of *akrasia* (insofar as my account is the right account, of course). I submit that the central tenets of Global Workspace Theory (GWT) provide an empirically informed account of having but not attending to knowledge that can serve to fill out a generally Aristotelian position with regard to an account of *akrasia* that also happens to be compatible with Davidson's notion of prima facie judgments.

The empirically informed account of *akrasia*

Now that I have presented the requisite background conceptual apparatus, I can present an explanation for the general case of *akrasia*. *Akrasia* occurs when a one-reason heuristic plays a prominent causal role in forming an action or intention, and when the one reason is judged by

[40] For demonstrations of formal simulations involving one-reason heuristics performing as well or better than more complex algorithms that make use of all available information see (Gigerenzer and Goldstein, Reasoning the Fast and Frugal Way: Models of Bounded Rationality).

the *akrates* to be inferior as a reason for action to some other reason for action that would counsel a different and incompatible action or intention.

Consider as an example, Aristotle's dieter. In this example, the very same object (the sweet) is an object judged to be valuable based on the criterion of its pleasant taste, which is likely to be noticed by a one-reason heuristic commending caloric items to our notice. The form of the judgment could be fairly stated as '*pf* I should eat this, sweetness' where a one-reason heuristic supplies one and only one reason supporting the prima facie judgment.

The same action (eating the sweet) is judged to be of negative value on the criterion of its healthfulness to a dieter already operating with a caloric surplus. Note that this does not simply set up a competing prima facie judgment of the form '*pf* I should not eat this, unhealthfulness'. In that case, *akrasia* would be no more explicable than in any other account of *akrasia* in which some force simply overpowers another. Rather, the dieter (again by virtue of being a dieter, and not being merely intemperate) regards the healthfulness of the food as the more important criterion of value when the two values (and perhaps more) are consciously compared. What makes one normative judgment qualify as *better* judgment is that the better judgment includes competing considerations and weighs them. So a judgment of the form '*pf* I should not eat this; sweetness, healthfulness' is a metajudgment in that it includes the considerations relevant to the one-reason heuristic, and in judging in favor of an action that is contrary to that commended by the one-reason heuristic the metajudgment holds the opposing criterion as a better basis for judgment, hence, 'better judgment'. When information is in the global workspace, it is attended to. When it is running in a background process it is incorrect to say that the agent does not know the information—they do, but do not attend to their knowledge.[41]

[41] See (Smith) for an account of the modal claim made when someone could have but did not attend to knowledge that they had.

This account of *akrasia* fits both the criteria for identifying *akrasia* as well as the examples of *akrasia* that I presented in Chapter 1. I have just described how a dieter might select a food for its tastefulness while the more reflective dieter who attends to his better judgment would consider the food's healthfulness as the more important trait and follow his better judgment. Likewise, the apostle Paul may often have found himself clinging to his worldly possessions, unmindful of his more important commitment to following Christ in a life of poverty and service. Certainly, science has documented a systematic tendency to value small, temporally nearby benefits to even very large benefits if they are temporally distant (Kirby and Marakovic). In many cases it makes sense not to delay enjoyment, but of course if the later rewards are sufficient enough, it is better to forego immediate enjoyment. So Paul may find himself unmindfully purchasing material comforts for the now when a more reflective Paul would eschew such immediate comforts in favor of an eternal reward in his future. Likewise, the lover in the throes of passion is prone to an attentional bias of the kind described in Tappolet's account and she does not attend to a more level-headed, but unquestionably better judgment to abstain. In all of these cases, a one-reason decisional heuristic plays a prominent role in fulfilling the first criterion of *akrasia*. That is, a one-reason decisional heuristic can, in some cases, select an action that better judgment would counsel against.

Criterion ii of *akrasia* is that the action is irrational. In none of these cases is the agent doing things for which they have no justifying reason at all, but all of them are betraying what they themselves hold to be of greater value in favor of something that they themselves hold to be of lesser value. That their actions are intelligible and amenable to the explanatory machinery of empirical psychology does not remove the sense that each agent has that they have behaved, by their own lights, irrationally. The cognitive and attentional heuristics to which I have appealed

in this account are generally useful, but are rightly classed as irrational when they get the wrong answer.

Despite wishing that they had acted differently, none of our exemplars are out of control of themselves in the sense that would eliminate culpability. The one thing that each would regret most about their behavior is that they could have (and by their lights should have) attended to their better judgment. The dieter could have avoided eating the sweet but did not. Paul could have given his money to the poor but did not. The lover could have abstained but she did not. In each of these cases, the agents are failing to attend to knowledge that they have, and would have acted in accord with their better judgment had they attended to that better judgment. In this way, all three act intentionally, though against their better judgment, fulfilling the third criterion.

Given that a failure to act in accord with better judgment is a form of lack of self-control, and given that a lack of self-control is generally blamed, *akrasia* is generally blameworthy (criterion iv). If it is within someone's ability to act more carefully, less hastily, or to consider some criterion that would change their behavior and they do not, they are deserving of at least some disapprobation.

Finally, nobody has better judgments that they never follow (criterion v). It is not the case that the dieter always breaks the diet, while maintaining in all seriousness and sincerity that they are on a diet. Paul does not sin at every opportunity while judging in favor of righteousness the whole time. The adulterer does not declare to him or herself, "I think adultery is wrong, and I plan to do it every chance I get!" The fact that someone is unable to change their behavior no matter what they think of their behavior, and no matter the considerations they undergo prior to acting, is evidence that the person is acting under compulsion, i.e. not intentionally. They may

be best advised to seek help, and a more extensive sort of help than that which causes the *akrates* to simply attend to their own better judgment.

An Observed Case

To this point I have been operating with some canonical examples of *akrasia* as it has been discussed in the philosophical tradition. To lend further support to the claim that *akrasia* is not only intelligible to contemporary empirical psychology, but in fact empirically observed, I shall now present an empirically documented case of *akrasia*. Of course, the term '*akrasia*' is not used, but I will attempt to show that this empirically observed case is a case of *akrasia*. As such I will describe the ways in which this case fits not only the traditional philosophical notion of *akrasia*, but the empirical explanation of *akrasia* contained so far in this chapter.

As it happens, the case to which I will appeal is a familiar one to many philosophers, and has most prominently been used to challenge the notion of character in ethics. The position that character has been shown by empirical research to be a fundamentally unstable notion is known primarily as situationism. I will not get into a full discussion of all of the various formulations of situationism in general, but instead shall appropriate one case used by one of situationism's most prominent proponents, John Doris[42]. Doris relays a number of often cited psychological experiments intended, when taken as a whole, to argue that character is an unstable notion that is not observed in fact.[43]

The case that I shall primarily utilize is a psychology experiment in which a group of seminarians, assumed to be persons possessing the character trait 'compassion' due to self-report and due to vocation, were told that they were late to give a presentation on Jesus's story of the

[42] For other prominent situationist accounts, see (Harman) and (Vranas).
[43] Most prominent among these are: (Isen and Levin) concerning generosity differences between subjects who had and who had not recently found a dime; (Mathews and Cannon) concerning the willingness of persons to act benevolently depending on ambient noise levels; (Darley and Batson) concerning the benevolence of people who are in a hurry.

Good Samaritan (Darley and Batson).[44] The story of the good Samaritan being a notorious example of altruistic compassion reflects a wryly humorous streak to Darley and Batson, the framers of the experiment. These provisionally compassionate seminary students who were in a hurry to give a talk were presented with a person suffering the symptoms of a heart attack (an actor hired by the psychologists) en route. The seminarians generally failed to respond to the heart attack "victim" on their path to the "presentation".[45]

Doris interprets this result as one among many that indicate that our notion of compassion as a character trait is an empty notion. However, in order to construct this study at all, one requires a reasonably robust notion of compassion as a character trait, including what sorts of things count as compassionate behavior and which situations involve clearly identifiable compassionate or non-compassionate outcomes. In this case, Darley and Batson (reasonably) assume that if anything is a compassionate action, it is stopping to assist the heart attack victim, speaking engagement or no speaking engagement. If anything is a betrayal of compassion, it is proceeding blithely onward to give a speech (even a speech extolling compassion) while ignoring someone gravely in need of assistance.

Additionally, the framers of the study even make assumptions concerning the sort of person likely to possess the trait of compassion (assumptions assumed to be shared by the reading audience) that belie a multifaceted, complex, and socially widespread idea of just what compassion is supposed to be.

For my purposes, it will work equally well to focus not on character traits, but on normative judgments. Surely, someone who is compassionate is one who judges that it is better

[44] Subjects of this study were also given less loaded topics to speak on, but there were not significant differences between the groups resultant from which topic they were to be speaking on.
[45] About 60% of those who were not in a hurry responded to the distress, while about 10% of those in a hurry responded.

to assist the heart attack "victim" than to keep a speaking engagement. I should also think that any reasonably moral person, not only a seminarian, would sincerely assent to this normative judgment.

So, to the case of seminarians in a hurry failing to do what is recognized as the compassionate act. The situationist's analysis of this study is as follows: the situation that the subjects find themselves in (being in a hurry) actuates a major difference in their behavior while the character trait that they are assumed to have (compassion) remains constant despite very different patterns of behavior. This is not taken to be an indication that character traits disappear when one is in a hurry, but rather that the situation of being in a hurry has a greater impact on what a person does than the character traits that they are supposed to have. This analysis can be easily redescribed in terms of judgment. While it is certainly the case that hurry has a substantial effect on the actions of the seminarians, we have no reason to assume that they abandon their better judgment. But how do we explain why people who are legitimately committed to helping those in need above their own immediate interests (at least and especially when the needs of others are dire compared with one's own needs) fail to do so?

As conceptual background to clarify my response to this question, it is useful to bring in some of Davidson's insights concerning intentional action. Since we are in part discussing the role of better judgment in causing (or failing to cause) action, Davidson's framework concerning how best to attribute causation in intentional action is informative. In "Actions, Reasons, and Causes" Davidson describes a simple event.

> "I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home. Here I need not have done four things, but only one, of which four descriptions have been given… [R]easons may

rationalize what someone does when it is described in one way and not when it is described in another." (4-5)

Combine this with the statement that "A reason rationalizes an action only if it leads us to see something that the agent saw, or thought he saw, in his action" (3)[46] and we have a position stating that in the doing of an action, an indefinite number of descriptions may apply to the action. In this case we are concerned with the *cause* of the action. And in explaining intentional action, it is by appealing to the agent's reasons (in the Davidsonian sense of 'reasons' specified above) for doing x that we identify the cause of their doing x.[47]

The difference between (intentionally) illuminating the room and (unintentionally) notifying a burglar of one's presence depends on what information one has when flipping the light switch. I wish to re-employ Davidson's reasoning in cases in which a person may have, but not attend to, information that relevantly differentiates intentional from unintentional action. My purpose in this is to respond to an assertion often made by those who are skeptical of the possibility of *akrasia*. The assertion is that, in the terms of the Darley and Batson seminarians, if they were *truly* compassionate, they would have stopped to help. So the fact that they did not is *ipso facto* evidence for a lack of compassion. This is reasonable as far as it goes. Anyone who intentionally ignored someone in dire need in order to make a speech on time would be acting uncompassionately. However, we must separate intentionally hurrying to a speech from intentionally ignoring someone in need, even though they are the same action.

I argue that the best interpretation of the results of the Darley and Batson study is that the case of the seminarians reveals an instance of *akrasia* more clearly than it reveals the instability of compassion. Moreover, I think that treating the data in the study as revealing an instance of

[46] See also a similar point, too lengthy to quote, in (Davidson, Intending 84-85)
[47] I realize that this is a contentious point, but will adopt it here without argument, simply referring the reader to the previously cited papers by Davidson, which I regard as persuasive.

akrasia is quite natural given the account of *akrasia* I have presented above.

Now that I have laid out the conceptual tools that I require to interpret this case as an example of *akrasia*, let us look in more detail at what happens in Darley and Batson's experiment. The seminarians do a number of things in the full course of participating in the experiment. They report that they are compassionate, they agree to speak about the Good Samaritan, they hurry to their presentation, they (some of them) miss an apparent heart attack victim, they speak about the Good Samaritan. The events in question that are simultaneous, as in Davidson's example of flipping the light switch, are 1) hurrying to the presentation and 2) missing an apparent heart attack victim. Notice that the description is as yet devoid of intentional explanation (rationalization). One of the features of Davidson's example that is important to note is that in acting intentionally with respect to one of these two descriptions of an action is not necessary to act intentionally with regard to the other. That is, in intentionally hurrying to the presentation, one need not intentionally miss the apparent heart attack victim. In deciding to walk quickly by someone who you think may be having a heart attack out of hope that someone else will notice and take action instead, you do not intentionally hurry to wherever else you are going, though you act intentionally and you do hurry to wherever else you're going.

When inserting intentional explanation from a third-person perspective one must take into account that descriptions of actions are ambiguous with respect to intentional explanation. It is equally plausible given an observer's evidence that: 1) the seminarians are intentionally ignoring the apparent heart attack victim, and 2) that the seminarians are incidentally ignoring the heart attack victim while intentionally doing something else (hurrying to their appointment). The facts that the seminarians do hurry to their presentation and that they do miss the apparent heart attack victim is silent as to the contents of their better judgments. The third person interpreter of this

experiment, in lacking certain knowledge of the intentions of the subjects of the study, can only say that the subjects in a hurry *more commonly* missed the heart attack "victim". We must appeal to the concept of *akrasia* to tell us the difference between those in a hurry who did and those in a hurry who did not stop to assist the "victim".

Causal reasoning is subject to the criteria of adequacy for explanations. Any explanation that is consistent with the data, and can explain not only why subjects who were not in a hurry assisted the "victim" more often than those who were in a hurry, but that can also specify the difference between those in a hurry who helped and those who did not, is a more powerful explanation than any explanation that can account for only some of those differences.

Among those in a hurry who did not assist the "victim", there are two plausible intentional explanations. At this time it is instructive to consider the merits of each. The first explanation is that those subjects noticed the victim and intentionally decided to ignore the victim and proceed to their presentation. This outcome shows the subject to be uncompassionate. To put it in terms of judgments, consider the following reasonable assumptions about the commitments of the compassionate person: a) the compassionate person judges that it is better to assist someone in a great need than to see to one's own lesser needs and b) making a presentation on time is a lesser need than getting swift aid when having a heart attack. In the case where this intentional explanation is at play, the difference between those who stop and those who do not is clearly the presence or absence of the judgment in favor of assisting the "victim".

How plausible is this explanation? I do not think that common moral experience contains so many persons who so callously prefer their own minor needs to the major needs of others in emergencies. There are undoubtedly some, but most persons would sincerely assent to assisting a heart attack victim even if late for an engagement. We would expect a majority of person who

hold this judgment then to act on it, and so something else must be going on in order to explain the results in a way that squares with other aspects of common experience.

The second plausible intentional explanation is that the hurrying subjects who did not stop to assist the "victim" did not notice the victim, but still intentionally hurried to their presentations. Some of the seminarians in fact noticed the victim, and stopped to render aid. The better explanation of the results of the experiment is that the experiment is crafted so as to induce *akrasia*.

Let us match up the criteria for *akrasia* outlined in Chapter 1 to the case of the compassionate subject (who is compassionate because she makes the judgments stipulated a few paragraphs above as a) and b)). The criteria for an *akratic* act are: (i) *Akrasia* is action against a belief about what is better or best to do, (ii) *Akrasia* is irrational, (iii) *Akrasia* is voluntary, (iv) *Akrasia* is blameworthy, (v) *Akrasia* is episodic.

The compassionate person is reasonably stipulated to make certain normative judgments, and in acting against them, fulfills condition (i), though they would insist even after acting against those normative judgments that they held those judgments (ii). Certainly they were not compelled to ignore the "victims" and intentionally hurried to their presentations (iii). When the seminarians report themselves as compassionate, the case provided does not give us evidence to conclude that they are not. Granted, the failure of the seminarians to respond to the apparent heart attack victim is blameworthy (iv), and also requires some explanation to preserve their claim that they are compassionate, and were compassionate the entire time. Further, it certainly sounds implausible that someone who was compassionate would *always* miss a chance to act compassionately (v) in circumstances like this experiment (it is worth noting that some of these compassionate people did notice the "victim" and both behaved compassionately and displayed

enkrateia, the opposite of *akrasia*).

I believe that I have provided such an explanation for the failure of these compassionate persons to act compassionately in my account of *akrasia* as involving cognitive or attentional bias. To briefly apply the outline of my account earlier in this chapter, the ability to narrow one's focus to accomplish a single, urgent task is an important cognitive ability, but can be misapplied. It is certainly not the case that the seminarians were *unable* to notice the apparent victim, as some experimental subjects in fact did notice the apparent victim. In Chapter 4, I shall be contending that the cognitive bias account of *akrasia* implicates a specifically cognitive weakness, but a vincible weakness.

What the case of the seminarians demonstrates is that when one is in a hurry, one's cognitive focus is narrowed toward identifying features in our environments that assist us in the goal of haste at the expense of information that does not. This means that one's attention while in a hurry would be biased toward, say, a quick way through a crowd as opposed to any features of the crowd that are not going to be germane to the goal of getting to where one is going on time. Specific experimental situations, narrowly tailored to take advantage of the cognitive biases of persons, generally influence those persons' behavior in a regular and predictable way.[48] What the Darley and Batson experiment succeeds in demonstrating is not so much that compassion is inherently irregular and unpredictable, but that our own cognitive and attentional tendencies are regular and predictable. It is far more unlikely for the better judgment favoring great benevolence over trivial prudence to be in the global workspace when in a hurry because in such situations the global workspace is narrowed to items selected by a one-reason heuristic that considers only what

[48] In fact, the original Darley and Batson study found that those told that they were late for something *very important* noticed the victim less often than those who were told that they were late for something not very important, indicating that strengthening the bias makes more frequent the actions associated with the bias.

will aid the subject in their haste.

Part and parcel to the classic virtue ethics approach is the idea that people have differing character traits in differing degrees from individual to individual. I submit that different individuals have differing cognitive abilities (to a point), but that the abilities to pay closer attention to more important things more often and to pay less attention to less important things are abilities that can be habituated and practiced, and can vary from individual to individual. In this respect, *enkrateia*, defined as the capacity to avoid *akrasia*, is a character trait much like any other, and matches up well with existing accounts of virtues and also with a common-sense notion of an enkratic person as more virtuous than an *akratic* person, and *akrasia* as distinct from, and better than, intemperance.

The subject who notices and decides not to help the "victim" is intemperate, but subjectively rational, preferring their own lesser good to the much greater good of another. The subject who fails to attend to their environment sufficiently to act on behalf of the commitments of their own character is *akratic*, but at least has it in their character to judge what ought to have been done. The *akrates*' defense of "Of course I'm compassionate, but I was in a hurry, I didn't notice the victim." makes their action intelligible, yet still not excusable, even if better than the intemperate person's action.

Conclusion

I take myself in this chapter to have provided an empirically informed account of *akrasia* that succeeds in demonstrating that *akrasia* is both actual and intelligible. I have responded to prominent claims to the contrary, and found that existing responses to those claims were desirable insofar as they were compatible with the more detailed, empirical account of *akrasia*. The empirically informed account usefully completes Davidson's defense of the logical

possibility of *akrasia*, and likewise for Aristotle's defense of its psychological intelligibility. Most importantly, the account of *akrasia* that I have provided answers these objections without changing anything about the traditional philosophical concept of *akrasia*.

If I have successfully argued that *akrasia* is actual (and thus possible) and intelligible, then I may now turn to examining the implications that *akrasia* generates with respect to some positions in meta-ethics and moral psychology.

CHAPTER 3: ACCOUNTING FOR *AKRASIA* IN VIEWS OF NORMATIVE MOTIVATION

In Chapter 1, I provided a set of features common to traditional examples of *akrasia*. Among those features is irrationality. Given that 'rationality' is an ambiguous term, it is my goal in this chapter to identify precisely what sense of rationality is impugned by *akrasia* in the description of *akrasia* as supplied in the previous chapter. This discussion will occur within the context of specifying the nature of the connection between normative judgment and motivation.

Central to *akrasia* is the role that normative judgments play in motivation. There are two basic positions concerning the relationship between normative judgments and motivation. One such view is the view that normative judgments have a necessary connection with motivation. This view is called, for whatever reason, normative judgment internalism. Its traditional opponent is the position that normative judgments have a merely contingent connection with motivation (Bratman 159). This view is called normative judgment externalism. I shall argue in this chapter for a modified version of normative judgment internalism.

Strong Normative Judgment Internalism

When we ask ourselves a normative question about what it is best for us to do, we regard that as equivalent to asking ourselves a practical question about what to do. Philosophical views defending this necessary connection between our judgments about what is best (our normative judgments) and our plans for action are known as normative judgment internalist views. These views most straightforwardly explain why questions of what it is best to do seem like questions of what to do. *Akrasia* provides a counterexample to a normative judgment internalist view. If there is a necessary connection between judgment about what is better or best and motivation to do what is judged, then *akrasia* as described previously cannot happen. Given this conflict, a defender of the possibility of *akrasia* and a proponent of normative judgment internalism must

either give one up entirely, or reformulate one or the other or both to accommodate what is plausible about each view.

I believe that no reformulation of *akrasia* is necessary, as the description of *akrasia* in the philosophical discussions that I have advanced and operated with thus far is, I think, the right one. I also believe that the evidence advanced previously in favor of the actuality of *akrasia* is sufficient to deny the desirability of doing away with *akrasia* on account of its conflict with normative judgment internalism.

Normative judgment internalism, though, contains a set of very plausible and widely accepted intuitions concerning the relationship of our normative judgments to our motivations, and I would like to preserve these to whatever extent I can. A short list of statements in favor of normative judgment internalism begins with Richard Price, writing "When we are conscious that an action is *fit* to be done, or that it *ought* to be done, it is not conceivable that we can remain *uninfluenced* or want a *motive* to action." (A Review Of The Principal Questions In Morals) Gilbert Harman states, "To think you ought to do something is to be motivated to do it. To think that it would be wrong to do something is to be motivated not to do it." (The Nature of Morality 33) C.L. Stevenson puts it "'Goodness' must have, so to speak, a magnetism. A person who recognizes X to be 'good' must *ipso facto* acquire a stronger tendency to act in its favor that he otherwise would have had." (The Emotive Meaning of Ethical Terms 13)[49] Simon Blackburn adds, "It seems to be a conceptual truth that to regard something as good is to feel a pull towards promoting or choosing it, or towards wanting other people to feel the pull towards promoting or

[49] Notice that this is the most weakly presented of the cited claims, but one that still considers judgment about the good (normative judgment) to be necessarily connected to motivation. This is still strong normative judgment internalism in the sense that a judgment about the good provides motivational force, but judgment about a greater good will provide greater motivational force.

choosing it." (Blackburn 23) [50] In Wittgenstein's "Lecture on Ethics", his only public lecture, he appears to be certain of only one thing in ethics, and that is the truth of something like normative judgment internalism, writing:

> "Now let us see what we could possibly mean by the expression, 'the absolutely right road.' I think it would be the road which everybody on seeing it would, with logical necessity, have to go, or be ashamed for not going. And similarly the absolute good, if it is a describable state of affairs, would be one which everybody, independent of his tastes and inclinations, would necessarily bring about or feel guilty for not bringing about."

Each of these phrasings of the thesis indicates something that is plausible about normative judgment internalism.

Price and Harman focus on the intuitive function of a judgment. If judgment were not sufficient to motivate action, one would wonder why we bothered making judgments at all. R. M. Hare argued that assenting to a sentence like 'It is moral for me to do x' entails assenting to the imperative to myself, 'do x'. For Hare, this is just a matter of the meanings of the relevant terms. He denied any possibility of *akrasia* on the strength of this kind of normative judgment internalism. Hare argued instead that any purported case of *akrasia* is either an exercise of "judgment" (note the quotes, indicating that the judgment in question is not *really* judgment) as if one were to recognize some social norm, but not to treat it as if it applied to her, or else that the action was compelled, or un-free, in some way.[51]

Blackburn, Stevenson, and Wittgenstein point to the magnetism of the good. When someone admits something is good, it seems that they *ipso facto* have motivation to pursue it. Socrates (as relayed by Plato in the Protagoras dialog) famously denied the possibility of *akrasia*

[50] I thank David Brink for a very helpful footnote here, (Moral Motivation 8)
[51] See (Hare, The Language of Morals Ch. 11)

on the ground that the good is so magnetic that anyone who knew what the good thing to do was could not act contrary to it.

There are many ways in which someone who defended normative judgment internalism could explain their view. One could have a view that a normative judgment by itself provides motivational force.[52] The necessary connection between normative judgments and motivation is in that case is explained by maintaining that motivational force is an essential part of normative judgments. One could also say that normative judgments express desires,[53] or else that normative judgments just are desires.[54] In the former of these two positions, the connection between normative judgments and motivational force is dependent on desires essentially having motivational force and there being a necessary causal relationship between having a particular desire and forming its corresponding normative judgment. For example, one could not judge that it is best to be a vegetarian unless they possessed an appropriate desire to be a vegetarian. The latter of these positions explains the necessary connection between normative judgments and motivational force by positing an identity relation between desires as the bearers of motivational force and normative judgments.

Strong normative judgment internalism provides a kind of sufficiency in terms of explanation of why some actions are taken or intentions formed. Imagine asking someone why she decided to stay home sick from work rather than go in on some particular occasion. If her response is that she judged that it would be better for her to stay in, we could view this as a sufficient explanation of her action in that the explanation is not enthymematic. We may make inquiries concerning the sense in which she judged it best to stay home, but all the while we recognize that whatever sense of 'best' is intended, it is explanatorily sufficient to account for

[52] A position like this is advocated by (Davidson, Intending), (Nagel), and (McDowell)
[53] This is the suggestion that many emotivists make.
[54] This viewpoint is open, but not promising, see (Lewis, Desire as Belief) and (Lewis, Desire as Belief II)

her motivation to stay home. If normative judgment internalism, or something very like it, were not true, then we would not be able to regard such explanations as complete explanations. Rather, we would have to supply the normative judgment *and* the connection between the normative judgment and motivation.

Even some notable skeptics of moral judgment internalism like Bernard Williams and David Brink reject *moral* judgment internalism while accepting a form of normative judgment internalism. In Williams' case, he argues that morality is only one of a number of normative concerns that might motivate a person, but only if that person himself held morality to be more important than competing personal projects (Internal and External Reasons). Brink's "principled amoralist" regards morality as of secondary or no importance when held against other normative criteria (Moral Motivation). What both of these positions have in common is the idea that 'is best' and 'is moral' are not necessarily synonymous. Both positions deny that someone must be necessarily motivated by their own moral judgment on the strength that they may be instead motivated by a *better* but not distinctly *moral* judgment.

I regard these considerations as sufficient motivation to preserve normative judgment internalism to the greatest extent compatible with the actuality of *akrasia*. If I have properly described the options open to one who wishes to accept both a version of *akrasia* and also a version of normative judgment internalism, then I am left with only one choice: to modify normative judgment internalism to account for *akrasia* as described previously.

Any modification of normative judgment internalism to allow for *akrasia* will be successful only in the case that it (1) is itself coherent and free from significant conceptual troubles, (2) allows *akrasia*, (3) preserves what is plausible about normative judgment

internalism, and (4) is defensible against a view that is not normative judgment internalism that already allows *akrasia*, a view called normative judgment externalism.

There are many potential varieties of normative judgment externalism because the position is defined negatively. It holds only that normative judgment internalism is false. The most common version of normative judgment externalism, and the one that I have in mind when I mention the position, is known as Humean externalism (Stroud, Weakness of Will and Practical Judgment). The "Humean" part of Humean externalism is due to adherence to some form of Humean psychology, which holds that beliefs (cognitive states) do not motivate action, but that only desires (conative states) motivate actions. The Humean externalist's take on *akrasia* is as follows. A person may *judge* (where a judgment is a cognitive state) that x is better than y, but if they have no antecedent *desire* to do x, they will not be motivated to do x, whatever their (cognitive) judgment about the merits of x. I shall return to this position after presenting my preferred version of weak normative judgment internalism.

Modifying Normative Judgment Internalism

My proposal for modifying normative judgment internalism will be to add a qualifying phrase weakening the strong thesis of normative judgment internalism that I have so far been describing. The strong normative judgment internalist thesis is:

Necessarily, if an agent judges it is best to φ, then the agent is motivated to φ.

The account of *akrasia* described in Chapter 1 provides a counterexample to the strong normative judgment internalist thesis. If the strong normative judgment internalist thesis is that it is necessary that if an agent judges it is best to φ, then the agent is motivated to φ, then any possibility of judging y better than x and doing x rather than y would deny the strong normative

judgment internalist thesis. To deny the strong normative judgment internalist thesis is to either opt for a weaker sense of 'necessary' in the thesis of normative judgment internalism or to adopt normative judgment externalism.

In contemporary philosophical writing, the way that internalists tend to weaken their positions to accommodate *akrasia* is to add to the strong normative judgment internalist thesis a term that in some way or other qualifies 'necessary'. Below, three examples of weak normative judgment internalist theses are supplied as paradigm cases of weakening internalism due to *akrasia*, specific weakenings underlined.

Wedgwood: "Necessarily, <u>if one is rational</u>, then, if one judges 'I ought to φ', one also intends to φ." (The Nature of Normativity 25)

Smith: "If someone judges that it is right that she φs then, <u>ceteris paribus</u>, she φs." (The Moral Problem 12)

Stroud: "..an agent's having reached a practical judgment in favour of φing is <u>normally</u> sufficient explanation of her intentionally φing, or intending to φ." (Weakness of Will and Practical Judgment 122)

There are three different weakenings of internalism present here, but given the surrounding text in each original work, each is intended to have roughly the same purpose, and that is to make room for (at least) *akrasia*.

Wedgwood provides a great deal of specificity with regard to the rationality he has in mind, and so it is worthwhile to start with his account of a weakening of internalism. Talk of reasons, reasonableness, rationality, and reason has historically been fraught with a myriad of distinctions. It is vitally important when speaking of the rational or reasonable to disambiguate one sense of the word 'rational' from others that may well be intended.

First off, Wedgwood means something stronger than the trivial sense of rationality (a sense so trivial, no one may even bother defending it[55]) that states simply that an action is rational if it is done for a reason (where a reason can be anything that explains the action, i.e. a cause). Even *akratic* action and compulsive action fit this description of rationality because the compulsive action is caused by the compulsion, and the *akrates* does act for a reason in the sense that their action is voluntary, though the reason for which they act is inferior to their better judgment.

Rather, Wedgwood claims that there are *synchronic* and *diachronic* requirements to rationality. Synchronic requirements state that to be rational, one must avoid combinations of beliefs and intentions that are intuitively in conflict with one another. Diachronic requirements state that to be rational one must follow all of the proper procedures for forming and revising one's beliefs and intentions. Synchronic rationality is the kind Wedgwood intends to be in play in his weakening of internalism.[56] In addition, Wedgwood tracks another distinction at play in this particular use of 'rationality'. He states that one could assess the rationality or irrationality of a thought or action in either its relation to the thinker's other mental states (subjective rationality) or else on the basis of its relations to the external world (objective rationality). *Synchronic* and *diachronic* rationality are meant to be distinct ways that a person can be described as subjectively rational. (The Nature of Normativity 27) For an agent to be completely subjectively rational would be for the agent to realize coherence among his mental states (be synchronically rational), and for the agent to follow all of the proper procedures for forming and

[55] It might even be that this sense of 'rational' is descriptive rather than normative. This might be the sense of rational that people invoke when saying that 'there's got to be a rational explanation for this' where what is rational is merely causal in some intelligible or non-supernatural way.
[56] See the final pages of (Davidson, Incoherence and Irrationality) for a similar view of the kind of irrationality implicated in *akrasia*¸ and Davidson's views concerning what makes synchronic inconsistency so much more puzzling than is diachronic inconsistency.

revising one's beliefs and intentions (be diachronically rational). For an agent to be objectively rational is to appear to choose correctly given an objective perspective.

To make these distinctions clearer, and to better isolate the form of irrationality implicated in the *akrates*, consider a pair of examples. The first is an example adapted from David Brink. If Biff is a light-bulb eater, his belief that light-bulbs are nutritious and his intention to eat that which is nutritious motivate Biff and explain why Biff is eating light bulbs. (Moral Realism and the Foundations of Ethics 39) Biff is synchronically rational so long as he recognizes no beliefs or intentions that contradict the above belief or intention. Biff is surely not diachronically rational, believing that light-bulbs are nutritious because of a failure to correctly interpret the ambiguity in the admonition to eat light meals.[57] In addition to having some degree of failure in his subjective rationality, Biff fails to be objectively rational; he fails to have a justificatory reason for his actions from the perspective of the third person, owing to the obvious disadvantages of being a light-bulb eater.

Compare this example to an example adapted from Bernard Williams (Internal and External Reasons 102). Chuck intends to drink gin, and believes the stuff in the bottle is gin. Assuming no belief or intention contradicts the above, Chuck is synchronically rational. Assume that the bottle is labeled 'gin' and looks much as a standard bottle of gin looks, and we would have no trouble saying that Chuck has followed the proper procedures for forming and revising his beliefs and intentions and that he is thus diachronically rational as well. As it turns out, the stuff in the bottle is, contra Chuck's belief, petrol. Though Chuck is subjectively rational, he has no objective justificatory reason to drink the stuff in the bottle (petrol). Wedgwood uses this

[57] Whatever Biff's reason for belief, it would seem to be a failure of diachronic rationality given the obviousness of the fact that light bulbs are not nutritious. I shall not challenge the ingenuity of philosophers to come up with a situation in which a decision to eat light-bulbs involves no plausible failure of diachronic rationality.

very example to mark the distinction between "choosing rationally" (being subjectively rational) and "choosing correctly" (being objectively rational). (Choosing Rationally and Choosing Correctly 201-202) Notice that neither of these cases could properly be called *akrasia*. Biff never judges against that which he intends to do and subsequently does. Chuck may recognize *after* his first sip of fizzy, lime-flavoured petrol on ice that he has made a mistake, but would not say that he acted against his judgment, but just that one of his beliefs was false in such a way that he should not necessarily have been expected to realize before it was too late. In both cases, even though Biff and Chuck are in some sense irrational, both are still motivated to act in accord with their own sincerely held normative judgments. Neither provides an example of *akrasia* and neither provides a counterexample to the thesis of strong normative judgment internalism.

The irrationality involved in the case of *akrasia* then, for Wedgwood, consists in inconsistency between one's beliefs (so long as judgments are beliefs or at least so long as one cannot make a judgment without believing that one has made a judgment, which are surely reasonable assumptions) and one's intentions. When considering the examples of *akrasia* given in the opening chapter of this work, we notice that the behaviors indicate a difference between what one judges in favor of and what one does. Wedgwood's qualification, 'insofar as one is rational' can have a number of readings, but since Wedgwood is explicitly interested in weakening strong internalism to make room for *akrasia*, we need only read 'rational' in his qualifications as subjective synchronic rationality. Lack of these other forms of rationality do not provide any reason to doubt the strong normative judgment internalist thesis; Biff and Chuck are both irrational in some way, though neither provides a counterexample to normative judgment internalism. As a consequence, there is no reason to include either subjective

diachronic rationality or objective rationality in the qualification of the normative judgment internalist thesis.

I stated earlier that an appropriate weakening of internalism should (1) be itself coherent and free from significant conceptual troubles, (2) allow *akrasia*, (3) preserve what is plausible about normative judgment internalism, and (4) be defensible against normative judgment externalism. Wedgwood's account is to my mind acceptable on each of the first three concerns. It seems free from any obvious conceptual inconsistencies, it straightforwardly allows *akrasia* (one might even say it allows *akrasia* and little if anything else), and it allows the good to be magnetic, and judgments to have a necessary connection to motivation for anyone not suffering from a failure of synchronic rationality.

I have addressed the first three concerns in what has preceded, but should say something briefly about the fourth here. This version of weakened normative judgment internalism is distinct from externalism in the following ways. First, Wedgwood's position regards a failure of the strong thesis as a failure of synchronic rationality. The externalist need not see any synchronic irrationality (or perhaps any irrationality at all) at play in one who simply does not desire to do that which she nonetheless judges to be the better thing to do. One might normatively condemn her as a hypocrite, or level on her some other normatively phrased disapprobation (including some other form of irrationality), but that would not necessarily have to be synchronic irrationality.

Additionally, Wedgwood's version of weak normative judgment internalism preserves a strong (though narrowly defeasible) connection between normative judgment and motivation. The *akrates*, in intentionally acting contrary to her better judgment, experiences an inconsistency between her own beliefs and intentions. This inconsistency, though, is not an unintelligible

inconsistency. The cognitive bias account of *akrasia* detailed in Chapter 2 lays out the sort of causal factors that are responsible for a breakdown of subjective synchronic rationality.

To begin to address the acceptability of Wedgwood's version of weak normative judgment internalism versus the corresponding externalist view, it will be useful to consider a different form of weak normative judgment internalism so as to develop a useful contrast with Wedgwood's view with respect to how well the view weathers an objection from normative judgment externalism.

To provide this contrast, I will now turn to Smith's account of weak normative judgment internalism. Smith is, unlike Stroud or Wedgwood, committed to the generally Humean view that beliefs and desires are distinct entities and that beliefs are in and of themselves devoid of motivational force. Note that this is also true of the most prominent form of normative judgment externalism. The way that the Humean normative judgment externalist accounts for *akrasia* is by making note of the very fact of the separation between beliefs and desires and the motivational monopoly held by desires.

When Smith invokes a *ceteris paribus* qualification of the strong normative judgment internalist thesis, what he means is that irrational behavior sometimes happens, whether it be *akrasia* or else psychological compulsion of some kind.[58] Compared with Wedgwood's narrow appeal to subjective synchronic rationality, Smith's is a very expansive weakening of the internalist thesis, open to the charge of being a weakening that makes the resulting thesis trivial. After all, what is the normative judgment internalist defending when he states "Normative judgments motivate actions (except when they don't)"? To Smith's credit, he does refine his

[58] It strikes me that this is not a strictly correct use of the phrase '*ceteris paribus*', although there may be a common usage of the phrase that takes it to mean simply 'barring strangeness' or 'normally'. Usually a ceteris paribus clause will be inserted into an empirical or scientific law in order to specify that normal conditions are assumed. This is generally a non-objectionable move for inductive principles, it is more objectionable for formal conceptual definitions.

statement of normative judgment internalism, later substituting the phrase 'insofar as one is rational' for the *ceteris paribus* clause. Eventually Smith's statement of normative judgment internalism is: "Agents who believe that they have a normative reason to φ in certain circumstances C rationally should desire to φ in C." (The Moral Problem 148) Unfortunately, Smith does not go on to explain just what kind of rationality he regards as necessary for his weak internalism to hold. What Smith does is analyze judgments about one's reasons for action into judgments about what one would desire if one were fully rational. It is tempting to read this move as an advocacy of the kind of subjective synchronic rationality that Wedgwood advocates, though considering rationality to be coherence of and between beliefs and desires rather than beliefs and intentions.

Reading Smith's move in this way is problematic because incoherence between beliefs and desires cannot account for *akrasia* in the way that incoherence between beliefs and intentions can. Smith relies on a common concept of "direction of fit" to specify the distinction between beliefs and desires. Beliefs are supposed to originate in the world, and changes in the world should lead to appropriate changes in an agent's belief states, while desires originate in the mind, and changes in desiderative states should lead to appropriate changes in the world. This distinction prevents any belief from being equivalent to any desire and vice versa. It is difficult to see how, on this view, any incoherence could arise between beliefs and desires.[59]

A possibility I shall mention just to put aside is that someone could believe that they held a desire that they did not hold or desire to hold a belief that they did not hold, but that is no actual incoherence between a belief and a desire, but rather simple a false belief or an unfulfilled desire. Alternatively, if normative judgments just are beliefs about what is better, then a person

[59] Indeed, throughout The Moral Problem Smith assumes, following Hume, that beliefs and desires are "distinct existences".

could have a belief that it would be better to φ without ever having a desire to φ. This person does not thereby have an inconsistency in their psychology. To say that such a state amounts to a failure of subjective[60] rationality would be to beg the question against the common externalist position because it just defines rationality as an agent having whatever desires they must have for their judgments to motivate them. For many externalists, and Humeans in general, there is nothing necessarily subjectively[61] irrational about having a belief about what is better and not having a desire to pursue that which is better. Smith relies on some beliefs having rational force over some desires, but without accounting for the source of that rational force. It is thus unclear how Smith's understanding of rationality as consistency between beliefs and desires doesn't simply beg the question against the externalist.

Consider another way that Smith states weak internalism (my paraphrase): if beliefs about what we have normative reason to do are beliefs about what we would desire if we were fully rational then judgments of practical reason should affect one's desires insofar as one is rational.[62] It is not clear to me that this formulation is any less question-begging than the former, but it may be at least a general attempt on Smith's part to state that our judgments have a sort of general rational force distinct from mere rationality as consistency between beliefs and desires.

Laying aside concerns of a question-begging notion of rationality, I shall introduce here an objection against Smith's rationality qualification that is due to Nomy Arpaly. Arpaly seeks to deny the position, common at least since Davidson, that *akrasia* is necessarily irrational by

[60] This use of subjective rationality is different than the sense in which Wedgwood uses it (as a combination of synchronic and diachronic rationality). This is meant to convey the sense of 'rationality' that is understood as 'rationality as consistency'. Wedgwood calls this synchronic rationality, and defines it as a consistency between one's beliefs and intentions. For Smith, it is not the consistency specifically of beliefs and intentions that he has in mind, but a consistency of belief states and desire states.
[61] Again, in this case, the use of 'subjectively rational' is meant to appeal to rationality as consistency not in the specific manner Wedgwood defines.
[62] For the extended statement, see (Smith, The Moral Problem 177)

denying that our judgments necessarily have any rational force of the kind suggested by Smith. Arpaly argues that if a person's beliefs or desires were irrationally attained, then not having a desire that we believe we ought to have may be the more rational state of affairs, though it be *akratic*. One of her several examples involves Sam, a college student who has left more work to do at the end of the term than he is comfortable with. He believes it would be best for him to be a hermit while in college so that he could get the most study time in and get the most benefit out of the college education that he is paying for. Imagine that while he is deliberating he leaves out facts about how his productivity would decrease due to loneliness. Failing to have the desire to become a hermit, Sam, reading Smith, would say "How irrational of me not to desire that which I believe would be best for me" (Arpaly 496-497). The force of Arpaly's objection is to deny that such a non-compliance with one's better judgment need be irrational. In fact, the more rational thing for Sam to do is to act contrary to that belief that it would be best to become a hermit during college. Arpaly's objection is well taken. Smith claims that a fully rational person would desire that which they believed they ought to desire, but Sam seems a clear example of a rational person not desiring that which they believe they ought to desire. Acting against his judgment is ultimately the more rational action.

Arpaly's example should be carefully considered. If the claim is that *akrasia* is essentially irrational, then a putative example of rational *akrasia* demands notice. To maintain that *akrasia* is irrational means to claim that Sam's case is not an example of *akrasia* or that Sam really is irrational. I think that the first response is reasonable because in not becoming a hermit though he judges that it would be best, Sam is failing to consider information that would certainly affect his views about the benefits of hermithood. In this sense it could be charged that Sam, instead of acting contrary to his better judgment is acting in concert with it, though without

realizing it at the time. In any case, the example gives us reason to doubt that becoming a hermit is Sam's sincere better judgment. However, we could recast the example subtly to preserve it as a clearer case of apparently rational *akrasia*.

Consider Samson, a university student who has to this point pursued social relationships for their ancillary benefits in terms of forming connections that may benefit him later. He has never enjoyed social interaction, and has often regarded the conversation of his fellows to be the height of dullness. While studying for exams, Samson is overcome with love of pure learning and resolves to swear off all social relationships in favor of living a life of entirely withdrawn scholarship. Samson genuinely believes that he would be happier as a hermit without all of this unpleasant socializing. He resolves to embark on hermithood immediately. But here's what happens next: Samson goes to sleep for the night and when his alarm wakes him in the morning he decides to go to his morning class, as he enjoys the learning, even if it means being around his classmates, and after class eats in the cafeteria as usual (without dwelling on the question of where to eat lunch and why), though it invites others to engage him in conversation while he is there. He follows the rest of his daily routine, perhaps avoiding some socializing that he would have pursued prior to his epiphany about the benefits of hermithood. Mostly by force of habit, Samson substantially continues his typical social routine, disliking it the whole while, and after some time realizes that he has failed to act in accord with his better judgment.

Recast in this manner, Samson's case is more clearly an example of *akrasia* than Sam's case, but now, one may wonder, if Samson would truly be happier as a hermit, then in what way is his failure to become a hermit rational? This is a fair question, and I shall add some stipulations to the example. Assume that though Samson genuinely dislikes social life, the burdens of recasting his lifestyle to avoid all social interactions would be more burdensome than

Samson could have reasonably supposed, given his ability to anticipate consequences. So burdensome as to overshadow the benefits of avoiding the socializing that he dislikes. I think that this preserves the intent of Arpaly's example. Samson judges it best to do something that he really shouldn't do, and then fails to do it. This failure saves him a great deal of trouble, so where do we get off calling Samson irrational?

Since I am claiming that one of the essential traits of *akrasia* is that it is irrational, I must in some way respond to Samson's case. Since I recast Arpaly's original example as more clearly a case of *akrasia*, I must now evaluate whether Samson is also acting rationally.

The way that Smith's position and Wedgwood's position are affected by this example shows a notable difference between the two positions. I shall begin with an answer compatible with Wedgwood's view. Making a sharp distinction between subjective and objective rationality allows us to grant an Arpaly-like intuition of Samson as rational (in one sense) and also to preserve the charge of irrationality (in another sense) given that Samson acted *akratic*ally. Wedgwood could say that Samson's case is one that involves a failure of synchronic rationality in that Samson had a belief that it was better for him to become a hermit which was inconsistent with Samson's intentions over the next several days. Samson is irrational in the sense that he is synchronically irrational (a kind of subjective irrationality), but has done something that is in his best interest (i.e. objectively rational).

I grant that Sam ends up avoiding something that is bad for him, but should we call a good outcome resultant from faulty processes rational?[63] Bringing in Wedgwood's distinction between choosing rationally and choosing correctly, we can see that Samson chose correctly in not becoming a hermit, that is, he acted in his objective interest, but Sam did not choose

[63] Aristotle mocks this position in Book 7 of the Nicomachean Ethics, considering it absurd that virtue should be the result of weakness combined with foolishness.

rationally. It is not clear what to say about the overall rationality (if such a phrase has any real meaning) of those who are subjectively irrational but are objectively rational given a certain case. Samson's case actually looks much like another example that I employed earlier to clarify Wedgwood's distinctions concerning the different forms of irrationality.

Consider again the man in Bernard Williams' example, referred to as Chuck above, who intends to drink gin, believes that the stuff in the bottle is gin, and so intends to drink the stuff in the bottle. There is no failure of subjective synchronic rationality in this man and no evidence of any violation of diachronic rationality (so long as the "gin" is in a bottle labeled 'gin' and is in a place gin might plausibly be found, and other reasonable assumptions). If the stuff in the bottle turns out to be petrol rather than gin, the man has chosen rationally, but alas has not chosen correctly. Samson, conversely, has chosen correctly but not chosen rationally. This appears to plausibly track our intuitions with regard to Samson's case. So for Wedgwood, Samson is objectively rational, but recognizes inconsistency between his judgments and actions.

Smith's account cannot access this approach. This is for two reasons. First, a belief and a desire, having as Smith says, "distinct existences" cannot be inconsistent with one another. Does a desire to do *a* and a judgment not to do *a* contradict one another? If so, what of their being "distinct existences"?[64] The second reason Smith's account cannot access Wedgwood's treatment of Samson's case is that Smith relies on this conditional notion of being fully rational (see note 64, above). I interpret 'fully rational' (I think reasonably) as objectively *and* subjectively rational. Smith's claim seems to be that the nature of the connection between normative judgment and motivation depends on the truth of the conditional 'If one is fully rational then if one judges that it is best to ϕ then one desires to ϕ'. I do not know whether Smith

[64] In fact, a hallmark of the Humean externalist position is the very distinctness between judging a thing and desiring a thing.

would strengthen the consequent to a biconditional, but Smith at least advocates the conditional relationship between judgments when fully rational and desires. This means that Smith is precisely targeted by Arpaly's objection. Samson cannot act against his better judgment unless he is not fully rational, and if he is fully rational, then he never judges that it is best to become a hermit in the first place. Arpaly's objection can only be overcome by a view which allows subjective and objective rationality to come apart.

Both Smith and Wedgwood rely heavily on some concept of rationality in their weakening of the internalist thesis (Wedgwood with more success than Smith), but the account I gave of Stroud's does not explicitly do so, relying on the qualification 'normally' instead. Stroud (Weakness of Will and Practical Judgment) provides some reasoning which indicates that what she believes she does when she weakens internalism is really to strengthen it. She begins by saying that the existence of *akrasia* denies internalism as an entailment from judgments to actions. This is quite uncontroversial, as the necessary entailment in the strong internalist thesis is directly contradicted by an action that instantiates the possibility that an agent could judge it better to φ without φing or intending to φ, as pointed out above.

Adding the qualification 'normally' isn't specific enough by itself. After all, what determines the normal conditions? The 'normally' qualification, without further specificity, would not necessarily preserve the distinction between externalism and internalism. The externalist's route to explaining why people most often (i.e. normally) act as they judge is that people *normally* desire to do what is best, or desire to act as they judge, or have some such appropriate desire.[65] In "Weakness of Will and Practical Judgment" Stroud means her use of 'normally' to presage an argument against externalism that concludes that internalism is to be

[65] I am skeptical about conative/affective states like desires having such broad content as a desire <to act as one judges> or to desire <what is best>. In any case, elaboration of such skepticism is not necessary for the present contention.

properly regarded as the 'normal' case, while externalism's claim to normality is improper (141-142).[66] It is an interesting argument, but not one that I shall address here.

Thankfully, Stroud addresses this issue in other work, describing the normal conditions as conditions of rationality. I believe that Stroud has at least a very similar idea to Wedgwood's of the kind of rationality qualification that best weakens internalism specifically against *akrasia*. In "Moral Overridingness and Moral Theory" Stroud writes, "…a charge of irrationality seems most at home when an agent acts against what *she herself takes to be* the balance of reasons. (Consider weakness of will.)" (174) This will be what I consider to be Stroud's view of what kind of irrationality *akrasia* represents, and a kind of irrationality that people *normally* lack. This is an internalist view in the sense that normative judgments have a *rationally* necessary connection to motivation. The explanatory sufficiency of normative judgment internalism is largely preserved because people are presumed to possess rationality (especially subjective synchronic rationality). Stroud describes *akrasia* above as a failure of subjective rationality, which can be read as a failure of synchronic rationality in that the agent's beliefs (what she herself takes to be the balance of reasons) are inconsistent with her intentions (she acts against that balance and is thus charged to be irrational). This general interpretation of Stroud indicates that she believes that *akrasia* represents a failure of subjective synchronic rationality, just like Wedgwood. Given what has preceded, I think that Wedgwood and Stroud's idea that *akrasia* represents a violation of subjective synchronic rationality is the correct idea. This can be contrasted to Smith's idea, which suffers from some conceptual difficulties, and doesn't differentiate itself from the externalist in a non-question begging manner.

[66] I think this argument is, more than an argument for internalism, an alternative means of effectively distinguishing weak normative judgment internalism from normative judgment externalism since externalism is just the denial of the strong normative judgment internalism thesis.

Strong Normative Judgment Externalism

An externalist may plausibly explain the entire range of behavior that internalism explains, but on the externalist account, saying that the normative judgment provided motivational force is enthymematic, leaving out (on one common account of externalism) that the person also had certain relevant desires which combined with the judgment to provide motivational force. Our general methodological preference for modesty in explanations would caution us not to claim more than necessary. Explaining the motivational force of judgments using only the judgments themselves is a modest explanation. The explanatory modesty of strong normative judgment internalism is a good reason to prefer it to normative judgment externalism.

However, *akrasia* provides a straightforward counterexample to strong normative judgment internalism. In adapting strong normative judgment internalism to allow for the possibility of *akrasia* (in light of evidence of the actuality of *akrasia*) the internalist must give up the superiority in terms of explanatory modesty that an unmodified normative judgment internalism possesses over normative judgment externalism. Explaining action simply by means of identifying a normative judgment must now be enthymematic under either view of normative judgments. In addition to the presence of the relevant normative judgment, the advocate of modified internalism must also assert that the agent who acts in accord with their judgment was not *akratic*. So explanations posit on the one hand a normative judgment along with an absence of *akrasia*, and on the other hand posit a normative judgment along with the presence of a requisite desire. So in terms of explanatory modesty, modified internalist explanation is on par with externalist explanation. In explaining why the woman stayed home from work, the explanation 'she judged it best to do so' must have 'and she was not *akratic*,' or the like, added to it by the normative judgment internalist, while 'and she had a desire to do what was judged

best' is added by the externalist. Given that weak internalism isn't explanatorily superior to externalism, that point becomes a draw, and the advocate of modified normative judgment internalism must turn to some other reason to prefer it over the externalist alternative.

Losing explanatory modesty is a significant cost to modifying normative judgment internalism. However, such modification does not force the internalist into accepting externalism. The internalist may cling to views that are not based on the explanatory sufficiency of strong internalism; examples include the magnetism of the good, or the role of judgments in motivation.

Since I will be defending a modification of normative judgment internalism, I will necessarily have to consider the alternative to normative judgment internalism, which is normative judgment externalism. Since externalism is simply the denial of internalism, no single positive characterization of it is forthcoming, as one could deny internalism in any number of ways for any number of reasons. Despite this, with regard to normative judgment internalism, there is one overwhelmingly prominent position that underlies normative judgment externalism. The position is often characterized as Humean because of its reliance on an understanding of belief/desire psychology as originally characterized by Hume.[67] The key feature of the Humean view is the status of desires as having a monopoly on motivational force. This Humean theory of motivation (HTM hereafter) is very common, and naturally suggests itself as a plausible reason for denying normative judgment internalism. If desires motivate while normative judgments (which are not themselves desires) do not, then there is no necessary connection between normative judgments and motivation. Of course, accepting HTM is not necessary for accepting normative judgment externalism because there may be any number of reasons to reject a necessary connection between normative judgments and motivation. Despite this, I regard

[67] See e.g. (Smith, The Moral Problem) and (Stroud, Weakness of Will and Practical Judgment 121-146)

acceptance of HTM as an important component of a theory of action that is intelligible, common, and not internalist with regard to normative judgments and motivation.

Accepting HTM, however, is only half of what is required to supply an account of normative judgment externalism. Recall above that one way of being a strong normative judgment internalist is to accept HTM and regard normative judgment as identical with desires. It follows that if desires hold a monopoly on motivational force, and normative judgments are desires, then there can be no instances of acting against better judgment, assuming that better judgment is stronger desire on this account.[68] The normative judgment externalist committed to the HTM must also hold a view of the nature of normative judgments such that normative judgments are not identical with desires.

If we limit our consideration of the candidates for moral judgments to beliefs and desires,[69] we have two possibilities concerning the identity of normative judgments. A view that holds normative judgments to be identical with desires is a strong internalist view, so long as all desires carry some motivational force. Since *akrasia* is a straightforward counterexample to strong normative judgment internalism, on the strength of the actual occurrence and observation of *akrasia*, I have been treating normative judgments as beliefs.

Normative judgment externalism is no more than the negation of the thesis of normative judgment internalism, but that negation contains some ambiguity. In maintaining the negation of 'necessarily, if an agent judges that it is better to φ, then the agent is motivated to φ', one might mean a variety of things. By contrast, the weak internalist does not negate the strong thesis, but

[68] I can imagine no other way of describing better judgment if normative judgments are identical with desires.

[69] A view that proposed some third form of mental state for normative judgments would be, in my opinion, unnecessarily novel, as the belief/desire distinction is both versatile, common in philosophical discourse, consistent with the cognitive/affective distinction operant in empirical psychology, and entirely sufficient for an account of *akrasia*.

rather qualifies the sort of necessity involved in the internalist thesis. I advocate the view inspired by Wedgwood that the connection between normative judgment and motivation is rationally necessary (where 'rationally' is understood in terms of subjective synchronic rationality) as opposed to logical or metaphysical necessity as in the strong internalist thesis. Smith appears to rely on a form of objective rationality to supply the necessary connection between judgment and motivation which opens his account up to Arpaly's objections as discussed above.

Since identifying normative judgments with desires yields strong normative judgment internalism, I shall assume for the sake of specifying the corresponding externalist view that normative judgments are beliefs. The HTM is committed to (at minimum) beliefs qua beliefs holding no motivational force, but there is a great deal of potential variety in the relationship between normative judgments and motivation (desires). At one end of the spectrum is a maximal disconnection between normative judgments and desires, and at the other end of the spectrum is a set of views that posit one kind or another of strong (but not necessary) connection between normative judgments and desires.

It is beside the point of this analysis to detail Hume's actual views concerning the relationship between normative judgments and desires, but it is in my opinion one of the most difficult interpretive challenges for those interested in Hume's moral psychology. I think that the textual evidence is ambiguous as to whether Hume thought that normative judgments were just desires, expressed desires, were caused by desires, or were beliefs not necessarily connected to any desires. So any view that grants desires motivational monopoly is by my reckoning entitled to the label 'Humean'.[70]

[70] See (Bricke 220-224) for an excellent discussion of the interpretive difficulties inherent in pinning down Hume's actual account of *akrasia*.

The view that normative judgments are independent of desires has no trouble explaining why some actions can be actions against better judgment. If the judgments themselves are motivationally inert, and the desires that actually carry motivational force are not necessarily connected to the presence or absence of any particular normative judgment, then it should be no surprise that some actions accord with our normative judgments and some do not. On the most extreme version of this view, the connection between normative judgments and desires is the most contingent connection possible: coincidence. I shall refer to this view as strong normative judgment externalism.[71]

Note that even strong normative judgment externalism does not imply that beliefs have no causal role at all in actions. After all, the most fervent desire for a taco is inert without some beliefs about where and how tacos are to be obtained. However, normative judgments have an element of evaluation that beliefs like "there is a taco shop down the street" lack. The interesting questions for my account begin with asking what the motivational picture is like for cases in which I have a desire for a taco, and a belief that there is a *good* taco shop to my left, and a *bad* one to my right (this is an explicitly evaluative belief). If this evaluation is accounted for as an expression of desires, as one form or another of normative non-cognitivism would have it, then intentional action contrary to such an evaluation is unintelligible. If instead I express the normative judgment that I *ought* to have a salad rather than tacos (It would be *better* to have salad rather than tacos), and the source of normative approbation is just an expression of desires that I already have, then again, it is unintelligible to consider intentional action against this judgment (which expresses my greater desires). This accords with what I have already pointed out: that *akrasia* is a counterexample to normative judgment internalism.

[71] This appears to be the "Humean" position addressed by Sarah Stroud in "Weakness of Will and Practical Judgment"

The strong externalist view (the view that treats even normative judgments as beliefs and that holds that these beliefs have only a coincidental connection to desires) has no trouble at all explaining action contrary to better judgment. Such views do, however, have trouble explaining why so many of our actions do conform to our normative judgments, and also why appealing to normative judgments is explanatorily useful (even if not sufficient).[72]

I believe that my account of the actuality of *akrasia* is damaging to strong normative judgment internalism, and I have made it my project to preserve what is plausible about internalism while accommodating *akrasia*. While strong normative judgment externalism can accommodate *akrasia*, it doesn't explain why *akrasia* is a problem. Specifically it doesn't account for *akrasia*'s being irrational in the relevant (subjective synchronic) way. In addition to this difficulty, I will argue in the next chapter that strong normative judgment externalism does not adequately cohere with evidence concerning the effectiveness of behavior modification strategies that are actuated by means of cognitive states independently of conative/affective states.

While I believe that my account of *akrasia* and its consequences are a very serious challenge to both strong externalism about normative judgments and also strong normative judgment internalism, I am at some level agnostic concerning the ultimate truth of the HTM taken on its own.

Weak Normative Judgment Internalism and the HTM

The Humean theory of motivation is very well-entrenched in philosophy, but it does not have a monopoly concerning intuitions about the kinds of mental states that carry motivational force. For my own part, I do not see that there can be no overlap whatsoever between cognitive

[72] Stroud, in "Weakness of Will and Practical Judgment" offers this concern as an argument against Humean externalism. See pp.141-146

and conative states, or that desires should hold a monopoly over motivation. My own intuition on this matter is that there seem to be both cognitively heavy and cognitively light motivational states.

Examples of cognitively light motivations are those that strike me as more immediate and bodily. Hunger is a desire for food, thirst a desire for drink, and the cognitive states required to facilitate action are limited to perception and memory. Cognitively heavy motivations, contrarily, have less propinquity and more abstract ends. If I am motivated to write philosophical work that defends the family of views in the philosophy of mind known as functionalism, or if I am motivated to argue that human-made entities could potentially be moral subjects so that in some distant future that includes autonomous androids (if there even is such a future) humans don't treat the androids as slaves, then these would strike me as cognitively heavy motivations. Such stories of cognitively heavy motivation do not easily reduce to more basic kinds of urges that human beings have.

Most pedestrian examples also seem to me to admit of both cognitively heavy and light motivations. Consider a shopper in a supermarket. People are naturally attracted to (that is, they naturally desire) foods that are high in sugar and fat because those foods tend to taste most pleasant. Our sense of taste tends to identify more caloric choices as more pleasant. This makes sense as an evolutionary mechanism for assisting us in selecting the most caloric foods in an environment in which calories are hard to come by. But at the same time, our supermarket shopper can recognize that they do not require the calories that our ancestors probably required, and that a more healthful diet is likely to be better for their health and happiness (but of course who knows which food expert to believe these days concerning what is and is not healthful). And of course the shopper might decide that they just don't care about health or their Omega-

6/Omega-3 ratio and end up with potato chips and cookies. The shopper might also leave with salmon, sweet potatoes, and pomegranates, and depending on how much attention they are paying to what they are doing, a couple of candy bars, or none.

I hope that it is clear from the shopper example that appeals to cognitive evaluation as well as conative evaluation can fit variously well with any of what goes on in the above example. The anti-Humean can appeal to the role of explicitly evaluative beliefs in the motivational story that are distinct from, and that even oppose desires. The Humean can just as easily reply that the presence of *some* desire is required in order to motivate the shopper to care about health in the first place.[73]

To further the dialectic between anti-Humean and Humean explanations of motivation, I shall introduce an example adapted from Shafer-Landau. Betty believes that it would be good for her to become a lawyer, so she applies to various law schools, selects the one that fits most of the criteria she considers to be important in a law school, and enrolls in that law school. She finds herself increasingly unhappy with her coursework, and while volunteering on a Habitat for Humanity worksite, discovers a love of carpentry and is convinced that she can do more good by serving Habitat than becoming a lawyer, so she pursues and lands a job as a Habitat site supervisor (Shafer-Landau 122-141).[74]

What makes this example work in the way that the anti-Humean wants is that it purports to resist the Humean position that all of our normative or evaluative judgments are at base desires that we currently hold. Betty is acting on the basis of what she thinks is best, or perhaps even trying to fulfill desires that she thinks she will have (but importantly, that she *does not* and *will not* have).

[73] This is the move that Hume makes in the EPM when arguing for sentiment as the basis of morality.
[74] Shafer-Landau provides an excellent discussion of issues in the humean/anti-humean debate, and seems to regard presumption as favoring the anti-humean.

The adherent of the HTM of course has a plausible response to cases like these. The Humean can point out that other desires (e.g. for prestige or high salary that she thought a career as a lawyer could bring, or a desire to please her parents, etc.) could well be implicated in adequately accounting for Betty's motivation to become a lawyer.

The call and response between the Humean and the anti-Humean in cases like Betty's illustrates something troubling about the dialectic. While the Humean explanation for Betty's motivation to enroll in law school is entirely plausible, only someone already convinced of the truth of HTM would look for motivations beyond what Betty *judged* would be best. On the other hand, only someone already committed to the truth of an anti-Humean position would regard Betty's normative judgment to be of itself sufficient for explanation. It is not clear that cases like Betty's can adequately contribute to determining presumption in the Humean/anti-Humean dialectic.

As a thought experiment, consider what would result if Betty simply had no desires. It might seem plausible to convince her to enroll in law school on the basis of some moral duty to provide for her parents as her parents provided for her, combined with making a case for law being the path most likely to provide financial success. This requires a sort of Kantian appreciation of moral duty as independent from desire ("inclination" in Kantian terms) but such a theory is at least on the table. While I am sympathetic to the anti-Humean view that normative judgments as belief states are sufficient to explain motivation, I can imagine no considerations that would prevent the Humean from denying that Betty could be motivated to do *anything at all* lacking desires. After all, why would duty be important to her? What could motivate her even to listen to arguments? Not only is the modified Betty case far from the actual world, I can imagine no way to conceptually or empirically test it.

The notion that normative judgments carry motivational force (an anti-Humean view) and the notion that normative judgments must generate, influence, stir-up, or access some desire in order to motivate (a Humean view) are both able to account for Betty's actions. The view that there is *at least some* law-like and regular (though defeasible) connection between normative judgments and desires could appropriately be termed 'weak normative judgment externalism'. I must admit here that it is not clear at what point a suitably weak version of normative judgment internalism is at all distinguishable from a suitably weak version of normative judgment externalism. Both views posit a law-like and regular connection between normative judgment and motivation for all synchronically rational agents. I regard adherence to the HTM as the main difference between the internalist and externalist views of normative judgment. That is, an externalist view, adhering to the HTM, requires the presence of an appropriate desire for motivation (a desire that does not fail to be present in any synchronically rational agent). An internalist view requires no desire, regarding normative judgment as itself motivationally sufficient for all synchronically rational agents.

My account of the causal factors that serve to undermine synchronic rationality in cases of *akrasia* appeals to cognitive states and cognitive factors, but does not rule out a rationally necessary connection between desires and normative judgments such as would be invoked by an adherent to the HTM. Because I do not subscribe fully to the HTM, I shall not supply an account of what kind of connection that this would be, and thus I supply no specific account of weak normative judgment externalism. Instead I will refer to this family of views as Humean externalism, and will direct my reply to this whole family of views, rather than any specific version of Humean externalism. I will argue in my next chapter that the weak internalist

understanding of the relationship between normative judgment and motivation is the correct one, given the account of *akrasia* that I have presented.

Conclusion

In this chapter I have specified the implications of the cognitive bias account of *akrasia* on several theories of moral motivation that are most significantly impacted by *akrasia*.

Any version of strong normative judgment internalism is at odds with the very existence of *akrasia*. On the strength of my account of *akrasia*'s actuality, then, I have supplied a good reason to doubt that strong normative judgment internalism is true. I have, however, weakened the strong thesis to hold that there is a *rationally* necessary connection between normative judgment and motivation, and in the process have provided a detailed account of the irrationality that *akrasia* represents while preserving as much as possible what is plausible about normative judgment internalism.

While my account is neither supportive of nor decisive against the Humean theory of motivation taken by itself, I do believe that a suitable analysis of *akrasia* demonstrates that strong normative judgment externalism must be false. Strong normative judgment externalism results from HTM's combination with the idea that normative judgments are beliefs and that there is nothing stronger than a coincidental connection between normative judgments and motivation (desires). This is a significant result due to the fact that strong normative judgment externalism is commonly regarded as a good explanation for *akrasia*. To wit, when someone judges that A is better than B and chooses B, the assumption is that the person who judged that A is better than B must simply not have desired A. A weaker normative judgment externalism that posits a more regular connection between judgment and motivation (desires) than coincidence is included as part of a family of views I term Humean externalism. The next chapter contains an

account of the incompatibility of Humean externalism with a preponderance of empirical evidence and conceptual analysis.

To summarize in preparation for the next chapter, a Humean may be an internalist or an externalist with respect to normative judgments. The Humean internalist is committed to normative judgments being non-cognitive and motivational (i.e. desires). This is strong normative judgment internalism, to which *akrasia* is a counterexample. The Humean externalist has no trouble allowing *akrasia* because she is committed to normative judgments being cognitive, but (per HTM) non-motivational. This is strong normative judgment externalism, and it has some conceptual difficulties both with explaining why we regularly follow our normative judgments and in distinguishing *akrasia* from compulsion. I have been advocating a weak normative judgment internalism that regards normative judgments as cognitive and motivational *insofar as one is synchronically rational* (the weakening clause). This position allows *akrasia* and allows for preservation of what is plausible about normative judgment internalism. This view specifies a strong but defeasible (causal) relation between normative judgment and motivation, which is the defeating of synchronic rationality by cognitive bias (the full account outlined in Chapter 2). A Humean could, however, hold that desires (defeasibly) cause normative judgments, which are cognitive states and (per HTM) non-motivational. This is weak (Humean) normative judgment externalism. The next chapter provides an argument against this view and in favor of weak normative judgment internalism.

CHAPTER 4: ON NOT BEING *AKRATIC*

The cognitive, weak internalist account of *akrasia* has, as I mentioned in the previous chapter, one chief rival in Humean externalism. Both weak normative judgment internalism and Humean externalism are compatible with the existence of *akrasia*, and each view is compatible with a reasonable explanation of *akrasia*. In this chapter I will contend that the cognitive, weak internalist account of *akrasia* is the better account because it is more compatible with an empirically informed means of *not* being *akratic*.

The chief difference between weak normative judgment internalism of the sort that I have been advocating and Humean (weak) normative judgment externalism is that the internalist view holds that normative judgments, regarded as cognitive states, cause motivation, while the Humean position regards motivation as essentially non-cognitive. In the case of *akrasia*, the truth of weak normative judgment internalism would imply that once cognitive bias is modified, *akratic* behavior is reformed, while truth of the Humean position would imply that only modification of desires would reform the *akrates*.

Suitably interpreted, this is an empirical question. In this chapter, I have assembled evidence and explanation that shows that modification of cognitive states as opposed to conative or affective states is more reliably indicative of behavioral change in the *akrates*. Much of this evidence is taken from studies of addiction and clinical approaches to reforming addicts. So I shall begin this account by describing the relationship between addiction and *akrasia*.

Akrasia and Addiction

I mentioned briefly in Chapter 2 that it would be improper to regard *akrasia* as a form of mental illness or psychopathology because it would mean that the traditional feature of blameworthiness would have to drop off of *akrasia* or be strongly mitigated. I wish to return to the blameworthiness of *akrasia* to discuss what it is that a person is expected to do in order to

avoid the disapprobation justly due the *akrates*. To this end, I would like to bring in a distinction between intemperance, *akrasia*, and compulsion as they are differentiated by blameworthiness.

Intemperance: An agent pursues a course of action, *c*, that is objectively incorrect (i.e. that by some reasonable account he ought not to do), while making no normative judgment opposing his doing *c*. Intemperance is a failure to be motivated to behave as one ought. The agent makes an objectively poor choice, and is generally blamed for choosing so.

Akrasia: An agent pursues a course of action, *c*, that by her own judgment is not the best course of action open to her. The agent chooses irrationally and is generally blamed for choosing irrationally, but may be given some credit for knowing better and regretting *c*, unlike the unrepentant intemperate.

Compulsion: An agent cannot help but pursue *c*, whether judging that *c* is the superior or inferior option. The agent cannot alter his own compulsive behavior, and this is why it is called compulsion. The agent doesn't choose and is not generally blameworthy (except insofar as he allows the conditions for compulsion to obtain, and/or does/did not seek help in redressing his compulsive behavior).

This distinction is important chiefly because our approbative responses to each of these phenomena are different. In the case of the intemperate, we blame the lack of motivation to do what one ought. The way we go about reforming the intemperate is by convincing them of what they ought to do or not do, and if that fails, we generally try to motivate the intemperate through reward or punishment to assist them in ceasing to behave recklessly. The *akrates* is culpable for behaving as they themselves would condemn, but the fact that they themselves condemn it often gets partial credit. Instead of having to convince the *akrates* of the best way to act, we need only assist the *akrates* in attending to her better judgment. The compulsive is a case in which

persuasion or another ordinary sort of motivational change is not effective. Generally, we regard compulsive behaviors as psychiatric pathologies of one kind or another, and attempt clinical interventions if the compulsion interferes with the subject's ability to live a normal life.

Normative judgment internalism characterizes the difference between the three cases in terms of the presence or absence of the relevant normative judgments. What intemperate and *akratic* individuals are blamed for is plausibly actuated on the status of their normative judgments vis a vis their actions. The intemperate simply has no normative judgment contrary to their action (but could have such a judgment and ought to). The *akrates* acts contrary to normative judgment (but could have and ought to have avoided such action). Compulsives cannot do other than they in fact do, whatever they judge, and so given the reasonable and common view that 'ought' implies 'can', they are not blamed.

Humean normative judgment externalism explains *akrasia* in terms of the permanent irrelevance of normative judgment (understood as cognitive states) to motivation. Action against normative judgment occurs when someone judges that x is better than y but desires y more than x, and so does y. Under this view, the strongest desires supply motivational force, so people do whatever they most desire to do. Many have found this explanation of *akrasia* plausible (Mele, Irrationality) (Stocker). If this is the best explanation for *akrasia*, then reforming the *akrates* consists in some form of desire modification.

If the Humean externalist explanation is really the best account of *akrasia*, then the approach that it suggests toward reforming the *akrates* should be the one that demonstrates the best success. If, on the other hand, a primarily cognitive approach is most effective, it is reasonable to conclude that *akrasia* is a primarily cognitive problem. In other circumstances, this kind of reasoning bears fruit. If a mechanic replaces part A, and the problem is subtly

affected, while replacing part B largely or completely fixes the problem, then the mechanic can reasonably conclude that part B is the largest part of the problem. What I intend to demonstrate in what follows is that success or failure in reforming frequent *akrasia* is actuated more by cognitive factors than desiderative factors.

This task is faced with the same problem that I faced in developing an empirically informed account of *akrasia* in Chapter 2. There has been, to my knowledge (and I have searched extensively) no study of reforming anything called '*akrasia*', so I am faced with the task of finding something that has been studied that matches the criteria for *akrasia*, even though the terminology under which it is studied in psychology, psychiatry, and physiology is not the same as philosophical terminology.

Akrasia per se has not been the subject of empirical study, but one sort of frequent *akratic* behavior that has received a great deal of empirical study is addiction. Instances of addiction as examples of *akrasia* are nothing new in philosophy, but even so, I shall go to some length drawing parallels between the cognitive account of *akrasia* and addicts continuing to engage in the addictive behaviors despite judging that they ought not.

The first step in this process is to recast the distinction between the intemperate, the *akratic*, and the compulsive (above) as a distinction between different sorts of addict. To this end, I appeal to a set of cases supplied by Gary Watson that instantiate the above definitions of intemperance, *akrasia*, and compulsion (Skepticism About Weakness of Will 324):[75]

(1) The reckless or self-indulgent (intemperate) case: the woman who knows that having another drink will likely result in her becoming drunk and unable to fulfill other obligations, but

[75] I do not know if Watson, if pushed, would accept Humean normative judgment externalism, but his account in this piece does seem to take seriously the main theoretical commitments of strong normative judgment externalism. See (Smith) for commentary on Watson's distinction.

who prefers the drink and accepts the consequences. She acts in accordance with her best judgment.

(2) The weak (*akratic*) case: the woman who judges that it would be better not to drink, who could have refrained, but did not. She acts contrary to her judgment.

(3) The compulsive case: the woman who judges that it would be better not to drink but who was unable to refrain. She also acts contrary to her judgment.

It squares with common experience that addicts are often, at various times, one of the three above. Of course, for my purposes, I shall focus attention on the addict who judges that they ought not behave as they do, and who is capable of avoiding that behavior.

The example Watson brings in has to do with choosing to drink another drink of alcohol. This is appropriate, as many instances of failing to do as we judge that we ought are often bound up in (at the mild end of the spectrum) bad habits or (at the more severe end) very serious addictions. Failures to change our habits, like when starting a new diet, are frequently cited as candidates for *akrasia*. In the context of discussing addiction, I shall provide an account of cognitive bias modification along with evidence of its effectiveness in assisting persons in breaking addictions. Again, this is an important part of the account because it demonstrates that the sort of weakness involved in *akrasia* is a cognitive weakness because it is most effectively remediated by addressing its cognitive aspects. In identifying the correctable weakness implicated in at least these cases of *akrasia*, I shall be identifying the weakness that the *akrates* is blameable for failing to correct.

Before I begin with the main discussion, I would like to point out an issue that may arise that might make the following account more likely to be misunderstood. I have made extensive use of empirical data from psychology in developing the cognitive bias account of *akrasia*, and I

shall make use of literature primarily from psychiatry in detailing cognitive bias modification as it pertains to reforming *akratic* behavior. I do not thereby mean to give the impression that I regard *akrasia* to be a pathology requiring clinical intervention. I hold what I believe is a common view of pathologies requiring clinical intervention. That is, I view such pathologies as examples of compulsion rather than *akrasia* or (at least ordinary) intemperance.

In fact, in denying the existence of *akrasia*, some have characterized reported cases of *akrasia* as instead being cases in which the dictates of a person's best judgment are psychologically impossible for her to follow (Hare, Freedom and Reason).[76] In any of these cases (some of which surely must exist) the action that takes place against better judgment is not intentional, and thus is not *akrasia*. It is instead a version of psychological compulsion. In adapting psychological literature to empirically inform the philosophical concept of *akrasia*, it would be tempting to identify *akrasia* with an existing mental disorder.[77] I encourage the reader to resist such temptation, and regard clinical pathologies as more akin to instances of compulsion (in the sense that compulsion operates in Watson's example) than to *akrasia*. Even if psychologists and psychiatrists do not generally regard pathological behavior to be unfree, the more common view of praiseworthy or blameworthy action involves action that is suitably under the control of the individual in question.[78] I shall avoid entry into any debates concerning metaphysical free will. The common view may or not be ultimately mistaken about pathological behavior, but at the very least I shall be able to provide an account of cognitive bias modification that addresses *akrasia* but that is not a strategy that requires or is restricted to clinical intervention.

[76] See especially Hare's Ch. 5. See also (Hardcastle) for a critique of an attempt to reduce psychological explanation of *akrasia* to neuroscience.
[77] This is largely what Kalis, et al. attempt. See Ch. 2.
[78] Aristotle, NE book III, agrees.

An addiction is a pattern that persists over time, while *akrasia* is episodic. Of course some are *akratic* more often than others, and addictive behaviors will be correlated with more frequent occurrences of *akrasia*. However, not all addictions are created equal. There appear to be many different sorts of addictions, some involving chemical dependence, and the strongest of these may appear better examples of compulsion than *akrasia*. Also, some addictions, like my own utter dependence on my morning coffee, are, if anything, examples of intemperance rather than of *akrasia* as most people do not care to break their mild to moderate caffeine addictions.

I shall like to leave aside both these most severe cases of addiction and the mild addictions that people generally don't regard as particularly bad or worthy of effort in breaking. I contend that there are sufficiently many examples of addicts who are capable of controlling and/or breaking their addictions, judge that it would be best to do so, and still sometimes fail to perform the individual actions that eventually lead to the breaking of a bad habit. These phenomena are rather well studied.[79]

Common experience tells us that at least some addictions that addicts wish to break involve instances of action against better judgment. It is part of our common knowledge of alcoholism, for instance, that most alcoholics do not think it best that they continue to be alcoholics.[80] As a necessary step in demonstrating that cognitive bias modification is an effective remedy for *akrasia*, I must demonstrate the role that cognitive bias plays in those addictions that include examples of *akrasia*. What follows then is an account of the cognitive aspects of addiction that match up with the account of *akrasia* as given in Chapter 2.

---

[79] See (Campbell), which is one of a very few articles that specifically identifies addiction with "*akrasia* or weakness of will".

[80] For voluminous examples of this see: (Alcoholics Anonymous), or any other collection of testimonies of alcoholics or recovering alcoholics. See also Chapter 1, for my contention that in order to properly distinguish *akrasia* from intemperance we must take at least some of the things that people say about their own judgments at face value.

Cognitive Bias in Addiction

Two kinds of cognitive bias are at play in both substance and behavioral addictions that do not involve substances. One sort of cognitive bias involved in addictive behavior is a bias that minimizes recall of the negative effects of the addictive behavior. Let us refer to this as recall bias.

Typically, rewards for behaviors tend to reinforce those behaviors, while negative consequences for behaviors tend to discourage repetition of those behaviors. Long experience with conditioning, incentives, and disincentives tells us that such a connection is as regular and reliable as any psychological law. Objectively, addictive behaviors are often harmful. A failure of the addict to reform his or her own behavior, or even to recognize the problem, is a cognitive failure—a form of subjective irrationality as well as a failure of objective rationality.

My account of *akrasia* has specified that *akrasia* is a failure of synchronic rationality, and not a failure of diachronic rationality. If addicts believed that their addictions were not harmful or if they misestimated the consequences of their addictive behaviors in a way that additional information or a different way of considering things would fix, then the addict would demonstrate a failure of diachronic rationality. Surely this is what happens some of the time, but it does not capture the full range of mental processes often associated with the persistence of addictive behaviors. Those who seek to give up their addictions often do so on the basis of the past negative consequences of addictive behaviors. It is the failure of their own past negative experiences to sufficiently motivate addicts that is, in a way, paradoxical. Being able but not disposed to remember negative consequences of addictive behaviors fits well with Aristotle's talk of having but not attending to knowledge, as well as the more empirically respectable talk of information that is or is not present in the global workspace.

The phenomenon of recall bias makes a charge of subjective synchronic irrationality (*akrasia*) intelligible and empirically verifiable. It is not that the addict believes things about their addiction that are false, or that they must revise. Instead, the past negative consequences of the addictive behavior are often not recalled at all.

William Campbell, a fellow of the American Society of Addiction Medicine, is one of a few who explicitly link *akrasia* in the case of addiction to a specifically cognitive impairment (recall bias). Campbell describes the causal relevance of cognition in addictive behavior as follows (the italics are my own):

> "Addicts appear to be acting at various times on 2 different belief systems. The first belief is that the addictive behavior is harmful and produces negative consequences…The addict appears to act on the basis of faulty reasoning, and the actions are such that *cognition does not appear to consider* the previous negative consequences of the addiction." (671)

This is a clear description of *akrasia* and its classification as a failure of subjective synchronic rationality. At this point it is tempting to ask what feature of addiction causes this lack of recall. This is a subtle confusion. It is like asking what it is about forests that causes trees to clump closely together. Campbell is arguing for cognitive bias as a causally necessary aspect of the etiology of addiction. It is not just clinicians who appear to hold this view. Campbell cites some literature from Alcoholics Anonymous, an organization with a wealth of practical experience that should not be discounted. In particular, Campbell singles out the statement that "…we shall describe some of the mental states that precede a relapse into drinking, for obviously this is the crux of the problem."

Recall bias is not the only kind of cognitive or attentional bias implicated in addictive behavior. Another sort of cognitive/attentional bias implicated in addiction occurs in the increased attention to addiction-related stimuli in the addict. We may refer to this as focus bias. Focus bias and recall bias serve together to make the addict more aware of the presence of temptation and less cognizant of its previous bad consequences. In Chapter 2, I discussed the general usefulness of one-reason decision heuristics, and also their propensity to be misapplied in certain situations. The heuristic gets labeled a 'bias' when it gets misused. Consider the ordinarily useful traits of selective attention and memory. Having our attention drawn to the fastest moving object in our surroundings can have survival value. Often, fast-moving things are dangerous (charging predators) or else are opportunities for food (fleeing prey). Caloric items present themselves readily to the attention because there has historically been value in knowing where the calories in our environment are. Generally, the ability to see what we want more readily than what we don't want is very useful. In the context of addiction, such tendencies are positively and powerfully counterproductive. It makes sense on these lights to regard addiction as a misapplication of the ordinarily useful cognitive tools that are selective attention and memory.

Medical and psychological researchers, in studying addicts and their characteristic behaviors, have noticed a number of ways in which addicts of various kinds share cognitive traits. These traits have become an integral part of understanding the cognitive aspects of addiction. Focus bias, as it is studied, consists in the following: a tendency of addicts to respond to certain cognitive cues more quickly than non-addicts, a reduced tendency of addicts to disengage attention from addiction-related cues and onto non addiction-related cues, and a

reduced tendency compared with non-addicts to distinguish target cues from distracters (Mazas, Finn and Steinmetz).

It is significant to recognize that these same sorts of cognitive biases contribute to a startlingly wide range of addictive behaviors, which includes both addictive behaviors that do and addictive behaviors that do not involve any psychoactive or mood-altering substances.[81]

A wealth of evidence suggests that increased attentional bias toward addiction-related stimuli predicts relapse of addiction among a startling diversity of addictions. As one example of attentional bias in addicts, a study by Liu et al. made use of what is known as a Stroop task to demonstrate focus bias in cocaine addicts (Liu, Lane and Schmitz). In a Stroop task, cocaine addicts and controls are contrasted in their abilities to identify the color of a word while ignoring the word's meaning. The word is presented, and subjects (both cocaine addicts and controls) are asked to press color coded buttons corresponding to one of the potential colors of the presented words as quickly as they can accurately do so. Some of the words are cocaine-related (e.g. 'cocaine', 'dealer', or 'freebase') while an equal number of words are neutral with regard to cocaine and length-matched with the cocaine-related words (e.g. 'cabinet', 'window', and 'armchair'). A significant difference in cocaine addicts' reaction times to neutral versus cocaine-related words is evidence of attentional bias to cocaine-related stimuli. Controls show no significant difference in reaction time to cocaine-related versus neutral stimuli. The Liu et al. study confirmed the results of other studies (Hester, Dixon and Garavan) (Vadhan, Carpenter and Copersino) that find an increase in what is above termed 'focus bias' among cocaine addicts.

Further, Liu et al. write "[I]mproving impulse control and remediating attentional bias may prove to be helpful tools in the treatment of cocaine dependence." (121) It stands to reason

[81] Though Seamus Decker and Jessica Gay claim that research into the role of cognitive bias in addiction is scarcer for "evidence about behaviors that do not involve drug use or other physiological factors". See their (Cognitive-bias toward gaming-related words and disinhibition in World of Warcraft gamers).

that if the remediation of cognitive bias would assist in the treatment of cocaine dependence, other sorts of chemical addictions should admit similar amenability to cognitive bias modification as effective treatment. Some evidence confirms this suggestion, indicating that higher attentional bias negatively correlates with the success of treatment outcomes for alcoholics (Cox, Hogan and Christian) and similar confirmation in the case of smokers (Janes, Pizzagalli and Richardt). The Janes et. al study is particularly interesting. The study measured brain reactivity to cues related to cigarettes and to smoking, and concluded that there was a strong negative relationship between brain reactivity to smoking-related cues and likelihood of continued tobacco abstinence among smokers who wish to quit smoking (our *akratic* addicts). They also found a correlation between brain reactivity (measured by fMRI data) and attentional bias (measured by a Stroop task). In concluding "…that prequit brain reactivity to smoking-related images is greater in smokers who eventually slip after attaining brief abstinence with NRT and that anterior insula and dACC fMRI cue reactivity correlate with an attentional bias to smoking-related words." Janes et al. provide a biological/neurological confirmation of the role played by attentional bias in addictions.

The empirical evidence for the important role that attentional bias plays in addictive behavior is not restricted to chemical addictions like alcoholism or addictions to cocaine or tobacco. Other studies have uncovered similar attentional bias (characterized by focus bias and recall bias) among overeaters (Nijs, Muris and Euser), pathological gamblers (Boyer and Dickerson), and computer gaming addicts (Decker and Gay).

Decker and Gay, studying computer gaming addiction, used an Affective Shifting Go/No-go Task (ASGNG) to measure cognitive bias toward gaming-related cues among habitual players of a particular video game against a control group of non-players. The ASGNG task is

similar to the Stroop task. A set of positively valenced common English terms as well as positively valenced jargon specific to the video game are targets, while negatively valenced English and jargon counterparts are distractors for some trials, vice-versa for other trials. Subjects are asked to identify the targets by pressing a button when they are displayed, and are instructed not to press the button for the distractors.

So each subject would be expected to press the button for a word like 'friend', a positively valenced English word, as well as for 'purple', a positively valenced word for World of Warcraft players.[82] Subjects would likewise be expected to leave off the button for negatively valenced English or World of Warcraft phrases, like 'betray' or 'nerf'[83] respectively.

The World of Warcraft players demonstrated cognitive bias toward game-related stimuli by more quickly and accurately distinguishing between game-related targets and distractors than English targets and distractors, and also distinguished game-related targets from game-related distractors more quickly and accurately than the control group of non-players distinguished English targets from English distractors. Decker and Gay conclude: "Similar to past research showing that recovering alcoholics had cognitive-bias to alcohol-related words, [game players] with high rates of time spent playing computer games showed cognitive-bias toward gaming-related words." (807-808)

It has long been clear that cognitive performance can be habituated—practice enough memorization and you will become better at memorizing things, even without intentionally trying to do so. The role of habit and cognitive bias in the case of the addict seems to be a kind

[82] The most powerful and desirable pieces of weaponry and armor in World of Warcraft are most easily distinguished by their names written in purple text (for rare or epic items) versus blue (for merely uncommon items) or green (for run-of-the-mill items). Players refer to receiving such an item as, e.g., 'getting a purple'.

[83] Blizzard, the company that maintains World of Warcraft, often makes changes in the abilities of certain classes of players' characters. Such changes that serve to make a class of character relatively more powerful are known as 'buffs' while such changes that make a class less powerful are known as 'nerfs'.

of feedback loop. The addict trains herself to recognize and seek addiction-related stimuli, and this makes the attentional bias toward addiction related stimuli stronger. If attentional bias really is as central to addiction as the evidence suggests, this feedback loop would explain why those who have been addicted for a greater period of time find it harder to break an addiction. The attentional bias is more highly habituated in the long term addict.

Because the same forms of cognitive bias are observed accompanying so many varieties of addiction, it is reasonable to postulate that these cognitive biases are central to what we mean by 'addiction'. Evidence that the degree of cognitive bias varies concomitant with the strength of the addiction (measured in rates of absinence from the addictive substance or behavior) is further reason to believe that cognitive bias is an essential element of addiction. Since addictive behavior is often contrary to the better judgment of the addict, addiction provides a rich field of examples for the cognitive bias account of *akrasia*.

It is worth noting that in the philosophical tradition, examples of people wanting to change their behavior but failing to do so often involve bad habits or addictions. Unifying an empirically informed account of *akrasia* with empirical evidence concerning the role of cognition in sustaining addictions is a philosophically and scientifically significant development. It is philosophically significant because it is the first appearance of a thoroughly empirical account of a *long*-discussed phenomenon. It is scientifically important because it serves to unify separate avenues of research under a broader aegis. Given a clear empirically informed account of *akrasia*, the interested empirical researcher has a starting point in further studying *akrasia* as such, rather than inadvertently revealing elements of *akrasia* while studying addictions, cognitive biases, or decisional heuristics.

As I am primarily interested in the philosophical importance of the empirically informed account of *akrasia*, I shall briefly point out how the empirically informed account contributes to, and in some sense completes prior philosophical perspectives on *akrasia*.

In Aristotle's diagnosis of *akrasia*, undertaken to refute the position that *akrasia* is psychologically impossible, Aristotle proposes that *akrasia* is the result of having but not attending to knowledge of the good. Lacking the vocabulary of modern behavioral psychology, Aristotle appears to have anticipated, albeit in a very general way, the empirically informed explanation of *akrasia*. Replacing vague notions of having but not attending to knowledge with detailed empirical accounts of cognitive/attentional bias preserves the spirit of Aristotle's feeling concerning an appropriate explanation for *akrasia* and adds an empirically verifiable phenomenon on which to ground an explanation of *akrasia*.

Similarly, Davidson, in developing an account of the logical possibility of *akrasia*, relies on a distinction between all-out judgments (judgments that consider everything relevant to the evaluation) and judgments with a prima facie operator that take the form pf(x is better than y, r) where r is the evidence considered. Davidson does not consider (as it is outside the scope of his paper's limited purpose) whether the difference between all-out judgments and prima facie judgments is empirically verifiable. The empirically informed account of *akrasia* that I have been advocating fills this gap in this overall story of *akrasia* as well as Aristotle's. Because human beings are incapable of simultaneously considering all relevant evidence at the same time, and frequently act upon judgments of the form outlined above, it is clear that we ought to see cases of action based on evidence that is more apparent or that is attended to *first* (prima facie

judgments) than judgments based on evidence that a more patient thought process would reveal as superior.[84]

The idea that a specific cognitive *weakness* explains the difference between the addict who sincerely judges that they ought to break their addiction and still relapses has both commonsense currency and also empirical verification. If a computer gaming addict (for example) is more apt than the non-addict to take notice of gaming-related stimuli, and also apt to respond more quickly to gaming related stimuli than the non-addict, then it should not be surprising for their decisions concerning computer gaming are more frequently made on the basis of prima facie judgments with gaming-related stimuli crowding out non gaming related stimuli, accompanied by a failure to recall past negative consequences of excessive gaming.

Treating *Akratic* Addiction

William Campbell, mentioned above, approaches the problem of addiction and *akrasia* from a treatment perspective. Campbell is motivated by what he sees as a problematic lack of a unifying definition of addiction that explains why chemical addictions (like alcohol and cocaine) should have so much in common with behavioral addictions (like gambling).[85] Campbell argues that the field of addiction treatment has been held back both by lack of a comprehensive etiology of addiction, and by an "accepted view" that treats addiction as primarily conative. He puts it briefly: "The accepted view is that craving causes the addict to act." (671) Campbell follows this claim with a brief refutation of the conative accepted view. First, if the craving were causative, then every time the cravings became sufficiently strong, an abstinent addict would relapse. In reality, sometimes they do, sometimes they don't. Further, sometimes addicts who experience

[84] See (Davidson, How is Weakness of the Will Possible? 40) for a formal description of better reasons supplanting inferior ones in judgment.

[85] "The present conceptualization of addiction inadequately explains addiction as an entity unto itself and does not provide any understanding of the relation between the substance and behavioral addictions" (Campbell 671)

severe craving stop their addictions. These events tell against cravings as a sufficient condition for relapse or as a necessary obstacle to recovery.

This is not an extended argument, and it is a bit simplistic, but I think Campbell's point has merit, particularly since the previously discussed evidence indicates a much more central role for cognitive states in addiction than conative states. But because the conative view is so prevalent, it is worth more detailed examination.

One might preserve the conative view against Campbell's argument and posit that whenever the desire to quit is strong enough, it can overpower even the strongest of cravings, and when it is weak enough, it can be overcome even by mild cravings. However, this idea, though common, has its own conceptual problems. The will (in this case, whatever accounts for the desire not to be an addict) is often taken to be the feature of psychology that resists or fails to resist desire, and the 'will versus desire' description of *akrasia* has been historically prominent enough to translate *akrasia* as "weakness of will".[86] Watson, who is skeptical of the view, puts the problem this way:

> "This talk of strength of desires is obscure enough, but insofar as it has meaning, there does not appear to be any way of judging the strength of desires except as they result in action…Isn't the only relatively clear measure of strength of desires [versus strength of the will] the tendency of those desires to express themselves independently of the agent's will?...If a sufficient condition of compulsive motivation is that the motivation be contrary to the agent's practical judgment,

[86] See also Davidson, "How is Weakness of the Will Possible?" p.27. Here, Davidson also characterizes a separation of "thinking we ought" and "wanting to" as the most common way of handling *akrasia*. This can legitimately be called the "received view" of *akrasia*. In (Paradoxes of Irrationality 175) Davidson expresses a similar worry to mine that the "will versus desire" picture (he calls it the Medea Principle) does not adequately distinguish *akrasia* from compulsion.

then weakness of will is a species of compulsion." (Skepticism About Weakness of Will 327-328)

In other words, the "will versus desire" picture of *akrasia* has difficulty distinguishing *akrasia* from compulsion. If some desire is so strong that nobody could overcome it, then it is a clear case of compulsion, but the evidence for this circumstance is identical to the evidence for someone with an extraordinarily weak will succumbing to a stronger, but still very weak (that is, resistable) desire.

Aside from this issue, the "will versus desire" theorist is constrained by their view to offer one of two remedies for the *akratic* addict. That is, the "will versus desire" theorist must provide some account of what it means to intentionally strengthen one's own will or else to intentionally weaken one's desires (both of which sound like things traditionally described themselves as "acts of will"). I need not belabor the inherent circularity of using one's will to strengthen one's will. Put into layman's terms, the addict who judges that they ought to quit and is unsuccessful in quitting needs to find a way to either want the addiction stimulus less or else to want to quit more. Such a view is dependent upon some successful method of desire modification.

Despite the status of the "will versus desire" view as the received view of *akrasia*, it appears that few actually endorse the view in its entirety, while many argue against it. I have no intention of building up the naïve "will versus desire" view[87] because my primary opponent is the Humean externalist. I bring up the "will versus desire" view because it shares one particular problem with the Humean externalist, and that is how to best account for reform of the *akrates*.

If the *akrates* is to act in accord with their judgment that x is better than y, then the Humean externalist must come up with an account for desiring x more strongly or desiring y less

[87] See the latter half of (Watson, Skepticism About Weakness of Will) for an attempt at this.

strongly. This is what Campbell has in mind in referring to a conative approach. It is my intent to show through other additional evidence and analysis that it is much more productive to approach addiction from a cognitive angle than from a conative angle, and that the cognitive approach to reforming *akrasia* has been attended with greater success than the conative approach. What I think that this demonstrates is that normative judgments, understood as cognitive states, have a much greater role in normative motivation than the Humean externalist can accept, and so my version of weak normative judgment internalism is the correct view.

Changing behavior without changing desires

Odysseus wished to hear the sirens sing, because their singing was said to be so beautiful that men would dash their ships upon the rocks pursuing the sirens who sang so. Knowing that his desire to pursue the sirens would be irresistible, Odysseus ordered his sailors to tie him to the mast and then to seal their own ears with wax, and not to let him loose until they were well clear of the sirens. As the story goes, Odysseus begged and pleaded and shouted for his men to untie him or to remove the wax from their ears, but they did not hear him, and followed his orders. So Odysseus changed what would have been his behavior without changing the desire to pursue the sirens. He did this by recognizing his interests, anticipating his future desiderative states, and then manipulating his environment to make the pursuit of an irresistible desire impossible so as to act in accord with his better judgment.

Examples of this combination of foresight, careful judgment, and manipulation of our future selves can be termed 'Odyssean self-control' in his honor.[88] People take similar, though less heroic measures every day. Not keeping candy bars in the house so as to avoid overindulging, not shopping for food while hungry, or seeking out a less distracting environment in which to work are all examples of Odyssean self-control.

[88] I first encountered this phrase in (Pinker), Chapter 9

Consider a more contemporary example germane to the current discussion. Ingrid is a recovering alcoholic. Let us stipulate that she is an addict who judges that it would be best not to be an addict, and so is an *akratic* rather than intemperate addict if she resumes drinking. She very much desires to drink, but of course has no trouble refraining from drinking while at work, as there is no alcohol available. Similarly, her husband helps her to ensure that she resists the temptation to keep any alcohol at home. The most direct route home from her workplace takes Ingrid by a pub where she has spent many an after-work hour drinking and socializing with her friends, some of whom she has had to break contact with because they have been insensitive to her efforts to stop drinking. She has even had her husband replace the phone numbers for these friends with the number of the local AA support line in her phone. Because she finds the temptation to stop at the pub nearly irresistible, she has stopped driving herself home from work, going as far as to sell her car and allow her driver's license to expire, replacing it with a mere government ID card. She takes the bus home, and there is no bus stop near her old pub.

The reason she goes to such heroic measures is to ensure that she would have to go to equally heroic measures to have a drink. She would have to solicit someone's cooperation which might not be forthcoming if they know she is a recovering alcoholic, and she tells everyone she knows that this is the case. She would have to call and schedule a cab or walk a long distance to get to her old pub, and both of those are actions that give her much time to reconsider or not follow through with these plans in the course of her ordinary work day. In other words, these obstacles to drinking and going to the pub allow Ingrid sufficient opportunity to attend to her meta-judgment as opposed to being in a situation in which recall bias and focus bias would have a significant causal role in her behavior. This is a good example of Odyssean self-control, and what is most notable is that it is an attempt to modify behavior not by diminishing the desire to

drink, but by Ingrid's reasoning out her likely response to environmental cues and then placing barriers in the way of encountering the cues likely to contribute to a relapse, while replacing some cues with cues likely to contribute to abstinence.

It would beg the question to say that either *of course* there is some desire not to drink in operation the whole time or that *of course* her judgment that it is best not to drink is sufficient motivation for her to work out the strategy that she has worked out. I do not deny that the Humean may be correct and that desires may hold a monopoly on motivation, but I think that Ingrid's case is one which, if taken at face value, demonstrates a form of cognitive self-manipulation. Whether there is some desire at play that counters the desire to drink or not, Ingrid's strategy is essentially one that is actuated on her ability to anticipate consequences and manipulate her surroundings to achieve results that she judges best. These are cognitive abilities. Further, her efforts are all steps that are intuitively consistent with her judgment that it is better to quit drinking, while a failure to do something to keep herself away from bars and alcohol would be intuitively inconsistent with her better judgment, opening Ingrid to the charge of subjective synchronic irrationality.

Consider only one more fabricated example. Alex judges that it would be best to quit wasting so much time playing video games. He decides to make use of the best behavior modification research available and visits the website www.stikk.com.[89] The Stikk system was born out of credible research on incentives and behavior modification, and chiefly makes use of the insight that it is more effective to give someone a reward (say, money) and then threaten to take it away if the subject doesn't complete a goal than to offer the same reward only once the goal is completed. Alex, in order to make the stikk contract, must set his goal: no more than ten

[89] This is an actual website, founded by Ian Ayres, Dean Karlan, and Jordan Goldberg, two Yale economics professors and a former Yale student, respectively.

hours of video gaming per week (hey, it's a start). Alex must then supply stakes. Most people choose to put money on the line, but the site allows a commitment contract without monetary stakes. Alex designates $10 for every hour exceeding 10 per week of video games that he plays. Of course, those at stikk do not wish to profit off of others' *akrasia*, so the disincentive for failure has an interesting twist. If Alex's credit card must be charged, the money goes to the Westboro Baptist Church, whose views and practices Alex absolutely detests. Alex then selects a referee, who keeps track of his progress. Alex's roommate, Beavo, who has been most vocal about the amount of time Alex has been wasting at video games, is the logical choice. Finally, Alex enlists several friends and family members to act in the role of supporters, whom he keeps informed of his progress and from whom he receives encouraging feedback.

If this method of behavior modification works, as the laboratory work on which the method is based would suggest,[90] then it is also an example of a form of self-manipulation that relies on the ability of the individual to predict their responses (including what their desires *will be*) in counterfactual scenarios. Would it be most accurate to say that Alex stopped playing so many video games because he hated the Westboro Baptist Church more than he loved video games? That makes some degree of sense, except that presumably Alex always hated the Westboro Baptist Church more than he loved video games, and that his hatred only mattered after he intentionally set up a system in which one was set directly opposed to the other. The cognitive anticipation of his future states is doing a great deal of motivational work. Even if the Humean is correct and desires hold a monopoly on motivation, there is at least room to pay much greater attention to cognitive states in a credible story of motivation, especially a person's normative judgments.

[90] The site, as of March 20, 2013, lists just over 300,000 workouts completed and over 2.5 million cigarettes not smoked.

As far back as Aristotle, the difference between the *akratic* and the intemperate is couched in their actions relative to their best judgment. The intemperate chooses in accord with their own best judgment (and thus are subjectively rational), but are disapproved of because their judgment runs afoul of some objective standard (and thus are called objectively irrational). This distinction has some consequences that are relevant here. The intemperate person might be persuaded to *change* their judgment, or might not, but the *akratic* is susceptible to correction of their behavior by simply having their best judgment more readily brought to their attention.[91]

Even in cases in which we do little or nothing to change desires that we have, we may change behavior. In the psychological literature, cases like the above are termed 'cognitive bias modification'. The term sounds more clinical and impressive than it really is. Actually, the kinds of strategies employed in the various forms of cognitive bias modification in the literature strongly resemble the above two examples.

Cognitive bias modification treatments have their genesis in research aimed at treating various sorts of anxiety and depression disorders. A significant part of the etiology of anxiety and depression disorders involve certain cognitive biases, and indeed these biases are common across many emotional disorders. As Matthews and MacLeod put it in their literature review:

> "Evidence has continued to show that, relative to emotionally stable individuals, those prone to emotional disorders preferentially attend to emotionally congruent cues, recall more unpleasant memories, and interpret ambiguous events in a more negative manner. The findings we have reviewed suggest that these emotional processing biases occur across emotional disorders, as perhaps might be expected

[91] Aristotle puts it "Moreover, the incontinent person is the sort to pursue excessive bodily pleasures against correct reason, but not because he is persuaded [it is best]. The intemperate person, however, is persuaded, because he is the sort of person to pursue them. Hence the incontinent person is easily persuaded out of it while the intemperate person is not." (111) (NE Book 7, Chapter 8, section 4)

> in view of their frequent comorbidity. The evidence also suggests that apparently different types of repeated negative ideation, including worry in GAD [generalized anxiety disorder] and rumination in depression, have more in common and are more similar across disorders than is sometimes supposed." (Mathews and MacLeod)

It is important to note that the cognitive biases specifically identified are focus biases[92] and recall biases, [93] which are both identified above as important causal factors in addictive behavior. Importantly, these biases disappear when emotional disorders are in remission (MacLeod and Mathews). This data has given researchers reason to wonder whether attempts to address these cognitive biases would improve clinical outcomes.

In the case of *akrasia*, the kind of cognitive bias modification that should be effective given the cognitive account of *akrasia* that I have supplied, is as follows. The key to avoiding *akrasia* is attending to one's own better judgment. The *akrates* needs some form of cognitive bias modification that has the effect of combating the focus and recall biases that crowd attention to better judgment out of the global workspace. Such approaches have commonsense support. I am not the first to propose that such cognitive approaches are effective remedies for *akrasia*. Alfred Mele, in discussing what *enkrateia* (the opposite of *akrasia*) consists of, writes:

> "An agent can, for example, keep clearly in mind, at the time of action, the reasons for doing the action which he judged best; he can refuse seriously to entertain "second thoughts" concerning matters about which he has just very carefully made up his mind; he can seek to add to his motivation for performing

[92] For more evidence concerning the causal role of focus bias in emotional disorders like anxiety and depression, see (Mineka and Sutton) and (Mathews and MacLeod, Selective processing of threat cues in anxiety states)

[93] For more evidence concerning the causal role of recall bias in many emotional and other disorders, see (Blaney).

the action judged best by promising himself a reward (e.g. an expensive dinner) for successfully resisting tempation" (Mele, Self-Control, Action, and Belief)[94]

The first part of the sentence refers to keeping better judgment in the global workspace, or as in Ingrid's case, keeping unwanted stimuli out of it. The second part of the sentence refers to strategies like Alex's, though his reward is supporter approbation while failure carries a penalty. Commonsense approaches to behavior analysis and modification are not always accurate, but where careful study confirms them, we have that much more reason to rely on such approaches.

In Chapter 2, I detailed the relationship between decisional heuritics and cognitive bias. What is worth noting is that a heuristic is a passive thing, while metajudgment is active, and requires attentional resources (i.e. space in the global workspace). It is also slower and more deliberate. Common remedies for attending to better judgment often feature a strategy of being more cognitively active than passive. Counting to ten before acting or speaking gives the agent opportunity to attend to metajudgment rather than acting out of anger or other impulse. Posting reminders to oneself where they will be seen during critical moments helps people to attend to factors that they at once consider most important and at the same time know they may neglect.

The success of some of these long-used attempts at cognitive bias modification is also observed in a more controlled setting. A recent study by Hoppitt, Matthews, Yiend, and Mackintosh examines the role of active training in cognitive bias modification (Hoppitt, Mathews and Yiend). The study is designed to reveal the effect of active (as opposed to passive) training on modifying cognitive bias.

[94] Mele's view is not a fully worked out view of the motivational role of normative judgments, but his focus on specifically cognitive "therapies" is apropos. He cites Alston, "Self-Intervention and the Structure of Motivation" *The Self: Psychological and Philosophical Issues* ed. Mischel, Oxford: Blackwell, 1977, p.77 and Brandt, *A Theory of the Good and the Right* Oxford: Clarendon, 1979, pp. 111, 126-27, 333ff.

The study takes two groups of volunteers who are not disposed to anxiety, as measured by a standardized assessment. One group is given active cognitive bias modification, while the other group is given passive cognitive bias modification. In the active training, the subjects are given a scenario that is emotionally ambiguous until the last word of the scenario. For example:

> "You have decided to go caving even though you feel nervous about being in such an enclosed space. You get to the caves before anyone else arrives. Going deep inside the first cave you realize you have completely lost your w—." (Hoppitt, Mathews and Yiend 75)

The framers of the study point out that such a scenario is emotionally ambiguous in the sense that the last word could sensibly be 'fear', but supplying the first letter of the word 'way' resolves the ambiguity. The subject is then asked if they envision themselves feeling afraid in the cave.

The passive training group is supplied with the entire passage above, complete with the final word, and the sentence 'You are feeling afraid of being in the cave' appended to the end of the original passage. Both groups are then given a filler task and then are both presented with an emotionally ambiguous passage such as:

> "You are finding that your sight is worse than it was and despite the risks you decide to try an experimental laser surgery you've read about. Afterwards as the bandages are taken off your eyes, you realize that your life will be affected radically by the results." (Hoppitt, Mathews and Yiend 75)

The point is to see if there is a difference between how the actively trained group and the passively trained group interprets the ambiguous passage. The study found a statistically significant difference in the tendency of the actively trained group versus the passively trained

group to interpret the ambiguous passage negatively. Presumably if the active training were positively valenced instead of negatively as the study write-up indicates then the active training would have increased the tendency of the active training sample to interpret the ambiguous passage positively.

Interpreting ambiguous evidence as valenced in a particular way is evidence of cognitive bias. If there were no cognitive bias present, the subject would interpret the ambiguous evidence as ambiguous. What the results of this study seem to indicate is that actively engaging the cognitive faculties to interpret data and envision one's own emotional response has an observable causal effect on future responses. Active cognitive engagement is at the heart of cognitive bias modification.

The study is carefully crafted to isolate the effect of active cognitive training, but the study interestingly confirms a great many common platitudes about behavior modification. For example, some form of "visualizing success" is a staple in self-help guides and guides to personal and professional success. The idea is that when you actively visualize yourself acting, thinking, or deciding a certain way, you become more likely to act, think, and decide in that way.

The treatment of a focus bias, especially in cases of addiction, would then have a strong effect on determining whether the addict would refrain or relapse. Most of the work in modifying focus bias is in the context of treatments for anxiety disorders. Part and parcel of the anxiety disorder is focusing unduly on negative or threatening stimuli to the exclusion of positive or non-threatening stimuli. There are two ways of measuring anxiety: trait anxiety and state anxiety. Measures of state anxiety are measures of the degree to which a person is in an anxious state. Trait anxiety is a measure of the effect of anxiety-producing stimuli. A recent review of the literature concerning attentional bias modification indicates that "Attention Bias Modification

Treatment produced a greater effect on trait than state anxiety measures. This suggests that ABMT might target the more enduring aspects of anxiety." (Hakamata, Lissek and Bar-Haim)

The message is encouraging for the treatment of *akrasia* by means of treating the cognitive biases that are implicated in the *akrates*. If the anxiety sufferer can come to diminish attentional biases that select threatening stimuli to the exclusion of positive and neutral stimuli, then it stands to reason that the addicted *akrates* like Ingrid or Alex may train him or herself to focus on more stimuli in their environments other than alcohol-related or gaming-related stimuli.

A study by Lester and others, similar to the Hoppit et al. study described above, but with a broader scope, details some strategies for cognitive bias modification designed to broadly treat anxiety and depression. What should strike the reader about their descriptions is that they are much more pedestrian in nature than the clinically impressive sounding phrase 'cognitive bias modification therapy' would suggest. It is a case in which at least some aspects of our common folk psychology have some empirical verification in a carefully controlled setting.

A sampling of the cognitive biases and their modification strategies are as follows (Lester, Mathews and Davison 300):

Cognitive Error	Definition	Clinical Example	Example Modification Item
Selective Abstraction	Focusing on a detail taken out of context, while ignoring other more salient features of the situation and conceptualizing the whole experience on the basis of this fragment	A recent graduate begins a new position and is eager to make friends with their colleagues. They ask their new colleagues whether they would like to join them for a drink after work and 2 people accept their offer. They focus on the fact that some people declined and think this means they aren't liked	You have started a new job and hope to be friends with your colleagues. At the end of your first day you ask whether people would like to go for a drink and 2 people offer to come out with you. You think this means you have probably been rejected/accepted Have you failed to make friends?

		rather than being pleased that some of their colleagues are keen to socialize.	
Dichotomous Thinking	Tendency to place all experiences in one of two opposite categories, e.g. flawless or defective rather than viewing them as existing on a continuum. In describing oneself, the extreme negative categorization is selected	You've been trying to diet but you've eaten a few sweets over the weekend. You tell yourself that you can never control yourself and that all your dieting and jogging over the whole week have gone down the drain.	You have been on a really strict diet for a few weeks and have totally cut out sweet things. However you couldn't resist a piece of cake on your friend's birthday. You think your attempts at dieting have been. futile/disciplined Have you completely failed in your attempts to diet?

Notice the overlap between the cognitive errors described in this table and cognitive errors involved in classic examples of *akrasia* discussed in Chapter 1 and throughout this work. The examples in the Lester et al. study are tailored to anxiety and depression, but consider different ways of fitting the definitions supplied.

Ingrid is at a party, and there is alcohol present, and several people near her are having an alcoholic drink. Ingrid focuses unduly on these examples and becomes anxious that everybody else is drinking, and she feels a great deal of social pressure that crowds out her resolve to stay on the wagon. Now imagine that a close friend is next to her to apply cognitive bias modification treatment. This interlocutor points out all of the people who are not drinking alcohol, and asks probing questions of Ingrid, asking whether she really believes that anyone notices or cares whether or not she has a drink. This line of questioning and pointing out of

external stimuli actively engage Ingrid's cognitive faculties and gives her a greater chance to attend to her better judgment of abstinence.[95]

Consider now Aristotle's dieter, who can be accused of dichotomous thinking with rather little modification of the above example. The dieter, though judging that it would be better to avoid the sweets than to indulge in them, recalls *akrasia* in his recent past, and considers his diet irrevocably lost. He indulges in the sweets, contravening his better judgment while making it even easier to continue indulging in the sweets. Again, an interlocutor could actively engage his cognitive faculties with probing questions about the real effectiveness of dieting and the comparative effectiveness of indulging less as opposed to more. Again, this would have the effect of not only allowing better judgment to prevail in this case, but (in accord with the evidence from the Hoppitt study) makes it more likely to prevail in similar circumstances in the near future.

The success of these strategies for cognitive bias (and therefore behavior) modification is also confirmed by Lester et al. In their words, "Cognitive Error Modification was capable of inducing systematic group differences in how hypothetical events were perceived in both a healthy and vulnerable sample." (305)

Of course, strategies for anti-*akratic* cognitive bias modification need not necessarily involve an interlocutor. Controlling one's environment (as in Odyssean self-control), setting reminders for oneself in places that they will likely be seen (being one's own interlocutor), habituating active engagement of cognition and metacognition (repetition of slogans, mottos, or

[95] Consider this from Aristotle: "For some people are like those who do not get tickled themselves if they tickle someone else first; if they see and notice something in advance, and rouse themselves and their rational calculation, they are not overcome by feelings, no matter whether something is pleasant or painful" (Nicomachean Ethics 110) (Book 7, Chapter 7, Section 8)

using the ‘count to ten’ strategy) are all examples of cognitive bias modification therapy that do not require a therapist.

I hope I have not belabored the point, but what I have been arguing is that the right way to reform the *akrates* is to focus on the cognitive aspects of the *akrates* rather than on their desires. If *akrasia* involves cognitive bias, as I argued in Chapter 2, and if the difference between being *akratic* and not being *akratic* is actuated on the modification of cognitive states, then this is good reason to believe that the cognitive account of *akrasia* is the right account. If the cognitive account is the right account, that indicates that normative judgments, understood as cognitive states, play a significant role in motivation and action. The evidence I have gone to such lengths describing is at odds with the picture of *akrasia* painted by the Humean externalist. For the Humean externalist, you can judge and cogitate all you like, but unless you have the appropriate desires, your behavior doesn’t change. The evidence indicates that cognitive states (which include normative judgments) have a much more significant role than the Humean perspective allows in motivation and action.

CONCLUSION

Akrasia has been a significant puzzle for philosophers for a long time. What I hope to have accomplished in this work is something new in the philosophy of *akrasia*. I have shown that *akrasia* as a philosophical concept has had a remarkable degree of consistency in addition to its longevity. In *akrasia*, we see a common part of human experience that baffles us. How can it be that we should be able to think that one thing is truly better than another and yet sometimes voluntarily do the other?

I have shown that recent discoveries concerning the way that people make decisions and evaluate options provide the groundwork for an explanation for why humans sometimes act intentionally against their better judgment. This explanation ought to settle the question of whether *akrasia* is possible or intelligible. Also, in providing an explanation for why we sometimes act contrary to our better judgment, I have shown why we may still hold the view that there is a necessary connection between normative judgment and motivation. This avoids the trouble of accounting for *akrasia* without accounting for why we so often *do* follow our own best judgment.

This account has implications that go beyond the immediate scope of *akrasia* in moral psychology and normative motivation. Very briefly, this account of *akrasia* relies on a notion of motivation that places a far greater emphasis on the role of cognition in motivation and action than is common in philosophy post Hume. In general, I think that philosophers ought to pay much closer attention to the role that cognition plays in motivation.

Also, my account of *akrasia* should have implications for public policy. Just for one example, consider public service announcements designed to encourage people not to smoke or to quit smoking. The vast majority of such advertisements concern the dangers of smoking, and often hyperbolize those dangers in a striking way. This kind of PSA aims only at persuading

someone to judge that they ought not smoke. A more effective public policy would consider *akrasia*. Perhaps there are many instances in which people agree that it would be best not to start or continue to smoke, but because of the situation that they are in, do not attend to such judgment. PSAs that rehearse situations in which peer pressure or impulsivity are common might be more likely to cause people to more frequently or easily attend to their best judgments and thus accomplish the goal of the public service advertisement more effectively.

I hope that these and other questions of moral and normative motivation can be informed by treatment of *akrasia* in this work.

BIBLIOGRAPHY

Alcoholics Anonymous. *The story of how many thousands of men and women have recovered from alcoholism*. 4th. Alcoholics Anonymous World Services, Inc., 2001.

Aristotle. *Nicomachean Ethics*. Trans. Terence Irwin. Indianapolis: Hackett, 1999.

Arpaly, Nomy. "On Acting Rationally against One's Best Judgment." *Ethics* 110 (2000): 488-513.

Austin, J. L. "A Plea for Excuses: The Presidential Address." *Proceedings of the Aristotelian Society* 57 (1956): 1-30.

Baars, Bernard. "The Global Brainweb: An Update on Global Workspace Theory." *Science and Consciousness Review* (2003).

Baumeister, RF, et al. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74.5 (1998): 1252-1265.

Blackburn, Simon. *Spreading the Word*. Oxford University Press, 1984.

Blaney. "Affect and Memory, A Review." *Psychological Bulletin* 99 (1986): 229-246.

Boyer and Dickerson. "Attentional bias and Addictive behavior: Automaticity in a gambling-specific modified Stroop task." *Addiction* 98 (2003): 61-70.

Bratman, Michael. "Practical Reasoning and Weakness of the Will." *Nous* 13.2 (1979): 153-171.

Bricke, John. *Mind & Morality*. Oxford: Oxford University Press, 1996.

Brink, David. "Moral Motivation." *Ethics* 108.1 (1997): 4-32.

—. *Moral Realism and the Foundations of Ethics*. Cambridge University Press, 1989.

Buss, Sarah. "Weakness of Will." *Pacific Philosophical Quarterly* 78 (1997): 13-44.

Campbell, William G. "Addiction: A Disease of Volition Caused by a Cognitive Impairment." *The Canadian Journal of Psychiatry* 48.10 (2003): 669-674.

Cosmides and Tooby. "Better than Rational: Evolutionary Psychology and the Invisible Hand." *American Economic Review* 84 (1994).

Cox, et al. "Alcohol attentional bias as a predictor of alcohol abusers' treatment outcomes." *Drug and Alcohol Dependence* 68 (2002): 237-243.

Darley, J. M. and C. D. Batson. "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior." *Journal of Personality and Social Psychology* 27 (1973): 100-108.

Davidson, Donald. "Actions, Reasons, and Causes." Davidson, Donald. *Essays on Actions and Events*. Oxford: Clarendon Press, 2001. 3-20.

Davidson, Donald. "How is Weakness of the Will Possible?" Davidson, Donald. *Essays on Actions and Events*. Oxford: Clarendon Press, 2001. 21-42.

Davidson, Donald. "Incoherence and Irrationality." Davidson, Donald. *Problems of Rationality*. Oxford: Oxford University Press, 2004. 189-198.

Davidson, Donald. "Intending." Davidson, Donald. *Essays on Actions and Events*. Oxford: Clarendon Press, 2001. 83-102.

Davidson, Donald. "Paradoxes of Irrationality." Davidson, Donald. *Problems of Rationality*. Oxford: Oxford University Press, 2004. 169-187.

Decker, Seamus A. and Jessica N. Gay. "Cognitive-bias toward gaming-related words and disinhibition in World of Warcraft gamers." *Computers in Human Behavior* 27 (2011): 798-810.

Doris, John. *Lack of Character: Personality and Moral Behavior*. Cambridge, 2002.

Frankena, William. "Hare on Moral Weakness and the Definition of Morality." *Ethics* 98.4 (1988): 779-792.
Gigerenzer, Gerd and Goldstein. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological Review* 103 (1996).
Gigerenzer, Gerd and Peter M. Todd. *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press, 1999.
Gosling, Justin. *Weakness of the Will*. New York: Routledge, 1990.
Hakamata, Yuko, et al. "Attention Bias Modification Treatment: A Meta-Analysis Toward the Establishment of Novel Treatment for Anxiety." *Biological Psychiatry* 68 (2010): 982-990.
Hardcastle, Valerie Gray. "Life at the borders: habits, addictions, and self-control." *Journal of Experimental & Theoretical Artificial Intelligence* 15 (2003): 243-253.
Hare, R. M. *Freedom and Reason*. New York: Oxford University Press, 1963.
—. *The Language of Morals*. Oxford: Oxford University Press, 1952.
Harman, Gilbert. *The Nature of Morality*. Oxford University Press, 1977.
—. "The Nonexistence of Character Traits." *Proceedings of the Arsitotelian Society* (2000).
Haselton, Martie G., Daniel Nettle and Paul W. Andrews. "The Evolution of Cognitive Bias." *The Handbook of Evolutionary Psychology*. Ed. David M. Buss. Hoboken: John Wiley & Sons, 2005. 724-746.
Hester, Dixon and Garavan. "A Consistent Attentional Bias for Drug-Related Material in Active Cocaine Users Across Word and Picture Versions of the Emotional Stroop Task." *Drug and Aclohol Dependence* 81 (2006): 251-257.
Hoppitt, Laura, et al. "Cognitive Bias Modification: The Critical Role of Active Training in Modifying Emotional Responses." *Behavior Therapy* 41 (2010): 73-81.
Isen, A. M. and P. F. Levin. "Effect of Feeling good on Helping: Cookies and Kindness." *Journal of Personality and Social Psychology* 21 (1972): 384-388.
Jackson, Frank. "Weakness of Will." *Mind* 93.369 (1984): 1-18.
Janes, et al. "Brain reactivity to smoking cues predicts ability to maintain tobacco abstinence." *Biological Psychiatry* 67 (2010): 722-729.
Kalis, Annemarie, et al. "Weakness of will, akrasia, and the neuropsychiatry of decision making: An interdisciplinary perspective." *Cognitive, Affective, & Behavioral Neuroscience* 8.4 (2008): 402-417.
Kirby, Kris and Nino Marakovic. "Modeling Myopic Decisions: Evidence for Hyperbolic Delay-Discounting within Subjects and Amounts." *Organizational Behavior and Human Decision Processes* 64.1 (1995): 22-30.
Lester, Kathryn J., et al. "Modifying cognitive errors promotes cognitive well being: A new approach to bias modification." *Journal of Behavior Therapy and Experimental Psychiatry* 42 (2011): 298-308.
Lewis, David. "Desire as Belief." *Mind* (1988): 323-332.
—. "Desire as Belief II." *Mind* (1996): 303-313.
Liu, Shijing, et al. "Relationship between attentional bias to cocaine-related stimuli and impulsivity in cocaine-dependent subjects." *The American Journal of Drug and Alcohol Abuse* 37 (2011): 117-122.
MacLeod, Colin and Andrew Mathews. "Cognitive Experimental Approaches to the Emotional Disorders." *Handbook of Behavior Therapy and Psychological Science*. Ed. Martin. New York: Pergamon Press, 1991. 116-150.

Mathews, Andrew and Colin MacLeod. "Cognitive Vulnerability to Emotional Disorders." *Annual Review of Clinical Psychology* 1 (2005): 167-195.
—. "Selective processing of threat cues in anxiety states." *Behavior Research and Therapy* 23 (1985): 563-569.
Mathews, K. E. and L. K. Cannon. "Environmental Noise Level as a Determinant of Helping Behavior." *Journal of Personality and Social Psychology* 32 (1975): 571-577.
Mazas, Finn and Steinmetz. "Decision-Making Biases, Antisocial Personality, and Early-Onset Alcoholism." *Alcoholism: Clinical and Experimental Research* 24 (n.d.): 1036-1040.
McClure, SM, et al. "Separate neural systems value immediate and delayed monetary rewards." *Science* 306 (n.d.): 503-507.
McDowell, John. "Virtue and Reason." *The Monist* 62.3 (1979): 331-350.
McGuire, M. C. "Can I do what I think I ought not? Where has Hare gone wrong?" *Mind* 70.279 (1961): 400-404.
Mele, Alfred. *Irrationality*. New York: Oxford University Press, 1987.
—. "Is Akratic Action Unfree?" *Philosophy and Phenomenological Research* 46.4 (1986): 673-679.
—. "Self-Control, Action, and Belief." *American Philosophical Quarterly* 22.2 (1985): 169-176.
Mineka, Susan and Steven K. Sutton. "Cognitive Biases and the Emotional Disorders." *Psychological Science* 3.1 (1992): 65-69.
Nagel, Thomas. *The Possibility of Altruism*. Princeton University Press, 1979.
Nijs, et al. "Differences in attention to food and food intake between overweight/obese females and normal-weight females under conditions of hunger and satiety." *Appetite* 54 (2010): 243-254.
Pears, David. *Motivated Irrationality*. Oxford: Clarendon Press, 1984.
Pinker, Stephen. *The Better Angels of Our Nature*. Penguin, 2011.
Plato. "Protagoras." Plato. *Plato*. Ed. Scott Buchanan. Trans. Benjamin Jowett. New York: Penguin, 1948. 45-120.
Price, Richard. *A Review Of The Principal Questions In Morals*. Kessinger Publishing, 1787.
Schenk. "Exploiting the Salience Bias in Designing Taxes." *Yale Journal on Regulation* 28 (2011): 253-311.
Schueler, G. F. "Akrasia Revisited." *Mind* 92.368 (1983): 580-584.
Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford: Clarendon Press, 2003.
Smith, Michael. "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion." *Weakness of Will and Practical Irrationality*. Ed. Sarah Stroud and Christine Tappolet. Oxford: Clarendon Press, 2003. 17-38.
—. *The Moral Problem*. Blackwell, 1994.
Stevenson, Charles Leslie. "The Emotive Meaning of Ethical Terms." *Mind* 46.181 (1937): 14-31.
Stocker, Michael. "Desiring the Bad: An Essay in Moral Psychology." *Journal of Philosophy* 76 (1979): 738-753.
Stroud, Sarah and Christine Tappolet, *Weakness of Will and Practical Irrationality*. Oxford: Clarendon Press, 2003.
Stroud, Sarah. "Moral Overridingness and Moral Theory." *Pacific Philosophical Quarterly* 79.2 (1998): 170-189.

Stroud, Sarah. "Weakness of Will and Practical Judgment." *Weakness of Will and Practical Irrationality*. Ed. Sarah Stroud and Christine Tappolet. Oxford: Clarendon Press, 2003. 121-146.
Tappolet, Christine. "Emotions and the Intelligibility of Akratic Action." *Weakness of Will and Practical Irrationality*. Ed. Sarah Stroud and Christine Tappolet. Oxford: Clarendon Press, 2003. 97-120.
Taylor, C. C. W. "Reply to Schueler on Akrasia." *Mind* 93.372 (1984): 584-586.
Tenenbaum, Sergio. "Accidie, Evaluation, and Motivation." *Weakness of Will and Practical Irrationality*. Ed. Sarah Stroud and Christine Tappolet. Oxford: Clarendon Press, 2003. 147-171.
—. "The Judgment of a Weak Will." *Philosophy and Phenomenological Research* 59.4 (1999): 875-911.
Thalberg, Irving. "Acting against one's better judgment." *Weakness of Will*. Ed. G. W. Mortimore. London: Macmillan, 1971. 233-246.
Tobon, Juliana I., Allison J. Ouimet and David J. A. Dozois. "Attentional Bias in Anxiety Disorders Following Cognitive Behavioral Treatment." *Journal of Cognitive Psychotherapy* 25.2 (2011): 114-129.
Trout, J. D. "Paternalism and Cognitive Bias." *Law and Philosophy* 24 (2005): 393-434.
Tversky and Kahneman. "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." *Psychological Review* 90 (1983): 293-315.
—. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185 (1974): 1131-1141.
Vadhan, et al. "Attentional Bias towards Cocaine Related Stimuli: Relationship to Treatment-Seeking for Cocaine Dependence." *American Journal of Drug and Alcohol Abuse* 33.5 (2007): 727-736.
Viens, A. M. "Addiction, Responsibility, and Moral Psychology." *The American Journal of Bioethics* 7.1 (2007): 17-19.
Vranas, P. B. M. "The Indeterminacy Paradox." *Nous* (2005).
Walker, Arthur F. "The Problem of Weakness of Will." *Nous* 23.5 (1989): 653-676.
Wallach, Wendell and Collin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press, 2009.
Watson, Gary. "Skepticism About Weakness of Will." *The Philosophical Review* 86.3 (1977): 316-339.
Watson, Gary. "The Work of the Will." *Weakness of Will and Practical Irrationality*. Ed. Sarah Stroud and Christine Tappolet. Oxford: Clarendon Press, 2003. 172-200.
Wedgwood, Ralph. "Choosing Rationally and Choosing Correctly." *Weakness of Will and Practical Irrationality*. Ed. Sarah Stroud and Christine Tappolet. Oxford: Clarendon Press, 2003. 201-229.
—. *The Nature of Normativity*. Oxford: Clarendon Press, 2007.
Wiland, Eric. "Trusting Advice and Weakness of Will." *Social Theory and Practice* 30.3 (2004): 371-389.
Williams, Bernard. "Internal and External Reasons." Williams, Bernard. *Moral Luck*. 1981. 101-113.