

1-1-2015

Weakness of Will: An Inquiry on Value

Michael Funke

University of South Florida, mbfunke@gmail.com

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [Ethics and Political Philosophy Commons](#)

Scholar Commons Citation

Funke, Michael, "Weakness of Will: An Inquiry on Value" (2015). *Graduate Theses and Dissertations*.
<http://scholarcommons.usf.edu/etd/5847>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Weakness of Will: An Inquiry on Value

by

Michael B. Funke

A dissertation presented in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Philosophy
College of Arts and Sciences
University of South Florida

Major Professor: Hugh LaFollette, PhD
Colin Heydt, PhD
Alex Levine, PhD
Stephen Turner, PhD
Brook Sadler, PhD

Date of Approval:
November 19th, 2014

Keywords: Action, Addiction, Akrasia, Compulsion, Ethics, Moral Psychology,
Motivation, Practical Rationality, Self Control, Strength of Will, Weakness of Will,
Willpower.

Copyright © 2014 Michael B. Funke

Dedication

For my mother.

Acknowledgments

I owe by far the largest debt of gratitude to Rhiannon for her support and love during the trying moments of this project. And, thank you to my friends Jarrett, JJ, Erik, Jim, Nigel and Brad for their conversation and confidence.

Thank you also to Hugh LaFollette who has been a great help over a long haul.

Table of Contents

Introduction	1
1. Willpower as Skill	6
1.1 Will as Brain	6
1.2 The Muscle Metaphor	10
1.3 Problems With the Muscle Metaphor: Regional Control	15
1.4 Problems with the Muscle Metaphor: Desire	16
1.5 Problems with the Muscle Metaphor: Strategic Success	21
1.6 Is Willpower Just Sugar?	23
1.7 The Skill Model of Willpower	24
2. Willpower Strategy	29
2.1 Distraction	30
2.2 Mindfulness	35
2.3 Avoidance	40
2.4 Precommitment Strategies	45
2.5 Self-Management, Self-Control or Willpower	48
2.6 Conclusion	54
3. Weakness of Will	56
3.1 Complex vs. Reductive Definitions	57
3.2 Defining Weakness of Will	62
3.3 Weakness of Will: The Ordinary Use	70
3.4 Addiction Research and Weak-Will	80
3.5 Conclusion	85
4. Freedom and Weakness of Will	86
4.1 The Consequences Argument	86
4.2 Skepticism About Freewill	88
4.3 Two Theories About Freewill	89
4.4 Libertarian (Incompatibalist) Freewill	89
4.5 Compatibilism: Determinism and Freedom	98
4.6 Semi-Compatibilism	108
4.7 Conclusion	121
5. Weak-Will and Recklessness	123
5.1 Practical vs. Theoretical Judgment	124
5.2 Willing and The Fact/Value Distinction	128
5.3 Strong Internalism: Logical Impossibility of Weakness of Will	136
5.4 Moderate Internalism: The possibility of weakness of will	137
5.6 Weak Internalism: The possibility of strict weakness of will	140

5.7	Watson's Skepticism About Weakness of Will	142
5.8	In Defense of Commonsense	144
5.9	Conclusion	147
6.	The Value of Willpower and Weakness of Will	149
6.1	Why We Need Willpower	150
6.2	Willpower Gone Wrong	153
6.3	The Practical Value of Weak-Will	157
6.4	The A Priori Value of Weak-Will	159
6.5	Weakness of Will as a Signifier	162
6.6	Weakness and Self-Compassion	165
6.7	Weakness of Will and Moral Licensing	171
6.8	Conclusion	180
	Conclusion	181
	Bibliography	185

List of Figures

Figure 1:	Chart of Willpower Strategies	30
-----------	-------------------------------	----

Abstract

One dominant scientific view holds that willpower is a type of muscle which can be weakened through use in the short term and strengthened through use over time. However, evidence from neuroscience, social psychology and behavioral economics suggest that willpower is regional, subverted through desire and strengthened by strategy—these are features a muscular account would not predict. It is better to think about willpower as a skill with a physiological component. Willpower strategies extend the brute effort of self-control through the use of reason and have the practical effect of increasing self-regulation. Willpower is "worth wanting" because there is a gap in our given desires and our evaluations. In general willpower is the skill responsible for extending the motivational force of evaluations to overcome the motivational force of other interests. Of course, willpower can be used in the service of evil, but in general it is a power we would prefer to have.

Interestingly, not all cases of weakness of will are, on balance, bad. As a practical matter weakness of will is a crucial element of developing willpower skills over time. Just as a skilled batter relies on failures to teach what is required for good hitting, willpower failures are an important element in developing habits for success. Additionally, the motivational failure of evaluation built in to weakness of will requires a commitment to practical claim that one can choose how to act in ways not dictated by given desires. This commitment to the importance and viability of evaluation is a crucial component of having a moral perspective in a natural system

and weakness of will is a signifier of this foundational element of a practical perspective.

Introduction

There is a huge gap between the world as it is and the world as it ought to be and they are both important. This dissertation is an effort to do philosophy in keeping with Wilfred Sellars' "synoptic vision" of orienting scientific knowledge and philosophical visions into a coherent arrangement sufficiently resonant with the linguistic community to support the practical effect of helping people to live well.¹ The world as it is cannot be explained by the way we happen to see the world, but we cannot abandon our own perspective in our pursuit of knowledge about the world. Since ancient Greece the aim of philosophy has been to extend knowledge in the natural sciences in the service of learning how to live well. Working out how the sciences make a difference in human lives remains the basic role of philosophy. Philosophy is a naturalistic pursuit in that it seeks to move human understanding closer to the truths of the natural sciences. Philosophy is also a practical matter because it seeks to answer the question "how should we live?"

This dissertation is about the gap between our judgment and motivation: it is about willpower and weakness of will. Weakness of will is common but it is poorly understood. My dissertation is an effort to respond to the issues of practical rationality raised by weakness of will in a way that is informed by modern scientific literature and in keeping with the philosophical tradition. On my view the account offered here brings together the central philosophical questions and recent

¹ Sellars, Wilfred (1962) "Philosophy and the Scientific Image of Man," in *Frontiers of Science and Philosophy*, Robert Colodny (ed.) (Pittsburgh, PA: University of Pittsburgh Press): 35–78

science in a way that is relatively consonant with popular usage. My central thesis is that willpower is best understood as a skill and this leads to some surprising insights about how willpower works strategically and how we might find value in willpower failures.

The leading account of willpower is based on the muscle metaphor--on this view willpower is like a muscle in that it is strengthened through use over time, but depleted through use in the short run. The muscle metaphor is right in some important respects, but overall its strengths are not enough to make up for its weaknesses. I discuss three central weaknesses of the muscle metaphor beginning with its advocates claims that willpower is global. Here I present evidence that suggests willpower operates in a regional way suggesting an alternative view. Further, evidence about the ways that willpower fails through appeal or desire, as opposed to failure resulting from outside pressure, suggests that willpower may not work in the way a muscle would. A third criticism of the dominant account of willpower is evidence that some of the strongest apparent cases of willpower success are rooted in strategies that seem to belie the sense of muscular strength.

In light of the weaknesses of the dominant account of willpower I argue that we should look for an alternative and suggest the view of willpower as a skill. The skill account is able to accommodate the empirical results of willpower depletion often cited by opponents because skills require physiological resources which can be depleted. However, a skill based account has the advantage of predicting regional strengths and weaknesses. A skill based account is also able to accommodate findings that the strongest cases of willpower are not cases of brute

effort but involve willpower strategies. I conclude that willpower is best thought of as a skill with a physiological component.

In the second chapter I offer an account of four distinct willpower strategies that augment a brute force account, including long-term, short-term, attention-based and avoidance-based strategies. Each variety of willpower strategy has distinctive advantages and disadvantages, but the focus here is on outlining some options to serve as markers for the intellectual landscape. There is some discussion here of the tendency to distinguish between self-management strategies and willpower strategies. This distinction seems based on accounts of action and identity that are somewhat restrictive in scope. I argue that action conceived of broadly is better situated to incorporate human character as developed over time and that such an account is more in keeping with willpower strategies that incorporate findings of modern social psychology and behavioral economics.

Willpower is necessary because motivation and judgment can come apart. When we are more motivated by given desires or inclinations than we are by our evaluations of those states, we are prone to willpower failures. In chapter three I argue that willpower failures take a number of forms and that we should not reduce weakness of will to any single variety. In spite of this generally expansive account I take care to distinguish weakness of will proper from an "inverted commas" account. The expression 'weakness of will' is sometimes used as an excusing condition, but this way of speaking is not in keeping with either the philosophical tradition or the other varieties of weakness considered here.

In the fourth chapter I outline the main arguments in favor of the possibility of freewill and skepticism about freewill. To begin with I explore the reasons for skepticism about free will and argue that those who are skeptical about freewill have a bigger worry than the particular trouble of explaining weak-will. I go on to describe a leading version of libertarian incompatibilism and explain how weakness of will is commonly thought to be compatible with this view. I also explain how weakness of will is thought to be compatible with compatibilist and semi-compatibilist positions on the freewill debate. What I never do is to take a stance on the freewill debate, that would be outside the scope of this project.

In the fifth chapter I turn to the central issues of motivation and judgment; particularly the ways that they come apart. I frame the discussion of weakness of will as on a continuum between compulsion and recklessness. The reckless agent, like Aristotle's Brute, does not make evaluative judgments he acts from appetitive desires without reflection. The compulsive agent may have some contrary judgments, but does not have the motivational capacity to act in line with judgment. Weakness of will happens when one is both making judgments about how to act and free to act in line with judgment, but does not do so. The debate between Mele and Holton has made clear that weakness of will can involve synchronic violations of judgment and violations of judgment that occur over time. The difference between types of weakness of will is important, but not as important as what they share. Consider by analogy the difference between types of substance dependence; addiction to nicotine is radically different from addiction to morphine in terms of outcomes, psychology and physiology. Nevertheless, there is good reason to keep

cases of substance abuse clustered together and a similar case can be made for clustering cases of weak-will.

One final consideration is that the ordinary account of weak-will suggests that it is entirely objectionable, but in fact there is something valuable to take away from weakness of will. Since being weak-willed means at least making judgments about how to behave the akratic is at least participating in the basic normative work of practical rationality. And, because the skill account of willpower allows that developing willpower is a practice developed over time, there is good reason to think that being akratic today is a necessary part of being strong-willed tomorrow. But, at the very least, we can say that the weak-willed agent is at least not acting recklessly. This should mitigate feelings of guilt or shame and should be seen as an important step in character formation.

Chapter 1

Willpower as Skill

There are several popular models which seek to describe the nature of willpower. The first part of this chapter focuses on the most popular of these models, the muscle metaphor for willpower. After some discussion I argue that there are serious limitations to the muscle metaphor. My strategy is not to wholly reject the view, but to say that it tells, at most, part of the story — and not the most important part at that. Willpower is best thought of as a skill. As Al Mele puts it in his recent *Backsliding* "In normal agents the capacity for self-control is not a mental analogue of brute physical strength...Our powers of self-control include a variety of skills and considerable savvy about which skills to use in particular situations."² Unfortunately, Mele never supports or explains his claim. I argue here that while the exercise of any skill involves at least some muscular action, the muscular element of willpower is built into the notion of a skill. Just as physical human strength is the skilled use of one's muscles, frame, etc. strength of will a skill which calls upon many faculties.

The Will and the Brain

We should avoid the mistake of positing some mysterious occult power called "the will" which resembles an impenetrable black box of judgment or decision making. The will is much less mysterious than all that. For instance, it seems that

² Al Mele, (2013) *Backsliding* pp. 93-94

efforts of will can be hugely affected by trauma to the prefrontal cortex. The seminal case of trauma to the prefrontal cortex occurred in 1848 when an accidental explosion sent a thirteen pound tamping iron through the frontal lobe of a railway foreman by the name of Phineas Gage. By all accounts before the accident Gage was a steadfast man with good impulse control, was not prone to becoming upset, and was an excellent employee in his leadership position with the railroad. However, after the accident Gage's primary physician, Dr. Harlow, describes Gages as "fitful, irreverent...capricious and vacillating, devising many plans of future operations, which are no sooner arranged than they are abandoned in turn for others appearing more feasible."³ Although Gage was still physically capable of performing his work with railroad Gage had lost the demeanor and impulse control required to perform his former role. Gage's case is of great interest because it shows that the prefrontal cortex plays an important role in affecting will power and trauma to this region of the brain has an extraordinary impact on the human volitional system.

Of course, massive trauma to the brain's prefrontal cortex is not the only way elicit measurable effects on human volition. Much less violent and more temporary measures also produce significant results in brain functioning. For instance, in a study measuring the relapse rates of recovering addicts it was found that one extra hour of sleep each night, shifting from seven to eight hours, was associated with significantly stronger resistance to relapse.⁴ The reason for this correlation, between

³ Harlow, John Martyn (1868). "Recovery from the passage of a bar through the head.." *Publications of the Massachusetts Medical Society* 2:327–347 [Republished in MacMillan, M (2000) "Restoring Phineas Gage: A 150th retrospective". *Journal of the History of the Neurosciences* 9 (1) 42-62.]

⁴ Britton et. al. (2010). "The contribution of mindfulness practice to a multicomponent behavioral sleep intervention following substance abuse treatment." *Substance Abuse*

additional sleep and increased will power, is explainable by comparing functional MRI scans of sleep deprived, people getting less than 6 hours, and a control group with seven or more hours sleep each night. It turns out that sleep deprivation impairs cognitive functioning especially in the prefrontal cortex and leaves some mid-brain regions over-stimulated. Because the prefrontal cortex is responsible for self-control and long range planning and the midbrain is associated more with basic impulses and instincts, the sleep deprived person is less capable of accomplishing high will power tasks.⁵

Other research involving the use of functional MRI brain imaging has continued to locate significant elements of the human volitional system reliably in the brain's prefrontal cortex. Some analyses further suggest that there is a strong correlation between specific brain areas and distinct types of will power. For instance, the left side of the prefrontal cortex shows the most significant activity when an individual is tasked with an activity which requires taking or continuing an action that is found to be difficult such as writing an essay or exercising. On the other hand, the right side of the prefrontal cortex is most active in resistance to temptation such as refusing dessert after dinner. The middle section of the frontal lobe, the ventromedial area, focuses on evaluating choices and making long range plans such as how to best achieve health.⁶ These imaging studies reveal that there is an important role of the brain and especially of the prefrontal cortex in common instances of willing. But they also suggest that it is a mistake to think about willing or

⁵ (Yoo, Gujar, Jolesz & Walker 2007)

⁶ Basten, Ulrike, et. al. (2010) "How the brain integrates costs and benefits during decision making." *Proceedings of the National Academy of Sciences* **107** (50) 217-67

will power as if it was a monolithic block. Rather, it seems that there is evidence in the brain for the conclusion that willing or will power occurs in several distinct ways associated with distinct regions of the brain. However, in broad terms it does seem that the brain is of two minds in cases of willpower challenges—the impulsive middle brain pushes us toward immediate desires and the calculating prefrontal cortex evaluates more distant interests.

In some ways the question of whether one will be weak-willed or strong-willed in any given situation comes down to a matter of which mind is currently ruling the brain. Another way of putting this would be to say that the difference between strength of will and weakness of will is a matter of identity, dependent upon which person one is at the time. When we are tired, hungry, drunk, preoccupied, or otherwise stressed it becomes increasingly likely that we will act in ways that we would ordinarily find unacceptable. In such cases our empathic friends often excuse our behavior, saying “He’s not himself today.” On the other hand our less sensitive friends may remark, “In vino veritas” suggesting that our ordinary demeanor is merely a mask hiding our true selves. The tension here appears to be between the impulsive middle brained self and the more regulated self motivated by the prefrontal cortex.

It is tempting to describe any action caused by the prefrontal cortex as strong willed and to describe any action caused by the middle brain as weak-willed, but that would be a mistake. Imagine a man standing before a firing squad waiting for his execution. The commanding officer offers the condemned man one last cigarette before he is shot dead. The man thinks about it briefly and responds “No thanks, I

am trying to quit.” If we suppose the condemned man is being sincere his statement is almost certainly a product of his prefrontal cortex and opposed to his impulsive self. It is also clear that the condemned man’s reasoning is seriously flawed—he no longer has a long range interest in not smoking, because he no longer has long range interests. The point is that strength of will is not simply resisting one’s immediate impulses; those impulses are quite important to our functioning as we should.⁷ Rather, strength of will is acting in accord with an appropriate weighing of one’s conflicting interests. Will power, like character is an achievement, not a given. If we can merge the different aspects of ourselves into a unified whole, then we have character and more likely, will-power.

The Muscle Metaphor of Willpower

According to one influential view willpower is analogous to muscular strength. On this view we can think about people who are strong willed as having similar features to those who are physically strong. Some people are naturally stronger than others, but exceptionally strong people have typically done a tremendous amount of work to strengthen their muscles. Sometimes this effort is designed to strengthen the body, but often it is the byproduct of some other activity which taxes one’s physical strength. In the short term such taxing activities have the effect of weakening one’s muscles, but over time the effect is that one develops muscular endurance and increased strength. Similarly, it has long been supposed that the willpower required for self-control is the kind of thing which can be depleted through use in the short term, while being strengthened with use in the long term. Since

⁷ For example people with damage to disgust or fear centers in the brain may sexually proposition family members or engage in wildly risky behavior.

1998, Roy Baumeister and colleagues have produced a steady stream of empirical data that seems to support many features of the muscle metaphor of willpower.

Baumeister's first study in the area of self-regulation aimed at establishing a phenomenon he described as "regulatory depletion," that is a relationship between engaging in a difficult task of self control and the weakening of self-regulation in subsequent tasks.⁸ In the first study researchers told subjects that they were involved in research on muscular strength. Subjects were then instructed to squeeze shut a grip strength spring, holding a piece of paper between the handles. After the initial evaluation subjects were divided into three groups and instructed to watch a movie. The groups included an increased emotional response group, a decreased response group and a control. Subjects in the manipulated groups were told to try to either increase or decrease their emotional response to the film and instructed that they should also increase or decrease the emotional response of their facial expressions. After completing the task of emotional self-regulation participants were again timed while holding a piece of paper between the handles of a commercially available grip developer. The control group averaged nearly an identical time as before watching the film, average time decreasing from 60.09 seconds to 58.52 seconds, only a 1.57 second difference. However, both groups asked to regulate affect showed a significant reduction in willingness to exert effort at the task. The group asked to decrease affective response for the film fell from an average time of 70.74 seconds to 52.25 seconds, a decrease of 18.49 seconds. More starkly the increased emotional response group began with an average time of

⁸ Roy Baumeister, et. al. (1998) "Self-Control as Limited Resource: Regulatory Depletion Patterns" *Journal of Personality and Social Psychology* Vol. 74, No. 3, 774-789

78.73 and fell to an average of 53.63, a reduction of 25.10 seconds. In a second study conducted to confirm these earlier results subjects were told about a white bear, some subjects were then told not to think about the white bear, while others were given no such instruction. After six minutes the experimenter returned and under the guise of a cover story had participants work on unsolvable anagram puzzles until they wanted to quit. Those who suppressed thoughts about white bears averaged 563 seconds trying to solve the puzzle, while those with no such instruction averaged 758 seconds.

In both studies participants were tasked with willpower intensive activities including controlling their thoughts and emotional states. In both cases these high willpower activities were followed with a subsequent, but unrelated task of will power. Subjects who had been previously engaged in self-control demonstrated less willpower on the subsequent tasks of working on a difficult puzzle or squeezing a grip strengthener. These results provide initial support for two theses; first they seem to show that our powers of self-regulation are depleted by exertion. The subjects seem to have exhausted their resources of self-control on the earlier task and had less willpower left for the next task as compared with untaxed control groups. And, second, these studies seem to show that different kinds of self-regulation draw on the same capacity. Both conclusions seem to illuminate the nature of willpower, and give us reason to adopt the muscle metaphor of willpower.

For instance, this has an empirical leg up on one common model of will as exemplifying sustained “focus” or “attention.” On this model, one would expect that simultaneous activities of willing to be mutually frustrating. But, on the attention

model, once someone has stopped one task and turned her attention to another, there should be little or no residual depletion of focus. It likewise has more explanatory power than more Platonic views that treat will as having the right structure of knowledge. On that account, we would expect earlier acts of willpower to prime or activate the relevant structure to produce increased will-power on subsequent tasks. To the contrary, on the muscle metaphor we would predict that earlier self-control would fatigue the relevant capacity and decrease expectations on subsequent tasks. This is precisely what the evidence shows. Finally, the muscle metaphor correctly predicts that just as short term depletion of muscular strength produces long-term muscular gains, short term depletion of strength of will should also produce strength the will long-term.⁹ Again, this is what the evidence shows.

In a series of studies on self-regulation exercises Oaten and Cheng have shown that the long term training of self-control has the effect of increasing willpower stamina and reducing fatigue following short term exertion of willpower. In each of the studies subjects were given a visual tracking task (VTT) in which they were to follow three identical black boxes in field of six, looking away for even a moment would cause participants to lose track of “their” boxes. This task was made more challenging by the fact that a distracting comedy reel played on a nearby screen. The VTT was performed a total of two times at the initial meeting, once as an initial baseline, and once again five minutes after a depletion exercise (not thinking about a white bear.) Participants were then assigned to a variety of training regimes, for

⁹ A long-term increase through use would also be predicted on the skill model of self-control.

instance one group enrolled in a two month physical exercise program¹⁰, another group enrolled in a four month financial monitoring program¹¹, yet another group participated in studying exercises.¹² Each group showed significant gains in the area of their training program, but according to the study hypothesis these long term training regimens should also exercise willpower and accordingly reduce fatigue or willpower depletion. This hypothesis was confirmed in each study when after training participants returned to the visual tracking task. Again participants engaged the VTT once as a baseline and then again following six minutes of not thinking about a white bear. After the long term training regimen participants were substantially more successful in self-regulation following tasks of self-control. This strongly suggests that strength of will can be improved through a consistent training regimen.

Baumeister's claim is that the results of these studies (those of Oaten and Cheng, Baumeister et. al., and others) also indicate that there is unified reserve of willpower which is exercised in a variety of self-regulation tasks. For instance, strengthening the will in activities of financial regulation helps to strengthen willpower against depletion on the white bear task and regulating one's emotional response to a movie dramatically reduces one's willpower in a grip strength exercise. The evidence here points toward the idea that strength of will is not sub-divided by task. That is, we do not have strong will about doing the dishes and weak will about chocolate cake, nor does one have strong will at work and weak will at home.

¹⁰ Oaten, M. & Cheng, K. (2006) "Longitudinal gains in self-control from regular physical exercise."

¹¹ Oaten, M. & Cheng, K. (2007) "Improvements in self-control from financial monitoring."

¹² Oaten, M. & Cheng, K. (2006) "Academic study improves regulatory strength." *Basic and Applied Social Psychology*.

Rather, willpower seems to be a single pool of mental energy from which we draw in the performance of various tasks of self-regulation. Of course we often succumb to temptations at home which we would resist at work, and manage to do the dishes only to find ourselves eating chocolate cake before bed, but these differences can be explained in part by the fact that willpower is a limited resource—we choose which areas of life will receive the greater part of willpower.

Problems With The Muscle Metaphor: Regional Control

Mele argues against the muscle metaphor on the grounds that "self control may be either regional or global, and it comes in degrees."¹³ Regional self-control is troublesome for a muscle-like conception of willpower because it seems as though a muscle can be applied toward different tasks. This is one of the central reasons Baumeister offers in favor of his account. If it turns out that willpower is regularly a regional phenomenon, one of the supports is removed from Baumeister's argument. And, on the other hand, finding that willpower is regional would support the idea that it is a skill specific to certain types of behavior or choice. The trouble at this point is that Mele offers little support for his claim; he attributes this view to A.O. Rorty, for whom it seems to be an element of Aristotelean psychology.¹⁴ As much as one might appreciate Aristotle's sense of human psychology he would likely want more of us were he working today and many parts of his science should absolutely be shelved.

While we have limited data Mele is currently working with a group conducting experimental research on the subject. But there is some reason to believe that self-

¹³ Mele (2012)

¹⁴ A.O. Rorty (1980) "Akrasia and Conflict" *Inquiry* 22:193-212

control is in fact distributed in regionally distinct ways based on animal research. I hasten to add that there is tremendous difficulty extrapolating from rats to humans, but some research on rats suggests that the mechanisms for immediate impulse control and those for long term planning are distinct. Rats were exposed to food towers which would dispense food if the rat put its nose in the bottom and press a button. A second tower would dispense five pellets instead of the traditional one, but it came with a brief delay. Researchers noted which rats were able to delay gratification in exchange for the increased reward. Other rats were then taught to cancel the action of pushing the button by refusing to give the pellet if the button was pushed after a tone was sounded. Some rats were very good at canceling the action, other continued to push the button in many cases despite indication that it would cause harm. Researchers examined the neural circuitry involved in these two kinds of self-control and preliminarily report that the mechanisms are distinct.¹⁵

Problems With The Muscle Metaphor: Failure through desire

Joel Feinberg argues against the metaphor of willpower as a muscle because he thinks that this metaphor locks one into the idea that failures of self-control are the product of literally irresistible desires. It is easy to see why such a concern might arise. After all, if willpower is a limited resource and is depleted through tasks of self-control, eventually one might literally have no willpower left. Feinberg writes, “The psychological situation is never—or hardly ever—like that of the man who dives

¹⁵ Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature neuroscience*, 8(11), 1481-1489.

Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Current opinion in neurobiology*, 6(2), 228-236.

from a sinking ship in the middle of the ocean and swims until he is exhausted and then drowns. Human endurance puts a severe limit on how long one can stay afloat in ocean; but there is no comparable limit to our ability to resist temptation.”¹⁶

Feinberg’s point is that in matters of willpower there is always something more that one could do, some amount of effort that would have proved successful beyond the point at which one’s self-control actually fails. That is, Feinberg believes that in almost all cases where willpower gives out one would have been successful, if only one had tried harder.

In part Feinberg is right, we often succumb to temptation in cases where we did have adequate willpower to resist and we are only occasionally in a state so depleted of strength that we could not resist an impulse. Consider for example two cases of an addict, Jill, trying to quit using heroin. In the first scenario Jill has been clean for nearly a day, she spent the majority of the day at work focusing on a monotonous and boring word processing task and in the early hours of the morning she has not yet been able to sleep due to withdrawal, add to this that Jill has not eaten a good meal all day. Jill’s mental reserves are seriously depleted and she breaks down. Jill scores a fix from Hank, her dealer, and returns home to shoot up in her apartment. Now imagine the exact same scenario, but imagine that a police officer was at Hank’s house asking questions. Would Jill still have tried to score from Hank, or, could she have resisted until the police had left? In all but the most

¹⁶ Joel Feinberg (1970) “What is so special about mental illness” in *Doing and Deserving* (Princeton University Press) pp. 282-3

extraordinary cases¹⁷ addicts can and do resist their urges when the addict recognizes a clearly hostile police presence. This fact suggests that even those with seriously compromised willpower are left with adequate strength of will to respond positively to large incentives.

It turns out that strength of will is depleted through use, but it is not often completely depleted. Baumeister compares this to muscular conservation of energy, pointing out that the muscle does not disappear completely, it simply weakens. On an alternative to Baumeister's view there remains an adequate amount of willpower for most tasks, but the desire to exercise that strength of will is supplanted by other alternative interests like giving in to temptation. As Gary Watson puts this point, "motivational obstacles work in part not by defeating one's best efforts, but by diverting one from effective resistance."¹⁸ Watson's point is that willpower is not defeated by brute force, but rather a conflicting appeal. This makes intuitive sense because strength of will is only possible when one is actively choosing between alternatives. There is no need for willpower unless there are conflicting choices. This is one place where the muscle metaphor of willpower breaks down. Muscular strength is often defeated by the brute force of external pressures. For example, when packing my office I know to use only small boxes for books not because I won't want to lift large boxes of books, but because I simply will not be able to do so. The weight of a large box of books is simply overwhelming. On the other hand, I know that if I want cheesecake it would be best to buy only one slice, rather than an

¹⁷ William James (1890,1950, 2:53) gives an account of addicts who reported that canon fire would not keep him from a keg of rum and another who drank the alcohol from six morbid specimen jars.

¹⁸ Gary Watson (2004) "Disordered Appetites" in *Agency and Answerability* (OUP) p. 65

entire cake. This is not because the entire cake will force itself on me in a way that I could not resist; rather, I will want to eat more than one slice. It is typically internal motivations that defeat willpower, not the external pressures that defeat muscles.

In opposition to a Watson-style account of willpower failures based on internal features, John Doris has marshaled a significant body of empirical research to the effect that situational pressures, not internal character traits, drive human action.¹⁹ On Doris's view character is so highly transitory and localized that he suggests character devolves into responsiveness to situational pressures. Doris recounts prominent results from social psychology including Milgram's results on obedience to authority, Zimbardo's Stanford prison experiments, and several studies on an effect known as the diffusion of responsibility. These action-guiding psychological phenomena represent in part situational pressures that are clearly capable of leading the vast majority of people to engage in morally reprehensible behavior. Milgram showed that ordinary people no different than you or I would shock other human beings to death simply because they were asked to by a researcher.²⁰ Zimbardo asked college students who had been screened for psychological stability to play prisoner and guards for two weeks, but the experiment was canceled after only a few days because the situation became too dangerous.²¹ Likewise, the diffusion of responsibility seems to account for a large number of people standing by passively while Kitty Genovese was murdered.²² Results like these lead Doris to

¹⁹ John Doris (2002) *Lack of Character* (Cambridge Press)

²⁰ Milgram, S. (1978). Obedience to authority.

²¹ Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison.

²² Manning, R., Levine, M., & Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping: the parable of the 38 witnesses. *American Psychologist*, 62(6), 555.

conclude that situational framing conditions shape human behavior far more strongly than do individuating global dispositions. If Doris is right about the weakness of character in determining action and external pressures often guide our action, the muscle metaphor is strengthened against a Watson-style objection. If external pressures what defeats the will, then it would make some sense to describe the will as a muscle. If internal pressures defeat the will, it seems less like a muscle.

However, the muscle metaphor does break down when it comes to the sort of pressures to which willpower tends to respond. Of course situational pressures effect our decision making process and these effects very often take place outside of our awareness, but they reflect a series of reasons for action about which we are typically unaware. For instance, obedience to authority and the diffusion of responsibility mentioned earlier, could both be described as participating in a sort of in-group behavior that is a normal part of the human disposition—for better or worse we are a herd animal. Contrary to some misconceptions herd animals do not act in the best interests of the group at their own expense, rather they follow the leaders and others within the herd. Sometimes this is in the interest of the individual and the group, sometimes it is in the interest of neither.²³ Entire herds of buffalo following one another off the edge of a cliff is not exactly individuating behavior, but it is indicative of an internal motivation to stick with the group. It should come as a shock to no one that the people are much more alike than they are different, so individuating characteristics will be significantly fewer than our similarities. The real value of situationist results in social psychology is that they force us away from a

²³ Grandin, T. (1980). Livestock behavior as related to handling facilities design.

tendency to look at our interesting differences and to confront the often overlooked similarities in our motivational systems. Confronting these overlooked impulses is extraordinarily valuable because *ceteris paribus* we make better decisions with more complete evidence. As we become more aware of the ways small situational differences can lead us to act badly we begin to have some hope of overcoming them.

Problems With The Muscle Metaphor: Strategic Success

A related consideration reveals a third difficulty with the muscle metaphor, specifically, it turns out that the most effective way to incorporate responsiveness to hidden motivational forces into our deliberation is to attend to the sort of situations into which we place ourselves. In a recent book on willpower Baumeister and Tierney call this strategy “playing offense,”²⁴ John Doris calls it “avoiding ethical brinkmanship.”²⁵ In both cases the recommendation amounts to an empirically based strategic implementation of willpower. The most strong-willed people it seems are those who do not struggle and fight contrary inclinations the hardest, rather strong-willed people are those who fight effectively and fighting effectively often means avoiding conflicts before they arise. But, this sort of strategy doesn’t sound much like the way a muscle operates; it is rather more like a skill.

Consider the contrast between two people, Tom and Susan, each with an equal desire to drink heavily at a summer party with friends. Both Tom and Susan also have other goals with which heavy drinking conflicts. Tom decides that he will attend the party and drink moderately; Susan decides that it would be best to avoid

²⁴ Baumeister and Tierney (2011) *Willpower* (The Penguin Press; NY)

²⁵ Doris (2002) p.147

the party altogether. Who is more likely to end up drinking heavily? It is pretty clear that Tom is more likely to drink heavily. Tom is engaging in risky behavior by attending the party despite his urge to over-indulge. His strategy relies on the heavy use of brute strength of will—Tom’s strategy requires big "willpower muscles"—especially if he intends to drink moderately which statistically would weaken his self-control. Susan on the other hand employs a risk averting strategy and dramatically lowers her chances of over-indulging. Susan’s strategy relies not on big willpower muscles, rather the strength of her choice lays in employing a sound technique. Recent empirical research suggests that effective self-controllers are much more like Susan than they are like Tom.²⁶ It seems that the effective use of willpower has more to do with strategic deployment, than it does with brute strength.

In an important respect strong-willed people are like the people in strongman competitions, they exercise regularly, but they also develop techniques which maximize the effectiveness of their developed capacity. Strongman competitions involve participants who attempt lift, carry, hold or throw extremely large and wieldy objects, such as beer kegs, tractor tires, telephone poles and cars. These competitions involve tremendous physical strength which can only be developed over many years of training. However, what distinguishes the best strongmen from the average ones is not that they are overwhelmingly stronger; rather, it is that they are familiar with the apparatuses and have developed a good technique.²⁷ The strongest strongmen are just as strong as average strongmen, what sets them apart

²⁶ Baumeister et. al. (2012) "Everyday Temptations: An Experience Sampling Study of Desire, Conflict, and Self-Control" *Journal of Personality and Social Psychology* Vol. 102, No. 6, 1318 –1335

²⁷ Or so the commentators, themselves strongmen, report. Also, leverage & weight > weight.

is that they are more skilled at employing their muscular strength. Likewise, what sets strong-willed people apart seems not to be that they have a more developed capacity for self-control, but that they are more skilled in deploying their resources.

Is Willpower Just Sugar?

Before discussing the skill of willpower, consider some recent work closely associating glucose levels with performance on strenuous tasks of self-control. One might conclude that the will-power “muscle” is tied closely to adequate levels of glucose in the pre-frontal cortex. In one such study participants were required to attend to a boring data feed while being tempted with videos of either classic skits from Saturday Night Live or a Robin Williams standup routine. The operational variable in this study was the glucose, or sugar, levels available to support neural functioning. Participants with low levels of glucose performed significantly worse on the demanding will power tasks. However, half a can of soda was able to resolve glucose deficiency and produce improved results.²⁸ These results have been widely supposed to indicate that blood glucose is an important mechanism of self-control. Further study of glucose levels on willpower reveal that acts of self-control have the effect of depleting subject’s levels of glucose and lowering their levels of self-control. It seems that willpower and failures of willpower are reliably linked to human physiology, or a willpower “muscle.” But brute exercise of self-control is notoriously ineffective and serious efforts of self-regulation require more than a robust physiology, they require sound technique.

²⁸ Baumeister, Roy (2007) “The physiology of willpower: linking blood glucose to self-control” *Personality and Social Psychology*. Vol. 9 No. 2 pp. 325-356

In short, although treating willpower as simply a metaphorical muscle is mistaken, it is not a mistake to treat the metaphor as capturing part of the story of willpower. Its flaw is that it does not adequately account for the strongest type of self-control—the strategic avoidance of temptation. Nor, does the muscle metaphor adequately describe the sort of pressures that tend to undermine strength of will. Rather, it is best to think about willpower as a skill reliant on a capacity one might describe as “muscle-like” but also dependant on the exercise of sound strategic planning. Thinking of willpower as a skill with a physical component has several advantages. First, the willpower as skill model is able to account for strategic instances of willpower. Second, the skill model of willpower is more amenable to an explanation of weakness of will in terms of poor decision making as opposed to connoting irresistible forces. Finally, thinking about willpower as a skill has the advantage that it describes self-regulation in terms which describe an activity or practice of agents, rather than a part of agent’s physiology.

The Skill Model of Willpower

Several researchers have argued that when it comes to examining the empirical data the skill model of self-regulation is at an explanatory disadvantage.²⁹ This claim is false, but warrants examination. The trouble, they say, is that the skill is model would both, fail to predict the phenomenon of ego depletion, and, would predict distinct varieties of willpower applicable to different spheres of influence. Consider first the phenomenon of ego depletion. As we saw earlier a person’s performance on tasks which require self-regulation can be manipulated for the worse

²⁹ Roy Baumeister, et. al. (1998) “Self-Control as Limited Resource: Regulatory Depletion Patterns” *Journal of Personality and Social Psychology* Vol. 74, No. 3, 777

by requiring them to exercises self-control before completing the task. This is why not laughing at a Robin Williams skit makes it less likely you will be able to track objects on a screen. Critics argue that someone who is skilled in a task does not get appreciably worse at the task as they perform it over several iterations. The critic's reasoning is that skill level remains essentially constant over multiple trials and should predict a relatively constant performance across iterative tests of self-control. They argue, that in the short-term muscle's weaken through use, while in the short-term skills remain constant through use and since willpower weakens through use in the short term it is better described as a muscle. What these critics fail to consider is that it may be best to think of willpower not exclusively as either a skill or muscular-type activity, but as a hybrid. The exercise of willpower in the real world requires that people act and as has been shown people tire, even when they retain their basic skill set. Putting skills into action requires the sort of exertion that depletes muscles and the supply of glucose to the brain. But, this does not indicate that willpower is glucose or any other deplete-able element. Rather, it means that willpower requires as an element of muscle-like capacities which are depleted through use.

Take for example, the difficult time one might have resisting the temptation to eat all of the vegan coconut, cardamom and lime cupcakes sitting in the refrigerator. In part this resistance is governed by one's available stores of mental energy in terms of rest, glucose and other factors. But given enough time and the requisite urges, one is very likely to eventually become depleted and eat those cupcakes. On the other hand, one might realize how difficult brute resistance will be and choose to

use his limited strength to make a decision with positive long-term effects. For example, rather than simply fight the urge to eat a cupcake, one might bring the cupcakes in to work to share with colleagues. The act of giving away delicious cupcakes may be a challenging test of willpower, but it is a brief challenge on par with resisting eating the cupcakes for only a moment. We would expect that both activities would lead to some ego depletion or weakening of the willpower “muscle.” However, in the case of giving away the cupcakes one’s efforts have been used far more efficiently and produce a significant advantage in terms of avoiding the cupcakes long term. Once one has given the cupcakes away the problem of resistance is solved. An advantage of the skill model of willpower is that it is able to incorporate strategic elements such as giving away the cupcakes. Suppose we ask “Who has more willpower or exhibits a greater degree of self regulation the person who sits for 5 minutes not eating the cupcakes, or the person who spends five minutes giving the cupcakes away?” These may be basically indistinguishable on the view of willpower as a muscle—both require resisting a desire for 5 minutes. However, on the view that willpower is a skill we can account for the likely effectiveness of resistance and say that giving away the cupcakes exhibits a greater power of self-regulation. This is important because strategic thinking is a crucial element of willpower.

It turns out that, all things being equal, making decisions about how one ought to act also has the effect of depleting one’s capacity for self-regulation.³⁰ This is in

³⁰ Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2014). Making choices impairs subsequent self-control: a limited-resource account of decision making, self-regulation, and active initiative.

part why sometimes all we want is not to have to decide about yet another issue. If you have ever said, "I will eat absolutely anywhere or anything as long as you choose," it may be that you were storing whatever reserves of self-control you had left for more important decisions. This phenomenon is often described as decision fatigue and it explains why planning an ornate wedding, for instance, is so difficult. The sheer number of decisions to be made is literally exhausting to one's willpower. However, some decisions tend to reduce the choices we will need to make in the future, while others leave future decisions to be made or even open more choices. When confronting a mountain of decisions a wise strategy for preserving one's ability to effectively deliberate about that mountain of choices is to eliminate as many unimportant decisions as possible.³¹ By wittling down the number of choices one confronts, it is possible to maximize the effectiveness of one's powers of self-control. A process of explicitly making narrowing decisions in the face of overwhelming choices is a learned skill, but it is a skill which depends on adequate supplies of mental energy and which can therefore be depleted.

The simplest way to understand the difference between strength of will as a muscle or a skill is to consider the parallel to physical strength. What is the appropriate measure of physical strength? It seems uncontroversial that physical strength (as opposed to health, endurance, etc.) is best measured by the ability to lift heavy objects. What it means to be physically strong is to be able to lift heavy loads. But in any test of this capacity it is practically impossible to separate the skill associated with lifting from the muscular requirement. This is not an artifact of

³¹ Of course, limiting important choices is also a valuable strategy. Odysseus' choice to lash himself to the mast, for example.

testing. Rather, it is due to the fact that human strength is the skilled use of one's muscles, frame, etc. For lack of a better way to put it, our mind and bodies are not the sort of distinct entities that can be adequately understood independently of one another. Similarly, we ought to ask the question, what is the appropriate measure of strength of will? Again, the answer will need to relate to how one is able to perform. What it means to be strong willed is to be able to regulate oneself in the service of regularly making and implementing reasonable choices about action. Any test of willpower will involve both a strategic and physiological element. Ego depletion highlights the physiological role of willpower, but we have at least two explanatory options. Either we use the muscle metaphor and begin to reduce will power to the physiological, or we describe willpower as a skill of physiologically grounded beings incorporating both cognitive and physiological components. Both accounts explain depletion, but the skill account has the advantage of also explaining strategic decision making.

Chapter 2

Willpower Strategies

Because, willpower involves more than brute psychological force it is important to get some idea of the strategies that tend to generate strength of will. Fortunately there a number of strategies with demonstrated success. It is helpful to think about the skills of willpower as dividing along two axes: long-term vs. immediate strategies, and strategies of engagement vs. strategies of avoidance. The strategies I mention here should not be taken as an exhaustive list, rather they are examples of the types of willpower strategies available. Short term strategies include mindfulness and distraction, these are short-term in the sense that they require the agent to exert some effort at the time of temptation. Of course, there may be some extended training in mindfulness or distraction at work, but the crucial element of immediate strategies is that they are deployed once a temptation arises. Long term strategies involve advanced planning for temptations that one expects will arise. Examples of long term willpower strategies include avoiding temptation by not engaging in brinkmanship and implementing precommitment strategies to prevent oneself from succumbing to temptation. Along the second axes the types of strategy divide differently. Mindfulness and pre-commitment strategies operate by encouraging engagement, but implementing procedural safe guards against a temptation overwhelming one's resistance. Conversely, strategies of distraction and

avoiding brinkmanship discourage engagement with temptation and operate by removing the temptation to the greatest degree possible.

Willpower Strategies	Immediate	Long-Term
Engagement	Mindfulness	Precommitment
Disengagement	Distraction	Avoidance

Chart of Willpower Strategies, Fig. 1

This chapter begins by describing the four types of willpower strategy mentioned above. Each type of willpower strategy is explained both theoretically and using a non-fanciful example. I then defend the idea that willpower can be extended over time. My view is that willpower can be an anticipatory activity involving the preparation for expected temptations for three reasons. First, skillful deployment of willpower is always linked to an agent whose interests and evaluations are extended over time. Second, in the strongest cases of willpower the exercise of willpower now not only affirms one's values now, but helps make an agent strong-willed over time. Third, if willpower is only thought of as immediate it seems that people who constantly walk on the edge of losing control actually have the most control. To some this conclusion will seem absurd. In any event, as a stipulative matter I see little harm in a technical vocabulary which distinguishes long-term self control from willpower. However, I argue that the terms "willpower" and "self-control" should be used interchangeably.

Distraction

A series of studies conducted by Walter Mischel in the late 1960's set the stage for thinking about willpower in terms of effective strategies in addition to those already discussed.³² Mischel and his colleagues tested the self-control mechanisms of ninety two 3-5 year old children by offering them a marshmallow and then telling the children that they could have a second marshmallow if they would only wait 15 minutes to eat the first one. The children responded with a range of interesting strategies including paying close attention to the treat at hand including holding, licking or nibbling the treat, some studiously avoided looking at the treat, others looked for distractions such as available toys, some sang songs to themselves reminding them that they were supposed to wait and one particularly cunning youngster simply fell asleep. Each of these strategies was effective in some cases, but there were some clear tendencies in terms of the effectiveness of strategies aimed at delaying gratification. Initially Mischel hypothesized that focusing on the expected reward would strengthen performance in the willpower task, but the results indicated just the opposite. It turns out that focusing on the expectation of the second marshmallow usually led the children to give in to temptation and eat the marshmallow early. On the other hand, children who distracted themselves by thinking pleasant thoughts or playing with toys were successful in delaying

³² Walter Mischel (1970) "Attention in Delay of Gratification" *Journal of Personality and Social Psychology* Vol. 16, No. 2, 329-337 and (1972) "Cognitive and attentional mechanisms in delay of gratification." *Journal of Personality and Social Psychology*, Vol 21 No. 2 pp. 204-218.

gratification better than half of the time. Mischel's results point to a promising strategy for self-control, distraction.

The overwhelming effectiveness of the sleep strategy in Mischel's study helps to explain why distraction can be such a powerful asset when resisting a desire. The sleeping child was no longer tempted to eat the single marshmallow and was thus able to wait patiently without fatiguing his capacity for self-regulation. Similarly, when we are trying to avoid eating that extra slice of pizza it is unwise to stare longingly at the leftover pizza. We might more effectively resist the impulse to overeat by closing the pizza box and turning toward the dishes to be washed. These actions help preserve our willpower by changing the way we view the situation in a way that requires significantly less self-regulation. It is much easier to wash dishes than it is to resist eating pizza, and if we can see ourselves as doing the one and not the other this strategy can help increase effective willpower. Self-distractions are especially effective strategies for enhancing self-control when they shift our actions from brute exercises of self-control like watching a wanted marshmallow, to more engaging tasks. By implementing distractions we take our minds off of the conflict we find difficult to manage and thereby preserve willpower.

Of course, it is important to remember that not all distractions are valuable strategies for maximizing willpower; in fact a lot of distractions are highly detrimental to effective self-regulation. Remember the white bear from Baumeister's studies? The whole point of that bear was to distract people from the task at hand by asking them to instead focus on not thinking about the bear. By distracting people from their primary task with another challenging task the white bear proved distracting in a

negative sense—it actually made the task of self-control harder. Most people in the modern world have experienced this phenomenon in one way or another. For example, when you are working hard to complete a project for work and email keeps coming in, your office phone rings, followed by your cellular phone, then a facebook message comes through, and all of a sudden you have utterly lost your momentum in the work project at hand. Such distractions seriously tax ones reserves of willpower.

The presence of willpower draining distracters gives us a hint that distractions work best as strategies of self-control when the willpower task at hand has to do with resisting a desire, as opposed to when the willpower challenge is to do something like accomplish a task. However, it would be wrong to suppose that distractions are always helpful at resisting desires. For example, Baba Shiv has shown that shoppers who are distracted were nearly 20% more likely to choose a product being marketed through taste testing than where less distracted shoppers.³³ Shiv had 224 people remember either a two digit number (low distraction) or an eight digit number (high distraction) and then had them each sample Lindt chocolate. The subjects were then asked to choose between a Godiva chocolate and a Lindt chocolate. Those with eight numbers to remember chose the Lindt at 66.1%, while those with only 2 numbers to remember chose Lindt at only 46.4%. The result contradicts the expectations of 92% of marketing executives who presumed that taste testing would be most effective without distractions. A second study provided added information

³³ Shiv and Fedorikhin (1999) "Heart and Mind in Conflict: The interplay of affect and cognition in consumer decision making." *Journal of Consumer Research* vol. 26 pp278-292.
Shiv and Nowlis (2004) " The Effect of Distractions while Tasting a Food Sample" *Journal of Consumer Research* vol 31 pp. 599-608.

to the subjects by telling half of them that Gourmet magazine had recently voted Lindt Europe's best chocolate. This information should lead to more people choosing the Lindt, and it does so in the low distraction group with the percentage choosing Lindt increasing from 46.4% to 69.6%. But the high distraction group shows an insignificant change from 66.1% to 65.4%. It seems that when distracted consumers not only are impacted by the marketing ploy at a greater level, but they are also less sensitive to information relevant to the choice at hand. Of course, we mostly want to make our own decisions based on good information and not marketing, but it seems that distractions weaken our resolve in this regard. Shiv associates this weakening of resolve with the power of distraction to emphasize pleasurable affective experience at the expense of considering other relevant information.

The idea that distractions can heighten rather than lessen affect is unusual given the widely supported research conclusion that distraction can lessen pain. In numerous studies it has been shown that among individuals in painful conditions those who are distracted report lower levels of experienced pain, while those who are undistracted or told to focus on the pain report higher levels of experienced pain.³⁴ However, research on the effectiveness of distraction in relieving pain is equivocal. For example, in a recent study eighty one 9-18 year old children took part in a cold pressor task and divided into an undistracted group and a group involved in an attention demanding tone detection task. The children in the distraction group

³⁴ McCaul, K. D., & Malott, J. M. (1984). Distraction and coping with pain. *Psychological bulletin*, 95(3), 516. Kleiber, C., & Harper, D. C. (1999). Effects of distraction on children's pain and distress during medical procedures: a meta-analysis. *Nursing research*, 48(1), 44-49. Gold, J. I., Kim, S. H., Kant, A. J., Joseph, M. H., & Rizzo, A. S. (2006). Effectiveness of virtual reality for pediatric pain distraction during IV placement. *CyberPsychology & Behavior*, 9(2), 207-212.

attended to the distracting task, and reported paying less attention to the painful sensation, but also reported experiencing higher levels of pain by ending the task sooner. This finding was significant across the group, but was especially the case for children who on an earlier measure tended to exaggerate the experience or importance of pain. So it seems that for children prone to catastrophizing distraction may enhance the experience of pain.³⁵

Mindfulness

It is unclear exactly how and when distraction increases or decreases willpower, but some have begun to suggest an alternative strategy of attending to difficult challenges such as resisting cravings or sticking with the cold pressor task. The strategy here is known as "mindfulness" and has the substantial advantage of applying to many traditionally challenging situations in which distraction would be either unreliable or unviable. Sarah Bowen has helped to develop the strategy of mindfulness, which she describes as "surfing the urge."³⁶ Mindfulness involves focusing on the details and nuance of experience involved with a difficult task without trying to hide or obscure the phenomena. The process is composed of three distinct elements, first accepting or attending to the urge or temptation; for example, this could mean paying close attention to the bodily sensation of wanting a cigarette. Second, take a deep breath and use this moment to pause and plan a response to the urge. And finally, look to your broader set of interests and look for the first opportunity to recommit to your goal.

³⁵ Verhoeven, et. al. (2012) "**Pain** catastrophizing influences the use and the effectiveness of **distraction** in school children." *European Journal of Pain*, Vol 16(2) pp. 256-267.

³⁶ Bowen, S. & A. Marlatt (2009) "Surfing the Urge: Brief Mindfulness-Based Intervention for College Student Smokers" *Psychology of Addictive Behaviors* Vol. 23 (2009) pp. 666-671

As a test of mindfulness training Bowen asked a group of active smokers to abstain from smoking for twelve hours and then to sit in a room with no distractions and to stare at a pack of their favorite cigarettes. They were then asked to pack the cigarettes, then to slowly unwrap the cellophane, then to slowly open the pack and to take time to smell the cigarettes, then to remove one cigarette and hold it for a while, then to rest a cigarette in their mouths, then to remove a lighter. This process took over an hour and the smokers were never allowed to light-up. Half of the smokers had received earlier mindfulness training, and half had not. After their training the smokers were asked simply to record the number of cigarettes they consumed in the subsequent week. They were not instructed to “surf the urge” nor were they instructed to cut back on their smoking. However, by the end of the week the group trained in mindfulness had reduced their consumption of cigarettes by 37% while the control group had not reduced their intake at all. In follow up studies involving the training of those in residential substance dependence programs Bowen had similar results with those trained in mindfulness showing an increased resistance to relapse.³⁷

Another study directly compared mindfulness training with distraction as strategies for emphasizing self-control. In this study ninety eight students were asked to carry around a clear plastic box of Hershey’s kisses for 48 hours and told that they should try not to eat any. The students were divided into three roughly equal groups. One group was trained in distraction techniques and instructed to

³⁷ Bowen, S. et. al. (2009) “Mindfulness-Based Relapse Prevention for Substance Use Disorders” *Journal of Consulting and Clinical Psychology* Vol. 78 pp. 295-305 See also, Witkiewitz and Bowen (2010) “Depression, Craving, Substance Use Following a Randomized Trial of Mindfulness-Based Relapse Prevention” *Journal of Consulting and Clinical Psychology* Vol. 78 pp. 362-374.

distract themselves whenever they had the urge to eat a kiss. The second group of students was trained in mindfulness and instructed not to suppress their desire, but to notice how the desire felt in their bodies, to pause and then to recommit to their goal. And the finally the control group was not instructed with any specific control strategy.³⁸ The results of this study were that most of the people who ate kisses came from the control group, with half as many in the distraction group eating kisses. The acceptance based group ate fewer kisses than the control, but more than the distraction group. However, the acceptance group did perform the best among those who reported high levels of cravings.

It seems that one effective strategy for self-control is to learn to accept one's inner conflicts and allow them to pass without acting on them. Underlying this counter-intuitive strategy may be the simple truth that we cannot eradicate our urges through distraction. In fact sometimes trying not think about the white bear, makes that almost impossible. Exercises of self-control are difficult in part because they keep coming to our attention despite (and sometimes because of) attempts to suppress them. At one time or another one must confront one's inner conflicts and when doing so it is best not to rely on brute willpower, but to have a strategy for effective implementation. Mindfulness training provides such a strategy.

It may be helpful to see how mindfulness works to compare it with the common advice of positive psychology to "visualize success."³⁹ The idea is that by

³⁸ Forman, Evan M. et. al. (2007) "A comparison of acceptance and control-based strategies for coping with food cravings" *Behaviour Research and Therapy*, Vol 45(10) pp. 2372-2386.

³⁹ Bandura, A. (1989). Human agency in social cognitive theory. *American psychologist*, 44(9), 1175., Weinberg, R.S. & Gould, D. (2007). Introduction to psychological skills training. In R.S.Weinberg, &

visualizing oneself performing the difficult task it will become easier to actually perform the task. However, recent research suggests that visualizing successes of willpower can lead to failure. Two things go wrong when we visualize success. First, we visualize succeeding, but not the steps we take and obstacles that we must overcome in order to reach success and second, we recall earlier successes and bask in them, coasting on our present challenge. It turns out that effective self-regulation is bolstered by imagining the ways that we might be challenged rather than imagining ourselves simply succeeding. Mindfulness training focuses on attending to exactly those impulses and urges which stand in the way of effective resistance. Imagining success is important in a motivational sense because it can help to overcome our inertia. But it is more important that the imagination of success be realistic and not an impoverished fantasy of success, otherwise the visualization becomes an impediment to progress.

A pair of researchers working in US business schools provides some compelling research on the way in which idealizing the future hurts our willpower and the degree to which we are prone to such idealizations.⁴⁰ Robin Tanner and Kurt Carlson asked people how much they planned to exercise in the upcoming week. They then asked another group how much they would exercise in the upcoming week in an ideal world. The two groups reported the same expectation of exercise, which Tanner and Carlson interpret as meaning that the default framing about the

D.R. Gould (Eds.), *Foundations of sport and exercise psychology* (4th ed.). Champaign, IL: Human Kinetics.

⁴⁰ : Robin J. Tanner and Kurt A. Carlson (2009) "Unrealistically Optimistic Consumers: A Selective Hypothesis Testing Account for Optimism in Predictions of Future Behavior" *Journal of Consumer Research*, Vol. 35, No. 5 pp. 810-822

future is in idealized terms. This means that when we imagine our future selves, we fail to recognize the commitments of time, stresses and conflicting desires that will affect us then. Recall, that this same difficulty arises when people are asked to contribute to a savings program for retirement or to delay gratification for a larger reward. Tanner and Carlson further these findings by showing just how resistant people are to viewing the future realistically. In one version of their study participants in the non-idealized group were asked specifically “Please do not provide an idealistic prediction, but the most realistic prediction you can.” Participants were then interviewed two weeks later to see how much they had actually exercised. The results showed an overly optimistic outlook in the early predictions. But how would new predictions come out if participants had this information. When asked to make another prediction about future exercise the participants predicted even higher rates of exercise than they had initially. When confronted with the hard truth that they had unrealistic expectations people did not adjust their expectations down, they adjusted them up to make-up for the failure—this is a strategy for failure.

Mindfulness training takes the basic insight of positive psychology, that focusing on pathology will never lead to a meaningful life and co-opts what has been previously pathological into the meaning of one's life.⁴¹ Personalizing the experience of temptation and resistance can strengthen will not by increasing the brute force of resistance, but by developing a sense of the resistance as meaningful. By attending to the embodied discomfort of resistance the cigarette smokers took

⁴¹ Seligman, M. E., & Csikszentmihalyi, M. (2000). *Positive psychology: An introduction* (Vol. 55, No. 1, p. 5). American Psychological Association.

ownership over the sensation. Further, the attention paid to ritual and detail of activity allowed the smokers to disentangle a cluster of smoking behaviors and sensations into constituent elements. This exploded view of a routine allowed the smokers greater control over their own behaviors.

It is important to note that the concept of mindfulness in resisting impulses with which an agent disagrees could develop into a sentimental account of temptation. One might mistakenly believe that temptation is somehow to be sought because resisting it gives life meaning. As I consider this potential mistake I cannot help but compare thinking about mindfulness of pain to the 16th century's "noble savage." It was thought by some that the "uncivilized" conditions of Native Americans and others allowed for the unadulterated communion with their own natural goodness. This reification of the natural was a romantic notion in the same way reifying pain or temptation would be. The experience of temptation is not itself a good; rather temptation is a necessary obstacle in living a life of one's choosing. On this account mindfulness is a strategy of attending to the details of the life one chooses in an effort to take control.

Avoidance (Avoiding Brinkmanship)

Brinkmanship is a cold-war term used to describe going to the edge of conflict in order to extract some gain or concession from an opponent.⁴² But being on the brink need not involve the possibility of nuclear war, indeed one can be on the brink of losing in an internal conflict with oneself. John Doris gives an example in which a colleague "with whom you have had a long flirtation invites you for dinner, offering

⁴² Schelling, T. C. (1980). *The strategy of conflict*. Harvard university press.

enticement of interesting food and elegant wine...with the excuse that you are temporarily orphaned while your spouse is out of town."⁴³ If we presume that marital infidelity is undesirable it is probably wise to avoid this situation. Even though you may sincerely value fidelity you may also "recognize that situational pressures may all too easily overwhelm character and avoid the dangerous situation."⁴⁴ The idea is that one can more effectively resist the impulse to infidelity while sober in the cold light of day, than would be possible with wine in candle light. A willpower strategy of avoiding brinkmanship means making the decision to avoid temptation before the temptation arises in full force.

It is not hard to see the value in taking steps to avoid temptation before it arises, we have all succumb to a predictable temptation. Of course, one could probably just exercise brute resistance in the case of a flirtatious co-worker, but avoiding difficult situations makes the task easier and shifts the odds in favor of success. Still an avoidance strategy can be important in ways beyond making a choice easier or improving the odds, it can make a difference in cases where we absolutely could not act as we believe we should through brute exercise of will power. There are some willpower challenges about which brute exercise approaches being useless because the trouble is with our own implicit judgments. In these cases we explicitly judge that one course of action would be best, but our judgment is regularly subverted by a contrary implicit bias. "Unlike explicit bias (which reflects the attitudes or beliefs that one endorses at a conscious level),

⁴³ ⁴³ Doris J (2002) Lack of Character: Personality and Moral Behavior. Cambridge University Press, New York p.147

⁴⁴ Ibid.

implicit bias is the bias in judgment and/or behavior that results from subtle cognitive processes (e.g., implicit attitudes and implicit stereotypes) that often operate at a level below conscious awareness and without intentional control."⁴⁵

Implicit association tests (IAT) work by measuring the "differential association of 2 target concepts with an attribute. The 2 concepts appear in a 2-choice task and the attribute in a 2nd task. When instructions oblige highly associated categories to share a response key, performance is faster than when less associated categories share a key. This performance difference implicitly measures differential association of the 2 concepts with the attribute."⁴⁶ For example, a subject might be asked to consider images of people of different races followed by positive or negative words. The rate differential between the time it takes to associate each race with positive or negative words is thought to give a measure of the subject's implicit racial bias. If these tests in fact measure what they report to measure, it seems that there is substantial implicit racial bias. Other versions of the IAT attempt to measure gender bias, age bias, disability, weight bias, sexual orientation bias and other categories.⁴⁷ These tests have become quite popular and results seem to indicate that most of us have at least some implicit biases with which we explicitly disagree.

No one who takes an IAT on race wants to discover that they are strongly biased against faces of African origin, but that is the most statistically common

⁴⁵ Banaji, M., & Heiphetz, L. (2010). Attitudes. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology*, 5th edition (pp. 348-388). New York: Wiley

⁴⁶ Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464.

⁴⁷ Project Implicit <https://implicit.harvard.edu/implicit/selectatest.html>

result.⁴⁸ IAT's reveal an especially difficult problem with implicit biases, namely that bias persists in our choices despite our explicit commitments and efforts to the contrary. As with other phenomenon, racial bias is not alleviated simply by awareness of the bias combined with the desire to act otherwise. This is not the kind of problem that is susceptible to correction through brute force of will. One could give oneself a moral hernia trying to lift the weight of racism to no avail. A long term strategy beyond moral strain is needed to overcome implicit biases.

Consider the sorts of situation in which an implicit bias may be particularly pernicious, in hiring for example. In a recent study participants were asked to imagine that they were a newly elected mayor seeking to root out police corruption by hiring a new chief of police. They were presented with applications including the names Brian and Karen. In the first condition Brian was described as well-educated and skilled in administration, but as having little "street" experience in terms of making arrests and pursuing criminals. Karen was described in opposite terms. In the second condition roles were reversed and Brian was "streetwise" and Karen was well educated. In both conditions the study participants overwhelmingly chose Brian over Karen. In the first condition participants claimed that education was clearly the most significant qualification for the job. In the second condition participants argued that experience was the most salient qualification. But, it seems that the qualification participants most valued was gender.⁴⁹

⁴⁸ Xu, K., Nosek, B. A., & Greenwald, A. G. (2015, February 25). Race IAT 2002-2013. Retrieved from osf.io/52qxl

⁴⁹ Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, 104(2), 207-223.

Participants in the "Hiring a New Police Chief" study agreed to act objectively when making their choice, but their implicit gender bias crept into their decision making nonetheless. Other studies reveal similar biases in housing with respect to stereotypically white and black names such as "Tyrone" and "Bradly."⁵⁰ Further evidence exists that this effect exists for teachers as well.⁵¹ Brute intention and even anti-discrimination laws are not enough to achieve fair decision making. We also require long-term strategies to make our intentions effective. One example of this sort of strategy would be to remove irrelevant identifying information such as race, or gender from job applications, housing applications and the like. This strategy works by allowing the decision maker to decide in advance what sorts of information is not relevant and then precluding consideration of that information by making it unavailable. By avoiding irrelevant information the decision maker who would be unable to resist her implicit biases through brute force is able to do as she judges best.

Avoidance strategies usually are not complete in the sense that they ensure or guarantee that self-avowed better reasons are acted upon. More typically an avoidance strategy simply increases the odds that one will be able to exercise control in the direction one favors at the moment of action. An avoidance strategy typically involves coordinating one's activities to avoid temptations and to choose instead situations which support our goals. Part of what this means is being aware of the physiological constraints on willpower and scheduling activities and goals

⁵⁰ Hanson, A., & Hawley, Z. (2011). Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *Journal of Urban Economics*, 70(2), 99-114.

⁵¹ Hinojosa, M. S., & Moras, A. (2009). Challenging colorblind education: A descriptive analysis of teacher racial attitudes.

accordingly. For example, because we know that hunger is a sign of low glucose and low glucose especially inhibits the self-control area of the prefrontal cortex, we should try to avoid making serious decisions (or grocery shopping) while hungry.⁵² Similarly, inadequate sleep causes the brain to process glucose inefficiently and leads to what some psychologists have called “temporary cognitive dysfunction.”⁵³ It is best to make difficult choices when well rested. A related consideration is that throughout a day of simply being awake, making even mundane choices, controlling one’s affect, directing one’s thoughts, etc. the physiological basis of willpower becomes depleted. For this reason it is often an effective willpower strategy to arrange one’s day so that the most challenging activities come early in the day when one is at one’s strongest.

Precommitment Strategies

It is of course not always possible to rearrange things to avoid temptation, so it is important that we can plan for predictable failures of brute willpower in other ways. Thomas Schelling won a Nobel Prize in economics for his work on the cold war which argued that given entirely open options temptations would be insurmountable. On his view the only way to avoid these temptations was for a nation to limit its options through a strategy of precommitment.⁵⁴ Precommitment strategies close off options, make some options more costly, delayed or difficult than

⁵² Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., ... & Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. *Journal of personality and social psychology*, 92(2), 325.

⁵³ Carroll, M. F., Burge, M. R., & Schade, D. S. (2003). Severe hypoglycemia in adults. *Reviews in Endocrine and Metabolic Disorders*, 4(2), 149-157.

⁵⁴ Schelling, Thomas C. (1980). *The Strategy of Conflict* (Reprint, illustrated and revised. ed.). Harvard University Press.

they would otherwise be in an effort to keep an agent from acting against her own better judgments.

The classic example of a precommitment strategy is the case of Ulysses ordering his men to strap him to the mast and then plug their own ears with wax. This way Ulysses imagined he would be able to hear the sirens song, but would be unable to fail himself and his men by wrecking on the island. Ulysses didn't want to avoid the sirens, but he also couldn't resist them by brute force of will. Having himself tied to the mast committed him to not wrecking the ship, while allowing him the preferred engagement with the sirens. This case is so central to the literature on precommitment that arrangements of this type have come to be known as "Ulysses contracts." The trouble with this classic case is that it seems unrealistic and may mischaracterize strong will as only about overcoming an individual temptation.

Consider instead the very real case of heroin addicts who opt for methadone maintenance treatment (MMT) as a strategy for overcoming addiction. Methadone treatment facilities work by substituting methadone (a synthetic opioid) for heroin. This substitution strategy works to reduce heroin use because methadone provides a similar high to heroin, cutting back or eliminating withdrawal, and at high enough doses the use of methadone effectively blocks the high of heroin—so junkies cannot reliably use both drugs. However, the real key to MMT is that a medical professional is in charge of administering the daily dose.⁵⁵ In MMT the patient is lashed to the proverbial mast and turns over the job of navigating to the treatment center staff.

⁵⁵ Garcia-Portilla, M. P., Bobes-Bascaran, M. T., Bascaran, M. T., Saiz, P. A., & Bobes, J. (2014). Long term outcomes of pharmacological treatments for opioid dependence: does methadone still lead the pack?. *British journal of clinical pharmacology*, 77(2), 272-284.

The point of this strategy is to avoid the frequently overwhelming temptations of addiction and to do so in a way that moves toward greater and greater self-control.

Of course, when we involve others in our precommitment strategies we do not normally expect that they will guide us as completely as Ulysses' men. Typically when precommitment strategies involve others they do so through what Howard Becker calls a "side bet."⁵⁶ As Ainslie describes the concept "We may just let a friend know that changing a certain behavior is important for us, so that the friend will be disappointed if we don't actually change it."⁵⁷ This strategy eschews physical compulsion for the more subtle controlling influence of social pressure. We desire the love and respect of those close to us and the idea is that by announcing our commitments in advance we may change the calculus of succumbing to temptation later. By making a public commitment when it is relatively easier we can make a harder choice down the road easier. In one sense the later choice becomes easier because the agent has more at stake than she otherwise would. But in a deeper sense the precommitment of publicity ties the agent to a community of support and helps to place the later decision in the context of the agent as she sees herself as a member of a larger social network.⁵⁸

Other precommitment strategies link the agent's future self more closely to the evaluations of her past self through the formation of personal rules. Personal rules are resolutions designed to maintain a course of action despite the fact that certain temptations are likely to arise. The trouble with personal rules is that they

⁵⁶ Becker H. Notes on the concept of commitment. *American Journal of Sociology*. 1960;66:32-44.

⁵⁷ Ainslie, G. (2001). *Breakdown of will*. Cambridge University Press. p.75

⁵⁸ Gilligan, Carol (1977) In a Different Voice: Women's conceptions of self and morality. *Harvard Educational Review* 47, 481-517.

seem to be derivative of brute exercise of will. As Ainslie puts it "This may be the most common way that people deal with temporary preferences, but also the most mysterious."⁵⁹ Elsewhere Ainslie asserts "personal rules operate most effectively with countable criteria."⁶⁰ As precommitment strategies personal rules are different because whereas Ulysses contracts rely on "external" controls, personal rules are "internal." Of course, this is why personal rules often fail and to the extent that they work it may be that the monitoring of "countable criteria" represents a pseudo external locus of control. In any event, the very notion that an individual self is rigidly distinct from her milieu is part of the problem to which situationist literature alerts us.

Self-Management, Self-Control or Willpower

In conversation with other philosophers it seems the majority philosophical opinion is that long-term strategies like strategic avoidance and Ulysses contracts are somehow distinct from short term examples of willpower.⁶¹ Among those I have discussed the matter with many favor the terms "self-control" or "self-management" for strategies that confront *expected* or *long-term* challenges. The term 'willpower,' they think, should be reserved for cases of overcoming *unexpected* difficulty whether through brute force of will or a short-term strategy like mindfulness. Because expected difficulties are typically planned for in advance, the labels "self-control" or "self-management" would be applied to what I have called "long-term" strategies. Conversely, because unexpected challenges cannot be combated in advance, what is labeled "willpower" might correspond to what I have called "short-term" strategies.

⁵⁹ Ainslie (2001) p.78

⁶⁰ Ainslie, G. (1999). The dangers of willpower. in *Getting hooked: Rationality and addiction*, edited by Jon Elster, Ole-Jørgen Skog p. 65-92.

⁶¹ This sense comes from discussion with Al Mele and others on my dissertation committee.

I think this standard way of talking is wrong, but understandable, given traditional conceptions of identity and philosophical skepticism about weakness of will.

I suspect that at the root of the supposed distinction between "self-management" and "willpower" is a worry about the implications of "self-management" for the conception of the self. In my view it is helpful to think about extended willpower strategies like prosthetic devices; they are not always to be thought of as part of oneself, but when sufficiently integrated into a life they become as authentic a part of oneself as any other. It is very difficult to suppress our cognitive biases and we use strategies like avoidance and Ulysses contracts as thinking tools to make our lives more like the lives we imagine for ourselves. Using tools to work around our short comings is not a gimmick to be considered external, coercive, or somehow less authentically a matter of will. To the contrary, we are deeply integrated with our tools to the point that they become part of ourselves.

In one way this is an argument by analogy, long-term strategies are thinking tools or devices which we integrate into our decision making in the same way that prosthetic legs are integrated into our walking or artificial hearts are integrated into the circulation of our blood. Language is perhaps the most powerful and widely used mental tool used to enhance willpower. Brute force of will often includes the strategic use of language as in case of chanting or repeating a mantra, perhaps, "I think I can. I think I can." The use of social tools, like identifying student exams with numbers rather than names, or calling on students using a random number generator, serve to work against implicit biases by incorporating tools into action. These may feel artificial at first, but so does an artificial leg. Over time willpower

strategies like these become incorporated, some more than others, and some become so central that they fit seamlessly into one's identity.

Chalmers and Clark have argued persuasively that there is no principled distinction between the mind contained within our body and our mind extended in various ways.⁶² They write, "Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle." Some of the long-term willpower strategies described here, such as announcing ones intentions to friends, seem to effectively recruit others into the act of willpower. This sort of strategic integration seems awkward if we are operating with an enlightenment view of the self, but it is powerful in the practical sense and coherent with a more permeable sense of identity. Cognitive bias can be very hard to suppress, but long-term willpower strategies operate not by suppressing neural pathways we oppose, but rather by strengthening pathways around our biases.

Precommitment strategies and long-term avoidance strategies have at their root work done by behavioral economists and game-theorists working mainly on problems of nuclear deterrence. Models of "self-management" typically presume "rational actors"⁶³ and "state actors"⁶⁴ in ways that are disembodied and mistaken about human rationality. It is reasonable to worry that these strategies reflect an automatic mechanism more than they do human character. But we should not fall prey to the genetic fallacy. Work in behavioral economics and game theory can help chart ways around the coercive and worrisome elements of situationism and develop

⁶² Andy Clark, David J Chalmers (January 1998). "The extended mind". *Analysis* **58** (1): 7–19

⁶³ Roger B. Myerson (1991). *Game Theory: Analysis of Conflict*, Harvard University Press,

⁶⁴ Schelling, T. C. (1978). Egonomics, or the art of self-management. *The American Economic Review*, 290-294.

what Rodgers and Warmke call a "*prescriptive situationism*." This form of strategic willpower is "forward looking self-control based on controlling one's situation as opposed to controlling one's character at the time of temptation."⁶⁵ Of course we could call this sort of forward thinking strategy "self-management" and describe it as opposed to "exerting willpower" but to do so has odd results.

Consider the case of two people who are tempted to become smokers. Alice and Robert each think that smoking will make them seem sophisticated and cool. Both Alice and Robert are self-aware enough to know that they are motivated almost entirely by appearances. As such they are both only interested in becoming social smokers; that is, they both want to smoke primarily when they are around other smokers and want to avoid smoking regularly. Of course, the fact of the matter is that social smoking is very likely to lead to nicotine addiction and a destructive pattern of future smoking. Alice realizes that the odds of her becoming only a social smoker are quite low and takes steps to resist her desire by avoiding friends and situations where social smoking would be appropriate. She does not step outside for "fresh air" with the smokers at the office and her dating profile is arranged so that only non-smokers are suggested as potential romantic partners. On the other hand Robert is able to follow through on his idea and becomes a social smoker: he smokes socially for a decade and never falls into addiction. To me it would be odd to claim that Robert is strong-willed and Alice is not. However, some will argue that because she is only using avoidance strategies to keep from smoking she is not strong-willed but rather just a prudent self-manager.

⁶⁵ Rodgers, T. J., & Warmke, B. (2013). Situationism versus Situationism. *Ethical Theory and Moral Practice*, 1-18.

On my view both Robert and Alice are exercising willpower but in different ways. Alice wanted to become a social smoker but decided it wasn't a good choice because the odds of remaining only a social smoker weren't good. This means that Alice has to resist her desire to smoke socially. Alice employs two techniques, first she resists the urge to light up with co-workers by not stepping outside with them when they go to smoke. This is an avoidance strategy. It is easier for Alice to politely decline to step outside when asked than it would be not to smoke once outside with the smokers. Alice wants to go outside with her co-workers, but she chooses not to because making this choice now will make things easier in the future. Indeed, over time things will become even easier because eventually the smokers will stop inviting Alice outside when they go to smoke. Alice's second strategy is to set up her online dating profile so that it is only displayed to non-smokers; this way she avoids any dates with smokers and her romantic partners will not tempt her to smoke. Importantly Alice wants to find a suitable romantic partner and she is not adverse to smokers. She knows that it would improve her chances of finding a mate if she made her profile available to smokers, but it would also make becoming a smoker more likely. Alice's choice to eliminate part of the dating pool is based not on her desires about a mate, but on her judgment about risk. Again, Alice employs a strategy of avoidance.

This gets us to another key difference those who distinguish long-term and short-term strategies may be drawing. Perhaps the idea is that when Alice resists the urge to step outside there is a feeling of difficulty, a feeling of strain, and that feeling is what some associate with willpower. In the case of checking the "non-

smokers" box on her online profile it is hard to imagine that Alice feels much difficulty or strain. This qualitative difference is perhaps what is essential to the proposed distinction between willpower and self-management. If this is what the distinction comes to it is a "hard problem" indeed, but not an impossible one. It is possible to retain the idea that there is feeling associated with willpower and to see it as strategic.

First, it is hard for many of us to see Alice's dating profile as providing much in the way of feeling. It may help to imagine that Alice was a lifelong smoker, that all the people she had ever been romantically interested in were also smokers, for her checking the "non-smoker" box could be an emotional commitment, an act of defiance and self-assertion in a qualitative sense. But the general idea that technological interfaces can provide qualitative feedback is not so distant. The feeling of a prosthetic leg can be emotional for an amputee, likewise the feel of a guitar in the hands of a player can provoke a range of emotion. The point is just that someone's using an external scaffolding for their own behaviors does not mean that the behavior loses its emotional or qualitative relationship to the agent. A strategy for action, even one involving extended mechanisms, can be personal.

Of course, one could reject "external" mechanisms and the hypothesis of extended cognition and still maintain that strategic mechanism are central to an identifiable will in strictly of the training they provide. Mechanisms like computer algorithms can work to train behaviors to be more in keeping with our best judgments. Alice might use an app that helps her avoid smoker friendly restaurants

for instance. Doing so is not turning over her will to the app, rather it is training herself to act more in keeping with her own judgment.

Second, the point of Alice's not stepping outside with her co-workers was to avoid a more qualitatively difficult choice further along. The entire point of a willpower strategy is to make one's judgment effective. It seems very odd to say that a strategy is no longer based on willpower because it is too effective. If Alice judges it best not to smoke and overcomes her desire to smoke without serious emotional strain, that seems like strength of will. Dating smokers, and fidgeting nervously while talking to her coworkers on their smoke break may also be strong-willed. But in the long run a strategy of avoiding the brink is more prudent because it develops the right sorts of habits. And it seems odd to say that someone who prudently controls her desires is not being strong-willed.

Conclusion

Willpower is a skill of self-regulation, one that we develop over time and that doesn't come easily to the young or the unpracticed. This is part of why the Greeks thought that the education of the young in the correct habits was so important. Learning how to choose the best paths around and through human psychology is a difficult matter and without a strategy the task of living well is impossible. Developing willpower means learning to choose routes through one's own psychology and one's environment that end up in keeping with one's judgment. Of course, the gentler and more fortunate the environment, the less concern one need have over strategy. But, for most of us, having the right kind of will, involves having the right strategies.

On Doris's view the primary motivation for human action is the stimulation of a situation. He thinks that character has very little impact on behavior, if character exists at all. This leads him to conclude "Given the practical risks there seems to be little reason for favoring strategies emphasizing 'steadfast exercise of the will' over strategies of 'skilled self-manipulation.'"⁶⁶ In truth strength of will, and especially its strategic facet, is more concerned with developing an agent's positive habits than it is with incorporating external forces of control. The techniques of extending willpower through strategy should be seen as a sort of scaffolding that allows human character and willpower to extend beyond the glitches of our reasoning.⁶⁷ Of course we have character, but it is something earned over time. Character is forged through the regular acts of willpower that lead us to check this box, or refuse to smoke with those people. When set out strategically seemingly insignificant choices collect and bundle together to create a character. This result is an act of will itself and so it should not come as a surprise that there are many people with very little character. Exercising willpower effectively is hard and it is complicated, but it is required for forming a character.

⁶⁶ Doris J (2002) *Lack of Character: Personality and Moral Behavior*. Cambridge University Press, New York p. 149

⁶⁷ Sterelny, Kim (2010) "Minds: extended or scaffolded?" *Phenomenology and the Cognitive Sciences*, Volume 9, Issue 4, pp 465-481

Chapter 3

Weakness of Will

Weakness of will is commonplace in our everyday experience, but it is poorly understood. Recently Neil Levy argued that "weakness of will does not correspond to a useful explanatory category."⁶⁸ On Levy's view what we call weakness of will is really just a single type of ego depletion. He suggests that we would do better to focus on the ways that self-control based in the pre-frontal cortex can be impeded and overridden by the more primitive impulses of our lower brain. Levy is right that we should focus on the ways that our self-control is overcome by fatigue, hunger and the like. However, he is wrong that the concept 'weakness of will' is reducible to one necessary condition. Levy's view that weak-will is not a useful explanatory category is based on the reduction to one type. If the concept is not reducible in the way he suggests, we have good reason to reject his claim about its usefulness. Further, since part of Levy's argument is that no evidence supports a broad account of weak-will, I offer some supporting evidence from the study of addiction. The aim of this chapter is to show that addiction research and a little bit of experimental philosophy can combine to shed light on a longstanding trouble in the history of philosophy.

The argument presented here is that ordinary language, theoretical reasoning and empirical evidence collectively suggest that there is no paradigmatic type of

⁶⁸ Neil Levy, [Resisting Weakness of the Will](#). *Philosophy and Phenomenological Research* 82 (2011): 134-155

weakness of will. Weakness of will is best understood as a polythetic or cluster concept. I begin briefly discussing definitions to show how a concept (especially of complex behavioral traits) can be meaningful without having necessary and sufficient conditions. The second section reviews the philosophical thinking about what weakness of will is and the range of characteristics are associated with the concept. The third section reviews linguistic evidence about the ordinary (everyday) use of the expression "weakness of will." An analysis of ordinary language shows that "weak-will" is used in a variety of ways best understood as a cluster concept. The final section presents evidence from addiction studies and treatment practice supporting the practical efficacy of treating willpower failure as a cluster of related phenomena.

Complex vs. Reductive Definition

My central thesis is that there is no paradigmatic case or type of weakness of will. This leads some people to conclude that the concept must be either amorphous, useless, or both. This is not true, but it is an understandable reaction and so it is worth offering some prefatory remarks about definition. The standard model of class definition follows the Aristotelian logic of setting primary and secondary conditions for natural kinds. For example, a bachelor is an unmarried man. Therefore, to be a bachelor it is necessary that one be unmarried and a man. Meeting both conditions is sufficient to describe a person as a bachelor. This example is a stipulative definition and is the sort that is especially useful for

mathematics.⁶⁹ However, we are not looking for a stipulative definition, instead we want to provide an explicative definition--that is we want both to respect language as it is used and to offer some direction as to how language can be clarified and improved. As Carnap put it, "The task of making more exact a vague or not quite exact concept used in everyday life...We call this the task of explicating, or of giving an explication for, the earlier concept."⁷⁰

Mill wrote about language that it is "not made but has made itself: it requires continual mending in order to be passable."⁷¹ An explicative definition of weak-will is bound at least to some extent to the ordinary use of the term, but it is also expected to "mend" thinking about the concept. On Mill's view improving the messiness of ordinary language entails that,

"a meaning must be found for the name, compatible with its continuing to denote, if possible all, but at any rate the greater or the more important part, of the things of which it is commonly predicated. The inquiry, therefore, into the definition, is an inquiry into the resemblances and differences among those things: whether there be any resemblance running through them all; if not, through what portion of them such a general resemblance can be traced: and finally, what are the common attributes, the possession of which gives to them all, or to that portion of them, the character of resemblance which has led to their being classed together."⁷²

Mill's call for definitions to bring together characteristic resemblances does not rely on a concept like weakness of will having a single unifying thread. Rather, the idea

⁶⁹ Frege (1914) argued that only stipulative definition was appropriate for mathematics.

⁷⁰ Carnap (1956) *Meaning and Necessity: A Study in Semantics and Modal Logic* p. 7-8

⁷¹ Mill System of Logic, Bk. 1, Chapter 8 pg. 118 <http://www.gutenberg.org/files/27942/27942-h/27942-h.html#toc23>

⁷² Ibid. pg. 119

is that some concepts are only useful when understood through a cluster of related attributes.

By the mid-twentieth century some members of the scientific community were beginning to develop reasoning about classes along the lines Mill suggests. For example, Morton Beckner writes,

"A class is ordinarily defined by reference to a set of properties which are both necessary and sufficient (by stipulation) for membership in the class. It is possible, however, to define a group K in terms of a set G of properties f_1, f_2, \dots, f_n in a different manner. Suppose we have an aggregate of individuals (we shall not yet call them a class) such that

- (1) each one possesses a large (but unspecified) number of the properties in G ;
- (2) each f in G is possessed by large number of these individuals; and
- (3) no f in G is possessed by every individual in the aggregate."⁷³

Beckner is pointing out that natural class distributions are not always represented by the neat categories of formal systems. He describes such groupings as "polythetic" by which he means that they share a number of characteristics that occur commonly in members of a group or class, but none of which is essential for membership of that group or class. This is in contrast to monothetic groups which are identified by universally common characteristics. Beckner argues that the existence of polythetic classifications is a typically biological phenomenon; not standard for say geological classification, but normal for living systems. Eleanor Rosch extended the idea of polythetic classifications into cognitive science by introducing evidence that children form concepts about the world through

⁷³ Beckner, Morton (1959) *The Biological Way of Thought* Columbia University Press, New York p. 22

resemblances of features, not necessary and sufficient conditions.⁷⁴ More recent work has confirmed the resemblance method of classification or concept development and has shown that this model is also primary in other primates.⁷⁵

The key difficulty in polythetic classification is to set limits for what counts as a sufficient resemblance. In other words, how much resemblance must two things have to be counted as members of the same class? We will not come to a complete answer here, but the general strategy is to group classifications by strength. For instance, Rosch asked 200 students to rate, on a scale of 1 to 7, whether they regarded an item as a good example of the category *furniture*. The students responses reflected a range of association from the most furniture-like object 'chair,' to 'stool' which was less furniture-like than chair and to the least furniture-like object on the list, telephone.⁷⁶ Each object fits more or less in the class "furniture" so that the class is based on the degree of correspondence between members and membership is not strictly binary. Rather, some things fit in a given category well, while others do not, based largely on the ordinary use the classification.

However, it is important that the use of a polythetic definition does not entail that the ordinary use of a classification is unambiguously correct. It is possible to maintain that the common classification of a thing is incorrect, or, at least that the association of the thing and the class is weaker or stronger than has been commonly thought. The grounds for such linguistic "mending" might in some instances be

⁷⁴ Rosch E. and Mervis, C. (1975) *Family resemblances: studies in the internal structure of categories*, *Cognitive Psychology* 7, 573-605;

⁷⁵ Couchman, Justin J.; Coutinho, M. V. C.; Smith, J. D. (2010). "[Rules and Resemblance: Their Changing Balance in the Category Learning of Humans \(*Homo sapiens*\) and Monkeys \(*Macaca mulatta*\)](#)". *Journal of Experimental Psychology: Animal Behavior Processes* 36

⁷⁶ Eleanor Rosch (1975) *Cognitive Representation of Semantic Categories* *Journal of Experimental Psychology* v. 104:192-233

related to the need for stipulative categories. For example, if an institution requires separate budgets for fruits and vegetables it will need to class tomatoes in one group or another. In such cases the distinction will likely be purely nominal. More often, mending linguistic usage will have to do with the correspondence of a particular classification with some real impact. For example, there are many reasons to classify a potato as a vegetable, but it may be more effective in conveying nutritional impact to class potatoes with grains.⁷⁷

As a philosophical example of a polythetic category consider the concept 'critical thinking.' Critical thinking can be defined in some very general ways, but it is really an amalgam of independent elements that cluster around related activity. Consider the following definition, "Critical thinking is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action."⁷⁸ This sort of definition is helpful because we know that someone who is having difficulty reasoning to sound conclusions might benefit from the study of critical thinking. However, until we know what specific sorts of trouble a person is having all "critical thinking" help will take a shotgun approach. We still need to know whether the problem is with statistical arguments, game theory, etc. The broad

⁷⁷ 'Potatoes are botanically classified as a vegetable, but they are classified nutritionally as a starchy food,' Department of Health Spokesperson quoted at <http://www.dailymail.co.uk/health/article-1372728/The-great-potato-debate-Nutritionists-insist-spuds-ARE-vegetables.html#ixzz3Gkfb8u2u>

⁷⁸ Michael Scriven & Richard Paul, at the 8th Annual International Conference on Critical Thinking and Education Reform, Summer 1987

concept is useful because it points toward a range of narrow considerations, not because it is reducible to just one of those considerations.

Defining Weakness of Will

At this point it will not come as a surprise to hear that there is more than one way to be weak-willed. Definitions of 'weakness of will' are varied and have a long history in philosophy dating from ancient Greece where the term used was *akrasia*. The ancient account of weak-will turned on a distinction Aristotle made between weakness of will that occurred because an *akratic* agent was disconnected from his sensibilities and weakness of will that occurred while the akratic agent was in full command of his faculties. In the first type of case the akratic agent is overwhelmed by his passions, maybe he is angry, in love or perhaps just drunk; let's call this drunken *akrasia*. In the latter type of case the akratic agent sees things clearly, but nevertheless decides to do something that he knows he shouldn't; let's call this clear-eyed *akrasia*. I discuss these at length in chapters three and four respectively, but some preliminary discussion is important here.

Clear-eyed *akrasia* is the sort that Socrates had in mind when he describes acting in a way that one knows is less than best. The clear-eyed akratic recognizes the wrongness of her action at the time she acts. There is no confusion or ignorance at play for the clear-eyed akratic. Rather, her weakness of will is a matter of her own judgment about what it would be best to do not being sufficiently motivational. Clear-eyed *akrasia* seems odd because there is clearly a connection between judgment and motivation. When I judge that I ought to do something, I am judgmentally motivated to do it. However, being motivated by judgment does not

entail that that will be sufficient to lead me to act appropriately. In clear-eyed akrasia a judgment about what would *also* be good to do, though not the best thing to do, wins the motivational struggle.⁷⁹

Drunken akrasia is the sort of weakness that occurs when one is incapacitated or overcome in the way that occurs when drunk. Instances of drunken akrasia do not require that an individual is literally drunken. Rather, drunkenness is a metaphor for a motivational element distinct from one's judgment which leads to action through an independent process. The drunken akratic has some judgment about what would be best to do but is led away from that course of action by a process which subverts the agent's judgment. It is not that the agent has conflicting judgments, or that her judgment shifts, rather on this model something overwhelms or subverts her judgment. Of course, just as one can be responsible for drunken acts, one can be responsible for weak-willed acts that are caused by a culpable or predictable subversion of judgment.⁸⁰

The sorts of structures or entities required to make good on drunken or clear-eyed akrasia are also plural. In the case of drunken akrasia there are several common pressures described as subverting an agent's judgment about how to act and thereby leading to weakness of will. For instance, a heroin addict may plausibly judge it best not to use, but be overcome by her addiction. A parent may judge it best to have her children brush their teeth, but under the sway of several glasses of wine forget to enforce the rule. A priest may plausibly judge it best to help his fellow man, but running late for a meeting fail to recognize an opportunity. What these

⁷⁹ I discuss the possibilities and issues surrounding clear-eyed akrasia at length in chapter three.

⁸⁰ I discuss the issues surrounding freewill and culpability at length in chapter four.

cases have in common is that an agent judges it best to do one thing, and yet does another. The weakness in these cases comes not from an effect on judgment per se, but from some other source bypassing judgment.

Cases of clear-eyed akrasia require a closer look at the mechanisms responsible for an agent's judgment itself. The most basic form of this is sometimes called the Humean account of judgment. On the Humean model judgment consists of a combination of belief and desire. An augmented version of this account includes intentions, which typically result from the combination of beliefs and desires and lead to action. The clear-eyed akratic might, as Al Mele suggests, form a decisive better judgment based on all relevant beliefs and desires and yet fail to form the appropriate intention. On Mele's view the motivational force of lesser beliefs and desires may act directly on the intention. This would cause weakness of will because the agent would assert that her better judgment suggests one course of action and yet be most motivated toward another.

Among philosophers it has become fashionable to distinguish between "weakness of will" and *akrasia*. *Akrasia* is a Greek word meaning incontinence, or the inability to control one's excretions. It also has a broader connotation of a lack in self control. Traditionally, *akrasia* was translated as weakness of will. The philosophical distinction en vogue at present is to describe *akrasia* as a peculiar, perhaps possible, curiosity of motivational failure, and to describe weakness of will as a common form of human fallibility. The basis of this distinction is that *akrasia* requires one act against her own beliefs while weakness of will requires that one

temporarily revise her beliefs despite a commitment not to do so. Historically, these would both have been called weakness of will.

The distinction between *akrasia* and weak-will is based largely on Richard Holton's innovative thinking about motivational failure. On Holton's view weakness of will is best characterized by a failure to act on one's resolutions, one's commitment to future action that one recognizes will be motivationally challenging. Hence, on his view, resolutions are "contrary inclination defeating intentions."⁸¹ A contrary inclination defeating intention is just an intention designed to address certain anticipated motivational difficulties. If, for example, Charlie forms the intention to stop using cigarettes and does so knowing that he will surely have an inclination to smoke, his intention is designed to defeat the inclination to smoke.

Of course, it cannot be the case that just any reason for failing to follow through with an intention is weak-willed. Intentions broadly considered are controlling in the sense that they standardly lead to action, but even stable intentions are open to revision based on highly significant features of the emerging situation. For example, if Charlie were to discover that smoking another cigarette would actually improve his health he would not be weak-willed for revising his intention. Simply reconsidering an intention in light of important new information is not something Holton wants to call weakness of will; rather he argues that only changes of resolution due to certain causes count as weak-willed. Specifically, one is weak-willed on Holton's view only if she revises her intention based on reasons her resolution was designed to defeat. Thus, on Holton's view, Charlie is weak willed

⁸¹ Richard Holton (2009) *Willing, Wanting, Waiting* (Oxford: Clarendon Press) p. 250

only if he violates the intention to quit smoking due to contrary inclinations that his resolution was designed to defeat.

One way to understand Holton's view is to contrast it with the well-known, and more traditional approach taken by Al Mele. Mele's view is described as "strictly" akratic to distinguish it from earlier accounts of akratic action, especially Donald Davidson's. Where Davidson described weakness of will as possible when so called *prima facie* judgments led to action at the expense of "all things considered" judgments, Mele maintains that weak-willed actions can be violations of unconditional judgments. As Mele puts it, strict akratic action requires an act be "performed intentionally and freely and, at the time which it is performed, the agent holds a judgment to the effect that there is good and sufficient reason for his not performing it."⁸² Thus, Mele maintains that weakness of will occurs when one violates an all out judgment and thus knowingly and freely peruses a lesser path.

Consider the following example to understand the contrast between Holton's and Mele's views. An imaginary Benjamin, a friend of Charlie's who also believes it would be best to quit smoking. On New Year's Eve Ben and Charlie are talking about their resolutions and they both express a genuine belief that they should resolve to quit smoking. However, whereas Charlie affirms his belief and forms a resolution to quit, Ben demurs quietly acknowledging that he should form a resolution but steadfastly declining to do so. Mele of course allows that Ben is weak-willed, but on Holton's view only Charlie is weak-willed, Ben is something else—akratic. The difference between the two is that Ben acts against his own

⁸² Al Mele (1987) *Irrationality* (OUP) p. 7

better judgment, while Charlie revises his judgment despite his resolution not to do so. Holton's aim to exclude cases of strict akratic action from the realm of what is called weak-willed is something of a revision of how anglophone philosophers have ordinarily used the term weak-willed. Because of this linguistic revisionism and because Holton's account of weakness of will depends on the revision of a resolution I deem Holton's view "revisionist weakness of will."

Holton offers both empirical and theoretical arguments in support of the primacy of revisionist weakness of will. The most persuasive theoretical argument is that the revisionist account can explain cases of weakness of will that apparently do not involve an agent's having inner conflict. Since the traditional account of strict akratic action can only accommodate cases of weakness of will involving inner conflict, if cases exist without such conflict, then the revisionist case appears plausible. Consider an example: James believes that he ought to stop smoking and resolves to do so. James knows that quitting will be difficult as he has tried to do so on many previous occasions. In the past he has always kept right on smoking when he has tried to taper off his consumption and only met with some success, three months without smoking, when he quit cold turkey. James knows that going cold turkey will be tough, especially for the first several days, but decides that it is his best bet. However, only twelve hours after waking up James finds himself desperate for a cigarette and decides that what would really be best is for him to gradually reduce his cigarette consumption, of course, this change allows James to smoke a cigarette now. The trouble with the traditional account of weakness of will as strict akratic action is that James does not appear to be weak-willed. If we take James's

change of mind seriously, that is, if we presume that he really has changed his mind about what method would be best, his behavior is not strictly akratic and thus not weak-willed. This is troubling because it seems that the individual who decides to quit smoking and then blithely changes his mind when the going gets tough is being weak willed.

However, the explanatory advantage that revisionist accounts of weak will enjoy come at a price. The revisionist account depends on a murky answer to the question "What constitutes an *unwarranted* revision of a resolution?" On one hand resolutions are a special kind of intention designed specifically to resist (unwarranted) revision and it would seem that they are only useful to the extent that they do resist revision. On the other hand blindly remaining on any given course of action is a recipe for disaster, new situations and information arise with great regularity and it is foolish not leave open the possibility that one's previous judgments were in error. Holton follows Bratman maintaining that it is "rational to reconsider an intention just in case doing so manifests tendencies that it is reasonable for the agent to have."⁸³ The tendencies that Holton has in mind here seem to be related to recognizing new and important information gleaned from emerging situations. This sort of response initially seems persuasive, but it is somewhat too vague to be of real use, as Holton himself admits.⁸⁴

⁸³ Holton (2009) Ch 4, p. 3

⁸⁴ Holton describes this level of response as "horribly vague." Ch 4, p. 4
Holton's rules of thumb for warranted resolution revision include revising "if one believes that circumstances have changed in such a way that they defeat the purpose of having the intention...if they [the intentions] can no longer be carried out...they will lead one to great suffering when that suffering was not envisaged."⁸⁴ He also offers several rules of thumb for when it is reasonable not to reconsider a resolution, including "in circumstances that prevent clear thought if those intentions were

Holton's advice then is to err on the side of non-reconsideration. His suggestion is that barring rather extraordinary circumstances one ought to refrain from even reconsidering one's resolutions. This is a powerful message and one carried widely and echoed loudly by the various "twelve step" programs, AA, NA, GA, etc. These programs require that one "turnover" one's life and trust the program. Rather than depending on the individual's judgment, which they suggest is deeply flawed, these programs admonish the individual to "work the steps." The value in this advice is that the individual need not make decisions about how to proceed and thus has apparently fewer opportunities for failure. The reasoning is that with fewer decisions made about a difficult topic, there are correspondingly fewer occasions for weakness. This is perhaps good practical advice in some situations, but, this solution to the problem of akratic resolutions merely moves the trouble back a level. A new problem arises in deciding when reconsideration is reasonable. Holton rightly says that when the reconsideration of a resolution is rational will be heavily dependent on the type of resolution, the individual's dispositions, situation, etc. and the matter is likely to remain a difficult matter of practical psychology. This is challenge particular to Holton's account and is not mirrored in the more traditional account of strict akratic action.

This history supports the earlier point about the nature of definition: there is no single paradigmatic version of weak will. Rather, the concept has historically involved several types and mechanisms. While there is no one core feature of all

made in circumstances that allow clear thought [and] intentions that were expressly made in order to get over one's later reluctance to act."⁸⁴

instances of weakness of will, it is possible to isolate some of the more prominent features. Weakness of will tends to involve a separation of judgment and motivation. The thing that one thinks he ought to do, or what he thinks he ought to do when he is thinking well, is not the thing that ends up happening. A central characteristic of many cases of weak-will is thus a sense of regret, or a sense that one has failed to live up to one's own expectations. Of course, self-deception can be quite complete in some instances, and passionate failures are often easy to overlook, so it is often the case that we do not notice our own weakness. The point is just that weakness of will involves failing to live up to our own judgments, sometimes synchronically, sometimes in retrospect and sometimes we never notice our failure at all. Classing these different events together is a historical reality in philosophy, and turns out to correspond with ordinary language as well.

Weakness of Will: The Ordinary Use

In addition his theoretical arguments for the importance of revisionist weakness of will Holton offers that “Whenever I have asked non-philosophers what they take weakness of will to consist in, they have made no mention of judgments about the better or worse course of action. Rather they have said things like this: weak-willed people are irresolute; they don't persist in their intentions; they are too easily deflected from the path they have chosen” (1999, p.241). Holton concludes that strict akratic action is less common and warrants the relatively foreign term *akrasia*. Albeit anecdotal, Holton's empirical claim has led to a growing body of substantive research about the expression 'weakness of will' in the ordinary use of language.

For instance, Al Mele conducted a series of 4 surveys based on the ordinary use of weakness of will by Florida State University undergraduates. Mele's first survey asked 72 subjects Holton's question "What is weakness of will?" only one mentioned Holton's revisionist view, while 11 mentioned Mele's traditional view.⁸⁵ In a second study Mele asked 119 subjects to answer the following question:

We're interested in what the expression "weakness of will" means to you. Please answer the following question by circling your answer. Which of the following descriptions of weakness of will is more accurate in your opinion?

- A. Doing something you believed or knew you shouldn't do (for example, going to a party even though you believed it would be better to stay home and study).
- B. Doing something you decided or intended not to do (for example, going to a party even though you decided to stay home and study).
- C. Neither. The descriptions are equally accurate or inaccurate.

In response 49% of subjects gave the "believed/knew" response Mele correlates with his traditional view, 33% gave the "decided/intended" response Mele ascribes to the revisionist view and 18% gave the third response. (2009, p.6) Mele suggest that on their face studies 1 and 2 seem to suggest that the traditional view is more ordinary than the revisionist account.

Joshua May suggests that we ought to heavily (if not entirely) discount studies 1 and 2 based on the unreliability of theoretical definitions offered by ordinary people in brief experimental conditions. He is probably right. People are notoriously bad at providing theoretical definitions for concepts with which they can competently use. However, research indicates that people are significantly better at

⁸⁵ Arguably 13 others described the traditional view, but Mele discounts them for various reasons in a footnote p.6.

identifying good theoretical definitions than providing them, think about understanding a foreign language compared with speaking it. So although we might heavily discount study 1, Mele's study 2 seems to provide some compelling evidence that the traditional view is more common than the revisionist view.

Mele's third and fourth studies satisfy worries about theoretical abstraction (at least somewhat) by offering some context within which subjects were asked to evaluate the presence of weakness of will. In the third study 25 subjects were given the following story:

Joe believes that it would be best to quit smoking cigarettes. He is thinking again—this time on New Year's Eve—about when to quit. He knows that quitting will be hard and unless he picks a good time to start he will fail. Joe judges that it would be best to smoke his last cigarette tonight and to be smoke free from then on. When he reports this to Jill, his wife, she asks whether this is his New Year's resolution. He says, —Not yet. I haven't yet actually decided to quit. Making that decision will be hard. To make it, I'll really have to psych myself up. I've been smoking for forty years. I believe I can quit, but I would definitely miss smoking. In the end, Joe fails to decide to quit smoking. Tomorrow, he smokes less than usual, but he has his first cigarette minutes after he awakes, as always. However, he could have decided to quit, and if he had he would have quit.

Subjects were given a 7 point scale with 1 labeled "strongly agree" and 7 labeled "strongly disagree" and asked to report their degree of agreement with the statement "Joe displays some weakness of will in this story." In the story Joe does not make a resolution to quit smoking so if Hotlon's revisionist view is ordinary we

should expect to see results tend toward 7 (Strong disagreement with the attribution of weakness of will.) However, the response average was 2.68 and 80% of subjects answered between 1-3; this seems to indicate agreement with an attribution of weakness of will in keeping with the traditional view.

Similarly, in his fourth study Mele presented 100 subjects with the same story about Joe and asked them to circle yes or no to answer the following question: “Does Joe display any weakness of will in this story?” Of the 100 subjects, 73 answered ‘yes’ (attributing weakness of will) and 27 said ‘no.’ Mele’s explanation for this is that Joe’s judgment that it would be best to quit combined with his failure to quit indicate a traditional form of weakness of will. Thus, Mele’s studies 3 & 4 apparently reveal a strong (80% and 73% respectively) ordinary use of weakness of will in the tradition fashion.

However, May brings up two reasonable objections to Mele’s interpretation of the data. First, May argues that Joe’s situation might be reasonably interpreted as involving a resolution-violation. On May’s proposed reading Joe has not resolved to quit, but he has “resolved to decide to quit.” (May, p.6) If Joe actually has formed a second order resolution to resolve to quit it is violated by his not actually resolving to quit. In such a case those who attribute weakness of will could be doing so on the basis of the revisionist view of weakness of will. I have several problems with this objection. First, there is little (if any) reason to think that Joe has resolved to form a resolution to quit for the New Year, he seems to be merely toying with the idea of quitting—thinking about what it might take and whether now is a good time. In the event that one focuses on Joe’s telling his wife “Not yet” when asked if he has

resolved to quit, one might think that Joe has formed a second order resolution. This is also supported by the early claim that Joe is thinking “about when to quit.” Note that Joe is not thinking about if he should quit. Apparently Joe has resolved to resolve quit, but, importantly, his second order resolution does not set a time for quitting. Thus, Joe’s failure to resolve to quit on New Year’s day is not a revision of his second order resolution and should not produce false positive results in Mele’s study. Nevertheless it is always possible that people have read Mele’s example as May thinks they might.

May’s second objection to studies 3 and 4 is on the basis of the relative weakness of the claims being tested, namely, “Joe displays *some* weakness of will in this story” and “Does Joe display *any* weakness of will?” It is, of course, easier to get some to agree that “Joe displays *some* weakness of will” than that “Joe displays loads of weakness of will.” May’s view is that the weakness of Mele’s attribution requirement accounts for the 80% agreement found with statements about Joe’s weakness of will. This is almost certainly right, but may be beside the point. All Mele really needs to show is that a large number of people think that Joe’s case is an ordinary use of the expression weakness of will and even low amounts of weakness of will should qualify as appropriate subjects of the term.

As an alternative to Mele’s studies 3 and 4 May suggests relying on Mele’s footnoted 3a in which he provides Joe’s story and the prompt “Joe does *not* display *any* weakness of will.” Subjects responded to 3a with about 58% disagreement (attribution of WOW) and 38% agreement. May contends that 3a is likely a more accurate reflection of ordinary usage because the attribution requirement is

comparatively strong. I disagree. If Joe displays trace elements of weakness of will one should be equally willing to both, agree that he has *some* weakness of will, and, disagree that he does *not* have *any* weakness of will. Affirming that a person, p, has *some* WOW is logically identical to denying that p does *not* have *any* WOW. The gap in results between Mele's 3 and 3a still warrant explanation, but something further is required.

If we take Mele's results from studies 2-4 as I believe we ought to we would see them as reliable small scale studies about the way ordinary people use the expression weakness of will. On the basis of these studies I think that it is safe to preliminarily conclude that Holton was wrong about our ordinary use of the expression weakness of will in at least one sense. The traditional view of weakness of will as intentional action against one's better judgment captures well an ordinary use of the term. However, Mele's studies say very little about other ordinary ways that weakness of will is used and some (2 and 3a) suggest that the revisionist view of the expression 'weakness of will' is also appropriate in ordinary use.

In a follow-up study May surveyed 97 people in and around the University of California, Santa Barbara and randomly assigned them to one of four conditions (stories) that follow:

1. "*Newman's Diet*" involves Newman, an agent who is weak-willed in both the traditional and revisionist sense. Newman believes it best to go on a diet, resolves to do so, and despite some apparent success succumbs to temptation and eats one each of his favorite doughnuts.
2. "*Christabel's Affair*" involves Christabel, an agent who is weak-willed in the traditional but not the revisionist sense. Christabel is a married woman who believes it best not to have an affair, but who resolves to have an affair anyway and having so resolved begins an affair with William, a man who is not her husband.

3. *“Rocky’s Loss of Nerve”* involves Rocky, an agent who is weak-willed in the revisionist but not the traditional sense. Rocky has promised his mother he will not play tackle football and thinks it best not to play yet he resolves to play anyway. When the time comes Rocky loses his nerve and violates his resolution but acts as he thinks best.

4. *“Kima’a Affair”* involves Kima, an agent who is not weak-willed in either the traditional or the revisionist senses. Kima doesn’t much care about her husband’s feelings and is attracted to Omar, Kima seduces Omar and they have an affair.

Using a seven point scale subjects were asked to rate their agreement or disagreement (this time with 7 being strongly agree and 1 being strongly disagree) with the following questions based on the assigned condition: “Newman displays weakness of will in eating the doughnuts. Christabel shows weakness of will in having the affair. Rocky displays weakness of will in not showing up for the game. Kima displays weakness of will in having the affair.”

May’s analysis of the results focuses on the average response between 1 and 7 for each of the four conditions. In condition one the average response was 5.57, meaning that on average people agree that Newman exhibits weakness of will in eating doughnuts when he thinks it best to lose weight and has resolved to do so. In condition four the average response was 3.08, meaning that on average people slightly disagree with descriptions of Kima’s affair (which violates neither her judgment nor her resolve) as weak-willed.⁸⁶ In condition two the average response was 4.29 and in condition three the average response was 4.19, this means that on

⁸⁶ Considering the fact that neither contender for weakness of will are present in this condition 3.08 should be an unexpectedly weak disagreement with the attribution of weakness of will. May attributes this unusual result to subject’s desire to attach stigma to Kima’s affair and/or an inclination to accept propositions generally. I would add that Kima’s affair likely violates a commitment to monogamy culturally implied by her marriage and might plausibly be read as the unwarranted revision of a resolution.

average Christabel's failure to act as she judged best and Rocky's failure to do as he resolved are viewed only very mildly as weak-willed.

The results of May's study suggest that the ordinary use of weakness of will is consistent with either Holton or Mele's view, with corresponding test conditions averaging 4.19 and 4.29 respectively. This might make it seem as though Mele's account is more in keeping with ordinary use than Holton's, but the slight distinction here is not statistically significant. It seems Holton's claim that the ordinary use of the term weakness of will is exclusively revisionist is off the mark, but perhaps more interesting is the strength of the responses. After all, 4.19 and 4.29 are only fractionally higher than a neutral response of 4.0. The relatively strong 5.57 average response in the case of Newman's Diet provides good reason to think that weakness of will is more strongly attributable in cases which involve violation of both intention and judgment. This leads May to conclude that "both variables are necessary, neither is sufficient." (May, p. 13) Thus, May describes the ordinary concept of weakness of will as a conjunctive cluster concept, requiring both violation of intention *and* evaluative judgment.

The notion that violations of both intention and judgment in a single case contributes to a stronger attribution of weakness of will than either one alone should not be surprising. If anything what surprises is that the attribution is so low, only 5.57, one wonders 'if Newman isn't weak-willed then who is?' However, the 5.57 average response is somewhat misleading. May reports that 74% of subjects agreed that Newman displayed weakness of will (providing answers between 5 and 7). Although May does not supply the exact figures 74% agreement with an average

of 5.57 suggests that those who did not agree were often on the fence (around a 3), and, those 74% tended to agree rather strongly (averaging around 6). This suggests that nearly everyone was at least somewhat comfortable with a description of Newman as weak-willed.

If we consider the percentages rather than average responses in conditions 2 and 3 we find that 50% of responses agreed with an attribution of weak-will (between 5 and 7), around 30% disagreed (between 1 and 3), and around 20% were neutral (4). This means that in each case around 70% of people would not object to an attribution of weakness of will and half would agree with such an attribution.⁸⁷ If 70% of people would not object to the attribution of weak-will to Christabel or Rocky, we should think that either condition would be sufficient for attribution, though neither should be considered necessary.

Of course, one might object to characterizing the neutral 20% as not objecting to attribution. If we suppose that the study treats weak-will as a binary property—it exists or it does not (in whatever degree)—then a neutral response should be seen as a disavowal of knowledge about weakness of will in the case at hand. On this view the neutral subject doesn't know if a weak-will is displayed or not and so would object to both attribution and denial of weakness of will. In any event 50% agreement with attribution represents a vast majority when compared with two other candidates garnering 20% and 30% respectively. It seems that ordinary people are 66% more likely to agree with attributions of weak-will than they are to disagree with

⁸⁷ Here I presume that neutral responses represent a neutrality about how strongly weak-willed the characters in question are. This presumes that subjects interpret the survey in a way May specifically aims to avoid, but one which I think is quite natural given the survey.

such attributions and they are 150% more likely to agree than they are to declare some form of neutrality. Construed in this way the results seem to indicate that ordinary people find either unwarranted revision of a resolution or violation of one's better judgment a sufficient condition for attribution of weakness of will. Thus, I propose that we consider weakness of will a cluster concept, but not a conjunctive concept with two necessary conditions. Rather, if May's findings were representative of all people, then weakness of will seems to be a cluster concept along the lines of an inclusive disjunction; both unwarranted revision of a resolution and violation of one's better judgment are sufficiency conditions, neither is necessary and the concept is strongest when both are included.

Thus far the research into ordinary use of the term weakness of will has been restricted to the traditional and revisionist interpretations. If the preceding account is correct we should see both traditional and revisionist accounts as sufficient to warrant attribution of weakness of will on the ordinary account. Further, we should think it especially appropriate to describe behaviors meeting both conditions as weak-willed. However, there is a prominent third perspective on weakness of will offered by Gary Watson, which warrants further testing. Watson describes weakness of will as a normative judgment that one has failed to develop the capacities of resistance commonly expected of competent adults. On Watson's account weak-willed agents are psychologically incapable of resisting the urges to which they give in but could have resisted if only she had developed normal capacities of resistance and thus the weak-willed agent is culpable for their compulsion in a way that merely compulsive agents are not.

I think that it is quite natural for people to use the expression weak-willed in the sense of normative failure that Watson has in mind.

Weakness of will should be thought of as a polythetic concept along the lines of an inclusive disjunction—that is in ordinary language weakness of will has many conditions none of which is necessary. Each of the following: unwarranted revision of a resolution, violation of better judgment, and the failure to develop normal capacities of resistance, are sufficient to warrant the ordinary use the term weakness of will. However, none of these are necessary—one could rightly be called weak-willed despite lacking any one or two of these conditions. Of course, there is likely to be much debate about what revisions are *unwarranted*, in what *better* judgment consists, and what constitutes a *normal* capacity for resistance, but that these questions all pertain to what we ordinarily call weakness of will seems clear.

Addiction Research and Weak-Will

As described earlier polythetic classes are sometimes used to offer explicative definitions in an effort to both conform to ordinary language and to help shape or clarify concepts. So far we have seen that there is strong evidence suggesting a cluster view of weak-will including revisionist akrasia and strict akrasia. But, is there any reason to think that the historical and ordinary uses of the concept 'weak-will' are problematic or somehow less useful than an alternative definition might be? To the contrary I believe that there is good reason to think that the philosophical and common understanding of weak-will is useful *because* of its breadth and not in spite of it.

To explain why I think this account of weakness of will is useful I want to briefly review the research on addiction and show how it illuminate the nature of addiction and gives us some guidance on how to best treat it. I do not mean to suggest that addiction and weakness of will are identical, but the two are sufficiently related that it appears that research on addiction could illuminate the concept of weakness of will. This final thrust toward a polythetic conception of weak-will involves two elements a review of some of the various theories about addiction and an account of the success rates of treatments based on these views. In this section it becomes clear that there is a significant relationship between addiction and weakness of will. The treatment success rates also reveal that addiction treatment is most successful when addiction is treated as a range of problems and that narrowing out conception of addiction decreases the success of preferred treatments. If we can extrapolate this evidence to weakness of will, it suggests that a broad account of weak-will is more useful.

As far as we can tell people have been drinking alcohol and taking drugs forever, but early thinkers tended to attribute substance abuse to immoral choices stemming from of sinful appetites. This sort of moral approbation did little to advance thinking on substance dependence, but it provided a singular and simple explanation of a set of relatively unusual behaviors. Treatment consisted of moral rectification, usually through the auspices of the church. Become a better person and your drinking will stop. This model of addiction remains prominent in today's 12 step programs, which continue to treat addiction as a moral failure which can only be resolved through the grace of God.

Unsurprisingly, the success rate of twelve step programs such as Alcoholics Anonymous (AA) is not high. According to one internal report the typical success rate of those enrolled in such programs is about 5%.⁸⁸ Compare this to the spontaneous recovery rate of addicts of about 5% and it begins to look like AA does not have a measurable effect on recovery.⁸⁹ Perhaps worse, there are some studies that indicate that AA's commitment to an addict's inability to control his or her addiction may lead to worse results than no treatment program at all.⁹⁰ Nevertheless the twelve step approach is appealing in its simplicity: the addict's sinful nature has failed her and salvation is possible only through communion with a higher power.

Fortunately, beginning in the 19th century and paralleling the expansion of 12 step programs there has also been an explosion in research on addiction as a medical phenomenon.⁹¹ This research has tended to move along two branches of thought. First, there is a robust literature on the psycho-social motivations toward addiction. This literature investigates the ways that psychological trauma such as childhood abuse, violent participation in war, etc. might cause substance abuse . It

⁸⁸ George Vaillant (1995) *The Natural History of Alcoholism Revisited*, (Harvard University Press; Cambridge, MA).

⁸⁹ R. G. Smart calculated that the spontaneous remission rate for alcoholism was between 3.7 and 7.4 percent per year. For supporting evidence see Charles Bufe, (1998) *Alcoholics Anonymous: Cult or Cure?* (Sharp Press, AZ)

⁹⁰ Brandsma et al., *The Outpatient Treatment of Alcoholism: A Review and Comparative Study*, Baltimore: University Park Press, 1980 and Ditman et al., "A controlled study on the use of court probation for drunk arrests," *American Journal of Psychiatry*, 124:160-163, 1967.

⁹¹ Because of the stark shift from very limited discussion of alcohol and narcotics abuse before the 19th century to wide confrontation with such abuse some believe that addiction is a modern invention. On such accounts substance addiction is a product of the hostility and desolation of modern western life, or, a faux difficulty exaggerated in an effort to obscure the real sources of social and personal malaise. To my mind these critiques are important in that they reveal the importance of our emergence into modernity for addiction. In this emergence we confront the death of God and what were once sacraments become pedestrian. But we should be careful not to suggest that the historical etiology of addiction fully explains its character or mitigates its harms. Addiction is very real despite the fact that it is socio-historical construction.

also identifies correlations between social and environmental cues and addictive behaviors. This body of research suggests that a substantial number of addicts are engaged in a process of self-medication and that significant numbers have trouble with addiction only in certain circumstances.

Of course, addiction doesn't have to be associated with PTSD or triggered by particular social occasions or environments. The psycho-social aspects of addiction take many forms, but carefully attending to the particulars of this type of impulse can lead to effective strategies for resistance. For example, George Ainslie (1992) and Jon Elster (1984) report that the formulation of specific private rules or precommitment strategies have been effective in resisting addictive behaviors. It has also been reported that elevation of circumstance from poverty to relatively comfortable living standards often has the effect of reducing addictive tendencies. Importantly, these strategies require specific knowledge about the triggers and dispositions of specific addicts; they are not a one size fits all approach to recovery.

The other main current of successful addiction treatment research has been the study of brain states and the effects of specific substances on the physical and chemical functions of the brain. Research along this line has produced a number of valuable and effective pharmaceutical treatments of addiction. One particularly promising therapy is use of the drug *Naltrexone* which blocks the high associated with alcohol and opiates. In clinical trials those taking Naltrexone were twice as likely as those on placebos to abstain from alcohol consumption. However, Naltrexone is useless in preventing cravings associated with cocaine or nicotine, and

may even make them worse. The point being that there is no one brain state associated with all addictions, and no one chemical which can treat them.

The fifth edition of the Diagnostic and Statistical Manual recognizes the complexity of addiction and describes substance abuse as meeting three out of seven criteria within a one-year period. The criteria include tolerance, withdrawal, overuse, unsuccessful efforts or desire to cease use, spending a significant amount of time finding, using or recovering from the substance, reeducation of important social, occupational or recreational activities because of use, and continued use despite having knowledge of harms caused by continued use. This is hardly a single marker and as it stands the DSM IV includes only psycho-social elements, none of the neurological elements associated with addiction are listed. This definition captures the part of the complexity of addiction by defining a cluster of conditions a number of which are sufficient and none of which are necessary. Robert West (2006) goes even further calling for a “synthetic theory of addiction that recognizes that addiction can involve any or all aspects of a motivational system that involves fine levels of operation.” (2006 p.192) In other words, West proposes a broad analysis, including internal, drug based, and external motivational stimuli to account for addictive behavior.

The result of admitting to a complex definition of addiction multiple elements of brain function and behavior is that there is conceptual space for a variety of treatment options. It is of course much simpler to suppose that one explanation and one solution are appropriate to everyone, but this is simply not the case. As addiction research has accepted the complexity of addiction and moved away from

simple moral condemnation it has become more effective at understanding and treating addicts. This is an important lesson for the related philosophical inquiry on weakness of will. Not only do we get closer to the truth of people's experiences by conceptualizing weakness of will as a cluster concept, but we also move closer to overcoming a troubling phenomenon in people's lives.

Conclusion

I have attempted here to show how addiction research and a little bit of experimental philosophy can combine to shed light on a longstanding trouble in the history of philosophy. The account here is aimed at urging others to consider the real possibility that the right way to look at weakness of will is as a complex cluster concept. The aim is to move philosophical thinking about weakness of will closer to the medical view of addiction. If this transition is successful, it will mean in part that we treat weakness of will as reflecting the myriad ways that people take action and not as a possible key to uncovering the one way that practical judgments are made.

It is also worth noting that correlating weakness of will with the medical model of addiction may have some practical advice to offer those of us who occasionally watch TV when we should be grading, or who decide to contribute to Oxfam and then do not, or even to those of us who want to be vegetarians but cannot seem to make the step. We need not think that these failures all come from the same sort of difficulty, nor must we be committed to the claim that each of them has the same treatment. I have not argued for this thesis here, but I suspect that attending to the particular nature of each problem and finding specific solutions will be the best way to make real moral progress.

Chapter 4

Freedom and Weakness of Will: The Compatibility Problem

A central attack against those who regard weakness of will as possible has been to argue that weakness of will is better understood as a form of coercion. I want to show that defenders of weakness of will need not get bogged down in debates over the existence of free will. First, I will briefly show that those who deny the existence of free will will also deny the existence of weakness of will, but since they do not offer any special reasons to object to WOW, we can set that aside for the purposes of this thesis. Then I show any plausible account of free will is compatible with the notion of weakness of will. I do not endorse each of the accounts of freedom described here, in my view some are quite suspect. Nevertheless, these are the main accounts of freewill and I show that each is compatible with the existence of weakness of will.

The “Consequences Argument”

Pierre Simon Laplace writes,

“We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it

would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes."⁹²

Laplace's demon is important because it would seem to undermine our ordinary belief in free will. Of course the actual existence of such a demon is not important, the force of the argument rests on the possibility of understanding all future events completely through earlier ones. If the history of the physical world leads inexorably to definite future events, the world is deterministic and determinism is at odds with our ordinary conception of freewill. Freewill is usually understood as incompatible with all of our actions having been determined by some earlier state.

Peter Van Inwagen describes determinism in simple terms as "the thesis that the past *determines* a unique future."⁹³ On Van Inwagen's view the truth of determinism is incompatible with freedom and responsibility because determinism does not allow agent's to cause events. Van Inwagen describes his central argument for incompatibilism as the "consequence argument." The consequence argument proceeds as follows: "If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of those things (including our present acts) are not up to us."⁹⁴ The meaning of an action being "up to us" is somewhat unclear, but at root Van Inwagen believes that it requires an event not be caused by any

⁹² Laplace, Pierre (1814 (Translation 1902)). *A Philosophical Essay on Probabilities*. London: John Wiley & Sons.

⁹³ Van Inwagen, Peter (1983) *An Essay on Free Will*, (Oxford: OUP) p. 2

⁹⁴ *Ibid.* p. 16

other object, i.e. physical process. The general challenge to freewill comes from the possibility of automatic physical processes, determinism, undermining the possibility of agents effectively making decisions about events.

Skepticism About Freewill

Both determinists and causal indeterminists are prone to be skeptical about the possibility of genuine human agency, because they are often skeptical about the possibility of human freedom. The argument from determinism or causal indeterminism to skepticism about free will is relatively straightforward. Skeptics argue as follows: If all human action is caused by antecedent events, people cannot cause actions.⁹⁵ (Here causing an action is taken in the ordinary direct sense. In some sense human actions are all a part of antecedent conditions, but this “butterfly effect” is not what we ordinarily mean when we talk about causing an action.) All human actions are caused by antecedent events. So people cannot cause actions. If people cannot be the ultimate cause of actions, then on some prominent accounts they are not free willed agents.⁹⁶ Therefore, people are not free willed agents.

Skeptics about free will are plentiful and they make a compelling case in physicalist terms, but they are not the primary audience for a defense of the possibility of weakness of will. Those who are skeptical of free will are skeptical of all intentional human action, not just one particular variety. We should expect that skeptics of free will would be skeptical of weak willed action, but only as a much larger subset of action generally. Philosophers who argue that purported cases of

⁹⁵ Further along I address the possibility of regional as opposed to global determinism.

⁹⁶ Call this the “agent-causation” view of free will. This view is discussed in greater detail further along.

weak willed action are really compulsive or unfree suppose that other actions are free—this is the force of their claims against the possibility of weakness of will. So for our purposes here we will focus on accounts of free will without undue concern for addressing general skepticism about freedom. If it turns out that free will has been an illusion all along, then the possibility of weak willed action is the least of our concerns.

Two Theories of Free Will

Among theories of free will there are two basic varieties, libertarianism and compatibilism. Libertarianism about freedom should be sharply distinguished from the political ideology by the same name—the two are not related. Libertarianism about free will, sometimes called metaphysical libertarianism, is perhaps the most common pre-theoretical view. Proponents of libertarianism maintain that people are free in a way that allows for moral responsibility and that this freedom is incompatible with either determinism or causal indeterminacy (and also randomness). The compatibilist agrees with the libertarian's first belief, which is that people are free in a way that allows for moral responsibility. But, the compatibilist rejects the libertarian's second belief, that freedom entails the falsity of determinism and casual indeterminacy. For this reason compatibilism is sometimes called "soft determinism."⁹⁷ On the compatibilist view, human freedom is compatible with determinism (although it need not embrace determinism,) while on the libertarian's view freedom and determinism are incompatible.

Libertarian (Incompatibilist) Freewill

⁹⁷ James, William (1884) "The Dilemma of Determinism," republished in *The Will to Believe* (New York, Dover 1956) p. 149

The strength of libertarian incompatibilism begins by accepting some forms of determinism as compatible with some forms of freedom. For example, being imprisoned is a form of coercion or un-freedom which is compatible with the opposite. Which is to say it is possible for some people to be unfree in the sense of being imprisoned while others are free in the sense of not being imprisoned. So, some forms of freedom and un-freedom are compatible on all accounts. But, the libertarian asserts that there are some types of freedom which are incompatible with the truth of determinism, we should think of these as the central cases of contention in the dispute about free will.

The type of freedom which is incompatible with determinism is present in the future being “up to us.”⁹⁸ At its core Libertarian free will requires that an agent genuinely have alternative possibilities. Having *alternative possibilities* simply means that although an agent acts in one way the agent can do otherwise or has the *power or ability* to do something different.⁹⁹ The principle of alternative possibilities holds when a person is free to choose between acting on alternative courses of action—this indicates that freedom involves a counter-factual ability or power to act in a way that is different from the way one actually acts. This is sometimes referred to as the ability to have done otherwise. On this condition, an agent’s action is free if and only if it is possible that she could have acted differently. Thus, on the possibility of acting otherwise is a necessary condition for an action’s having been free.

⁹⁸cf. Van Inwagen

⁹⁹ Ibid. p. 33

Harry Frankfurt challenges the principle of alternative possibilities writing, “A person may well be morally responsible for what he has done even though he could not have done otherwise.”¹⁰⁰ Consider Ellen Smith who intentionally crashes her car in an effort to defraud her insurance company. While she chooses to crash her car at a particular time, the conditions on the road that evening combined with the condition of her car make it such that she would have crashed at that time with or without having made such a decision. In this case there is no malevolent agent conspiring to coerce any behavior or action. Smith is responsible for crashing her car in an effort to defraud her insurance company, and this despite the fact that she could not have done otherwise than crash her car. Smith is responsible because she chose to crash her car notwithstanding the slickness of the road or her poor brakes. If this seems too fanciful still, consider that the best time to fake an accident would be precisely when one has bad brakes and the road is wet—these conditions make Smith’s fraud more believable and they force her crash even if she were to change her mind.¹⁰¹

What stories of this sort, so called Frankfurt-style examples, are meant to show is that the principle of alternative possibilities is irrelevant to moral responsibility. This thesis is sometimes called the irrelevance thesis and is a fundamental problem for incompatibilists because it presents a challenge to the

¹⁰⁰ Frankfurt, Harry (1969) “Alternate Possibilities and Moral Responsibility,” reprinted in *The Importance of What We Care About* (New York; Cambridge University Press) p. 1

¹⁰¹ Some people intentionally construct such situations explicitly to combat the possibility of their failure of nerve. Consider the addict who chooses to quit using drugs and implements a failsafe mechanism by burning bridges with her suppliers. If she quits successfully without ever trying to score more drugs we should say that she is directly responsible for quitting. On the other hand, if she succumbs to temptation and tries to score more drugs she will be denied and will avoid using despite her own immediate choice. Of course, in the failsafe scenario the addict may be responsible for arranging the failsafe and thus derivatively responsible for quitting.

linking of alternative possibilities with moral responsibility. If one can be morally responsible without having alternative possibilities, the consequences argument about determinism is irrelevant to the issue of moral responsibility. Since most of the historical defense of libertarian free will is based on the consequences argument, Frankfurt style cases represent a serious challenge to one argument for both libertarian free will and determinism.

Daniel Dennett endorses the idea that Frankfurt style cases can be defended against objection(s), but suggests that they do not go far enough in providing a convincing refutation of the principle of alternative possibilities. On Dennett's view "whatever 'could have done otherwise' actually means, it is not what we are interested in when we care about whether some act was freely and responsibly performed."¹⁰² His suggestion is that the principle of alternative possibilities has been mistakenly presumed to be a metaphysical stipulation on freedom and thus by extension a requirement for moral responsibility. Viewing the principle of alternative possibilities as a metaphysical condition is mistaken because it would be both unanswerable and would violate the sense that we are responsible for our characters. Thus, Dennett argues, when we inquire about alternative possibilities we really mean something more like, "how did I make my decision and was that process adequate? Can I modify my process to make it better?"

On Dennett's view the trouble with the question "could I have done otherwise?" stems from the fact that many of our actions stem from firm convictions, beliefs, and characters. Dennett uses himself as an example, suggesting that he

¹⁰² Ibid. pp. 131-132

could never be persuaded to torture another human being for a thousand dollars. It is, he claims, impossible that he do otherwise than reject the torture for money offer. This lack of alternatives is not because he is out of control, coerced or otherwise unfree, rather it is “because I see so clearly what the situation is and because my rational control faculty is *not* impaired.”¹⁰³ Dennett maintains that he would be responsible and perhaps praiseworthy for his action (refusing to torture) because “a considerable part of being a responsible person consists in making oneself unable to do the things one would be blamed for if one did them.”¹⁰⁴ Of course, having the responsibility of arranging ourselves such that we could not do otherwise in some certain situations implies that the principle of alternative possibilities is false.

The question “could I have done otherwise” opens the possibility of reflecting on what went wrong with an earlier action and this reflection is vital to self improvement. On Dennett’s view it simply doesn’t matter whether we were metaphysically free to otherwise or not. “Who cares whether...I could have done something else? I didn’t...and it’s too late to undo what I did.”¹⁰⁵ Contra-causal freedom is on Dennett’s view simply not a kind of freedom worth wanting because in it cannot ever matter to one’s actual actions. The sorts of freedom worth wanting are freedom from coercion, freedom to participate in the political process, not the freedom to do what one doesn’t actually do.

Kane argues that what is important in the question “could I have done otherwise?” is not only the possibility of taking corrective action, but also attributing

¹⁰³ Ibid. p. 133

¹⁰⁴ Ibid. p. 135

¹⁰⁵ Ibid. p. 142

responsibility for our past. This additional consideration is key to the libertarian (incompatibilist) response is the condition on freedom Kane describes as “ultimate responsibility.” Ultimate responsibility requires that an agent be personally responsible for causing an action. Ultimacy is a “backtracking” or historical condition about the origin of an agent’s action. Ultimate responsibility requires that an agent’s action ultimately or originally stems from a voluntary contribution or willing action. As Kane puts it, “ultimate responsibility lies where the ultimate cause is.”¹⁰⁶ This means that an agent is free only to the extent that her actions are ultimately caused by her own choices about actions, character, etc.

The condition of ultimate responsibility suggests the importance of looking backward in an effort to determine responsibility. Kane argues that some agents without genuine alternatives at time Θ could be free at Θ because they are ultimately responsible for their character at Θ due to voluntary actions at Θ_{-1} . Such an agent would be free at Θ because she had alternative possibilities at Θ_{-1} and her choice at Θ_{-1} is ultimately responsible for her action at Θ . In this way Kane thinks, “we salvaged the AP condition in response to Dennett by invoking the UR condition” and suggests “incompatibilists about moral responsibility should do the same in response to Frankfurt.”¹⁰⁷

Recall the Frankfurt-style case of Smith whose actions crashed her car under conditions which would have led her car to crash anyway. Kane argues that Smith is responsible because she puts herself in this situation “by virtue of other actions or

¹⁰⁶ Ibid.

¹⁰⁷ Ibid. p. 42

choices in the past with respect to which (s)he could have done otherwise."¹⁰⁸ This strategy requires that some actions may be both free and violate the condition of alternative possibilities, because the situation itself arises based on a having a character that the agent could have changed. Thus, in situations without genuine alternative possibilities the agent may be thought to be free in the incompatibilist sense when she bears ultimate responsibility in virtue of having had alternative possibilities in relevant past situations.

If, as Kane believes, the libertarian account of incompatibilism hangs on the notion of ultimate responsibility challenges to its possibility are central to a preference between incompatibilism and compatibilism. Here I offer two main objections to the notion of ultimate responsibility as Kane articulates it. First, the ultimacy condition is an elusive condition to pin down, a sort of unmoved mover, and the arche of an agent's action. If an agent's freedom is dependent upon earlier free choices one will wonder how responsible an agent could be for the development of their character at ever earlier stages. At the limit this line of thinking may produce a vicious regress of the kind supposed in the cosmological argument for God. Second, even supposing the ultimacy condition is both possible and identifiable, there is some serious concern about the base responsibility condition in a number of Frankfurt style cases. If accurate, either one of these objections is sufficient to derail the libertarian account of freewill articulated here.

Nevertheless, libertarians believe that many cases of action include both an agent's ultimate responsibility and genuine alternative possibilities. There are

¹⁰⁸ Ibid.

plausible reasons to believe that an agent sometimes is both ultimately responsible for her actions and could have done otherwise. There are also plausible arguments against such possibilities. As I said earlier my aim here is not to offer a thorough accounting of those arguments, nor certainly to settle the debate. Rather my aim is to sketch the main positions and controversies in the free will debate in an effort to see how they relate to discussion of weakness of will. Let us suppose for the sake of argument that the Libertarians are successful in defending the adequacy of both proposed conditions, the principle of alternative possibilities and ultimate responsibility. Further, let us suppose that Libertarians show us that there are cases of action which meet both conditions. In such a world we should all be prepared to admit that freedom exists and is (in an important sense) incompatible with determinism. But what would the truth of libertarian incompatibilism mean for weakness of will?

Accounts of weakness of will typically rely on a division in the desires, judgments, intentions, beliefs, etc. of an agent, with the effect of leading the agent to act in a way that she experiences as failing herself. The weak willed agent believes that she should do one thing, but does another. Moreover, the weak willed agent's failure is the result of an internal division, not external coercion or internal compulsion. Such an action can only be free on the libertarian account only if the agent is both free to have done otherwise, and ultimately responsible for the weak willed action. To see how this is possible first note that the weak willed agent's failure comes from a tension in her own judgments, intentions, beliefs, etc, what

Kane calls “internal motivational points of view.”¹⁰⁹ This suggests that an agent’s failure to act as she thinks best is an action for which the agent is ultimately responsible, because it is her inclinations that ultimately support both the preferred and inferior course of action.

Further, the weak willed agent’s internal tension between options leads to uncertainty about which course she will pursue. Kane writes, “uncertainty in the minds of agents...is reflected in the indeterminacy of their efforts to overcome temptation and in the consequent indeterminacy and unpredictability of the outcomes.”¹¹⁰ The conflicted agent’s actions are indeterminate because of the tension in her evaluative process. This tension and the resulting indeterminacy entails that she could act on either her better or inferior inclinations. It is the tension in the agent’s evaluation that allows her a plurality of options for action—evaluative tension allows that she could act otherwise. Thus, the tension experienced by the weak willed agent leads to the satisfaction of the condition of alternative possibilities.

Where the decision between possibilities is made based on the agent’s decision of “strong willed” options and “weak willed” options, it is her decision which “makes the difference” in her action. Of course, not all cases of action involving a conflicted agent are resolved on the basis of the agent’s decision, sometimes coercive influences, or immutable psychological forces “make the difference.” But, there is no reason to suppose that all failures or weak willed actions are of this latter type, any more than we should suppose that all actions are of the latter type. One’s better judgment has a clear motivational force, but better judgment is not the only

¹⁰⁹ Ibid. p. 132

¹¹⁰ Ibid. p. 132-133

thing with motivational force; lesser judgments are motivational as well. The point here is that what an agent voluntarily causes may stem from both better or inferior motivations and the agent still be thought ultimately responsible. To sum up, the libertarian position is that an agent's inner conflict can lead to indeterminacy of action, the possibility of doing otherwise, and that the final action either way may be based on the agent's decision, making her ultimately responsible. Thus, if correct, on the libertarian account "free will requires the recognition that failures or weakness ...can occur and need not always be compulsive or determined."¹¹¹

Compatibilism: Determinism and Freedom

Donald Davidson once wrote, "Hobbes, Locke, Hume, Moore, Schlick, Ayer, Stevenson, and a host of others have done what can be done, or ought ever to have been needed, to remove the confusions that can make determinism seem to frustrate freedom."¹¹² While many of us are willing to accept that the argument for determinism is inductively strong, this usually leaves us unfazed in our support for concepts like freedom and moral responsibility. Compatibilism is at root the view that freedom and moral responsibility are compatible with the truth of casual determinism. The view shared by all compatibilists is that if determinism turns out to be true, this will not negate all genuine accounts of free agency nor moral responsibility. The allure of compatibilism is the promise of showing how our intuitive sense of freedom can reasonably coexist with the possibility of deterministic forces. The main difficulty of advancing compatibilism is that determinism seems to rule out an agent's having genuine alternative possibilities. Thus, the compatibilists'

¹¹¹ Ibid. p. 155

¹¹² Watson p. 173 citing Davidson, Donald (1980) p. 63

strategy is to show that the truth of determinism does not entail the absence of alternative possibilities.¹¹³

Hume's early compatibilism continues through the work of G.E. Moore and A.J. Ayer, each of whom argued that the existence of alternative possibilities is compatible with determinism.¹¹⁴ The style of argument employed by Ayer and Moore focused on a particular understanding of the concept "could have done otherwise." If one sees the counterfactual *could have* acted otherwise as indicating something about the actual world, both argued, then one has gone astray. In fact the counterfactual condition, "could have acted otherwise," they thought, should be understood as referring to another possible world. On this view saying that someone could have acted other than she did amounts to saying "had she wanted to act differently, then she could have acted other than she did." The move here is to focus on the conditional aspect of a counterfactual.

The key to conditional analysis is the idea that a determined agent has alternative possibilities when she could have done otherwise if she had wanted to. On this account, the free agent is not actually free to do otherwise at the time of her action, rather she might have done otherwise provided that things earlier in her causal chain had gone differently. The trouble with this defense of compatibilism is that it seems to play kick the can with the burden of alternative possibility. The worry is that an agent's disposition is linked causally to some earlier and also determined

¹¹³ Some call this "strong compatibilism" because if successful the strategy produces a broad view of freedom. Campbell, Joseph Keim (1997) "A Compatibilist Theory of Alternative Possibilities," *Philosophical Studies*, 88:3 pp. 319-330

¹¹⁴ See David Hume (1748) *Enquiry Concerning Human Understanding* reprinted (1995) Prentice Hall, p. 104, G.E. Moore (1903) *Principia Ethica* Ch. III, Online at: <http://fair-use.org/g-e-moore/principia-ethica/chapter-iii> and A.J. Ayer (1954) "Freedom and Necessity" Online at: http://www.unc.edu/~dfrost/classes/Ayer_FreedomandNecessity.pdf Accessed: June 20, 2011.

event. Thus, in order to alter one's dispositions it seems one must be able to alter earlier events in the causal chain. And to alter those events, one must be able to alter still earlier events, and so on.

Donald Davidson argued that dispositions are not governed strictly by earlier events, but are states determined in part by natural laws, so no earlier event is required for the existence of alternative dispositions.¹¹⁵ The idea is to avoid a regress by demonstrating the existence of two distinct types of possible action, determinate causal action, and intentional agent causal action. For example, imagine Moshe who naturally prefers bagels to bacon. If Moshe chooses either breakfast food he will always choose bagels if he intentionally chooses either. However, given an unusual turn of events Moshe might choose a breakfast with bacon over one with bagels, just not intentionally. Davidson thinks that given Moshe's natural disposition it would be reasonable to say that he could have done otherwise--that is, he could have *intentionally* chosen the bacon breakfast. Moshe is free, just not very free. He is determined to eat the bacon, but he has alternative possibilities because if he had a different disposition he could do so intentionally.

In his 1956 essay "If and Cans" J.L. Austin describes kicking himself for missing a very short putt and thinking that he could have holed it in conditions precisely as they were. Of course, he tried to hole the putt, but he missed. Here Austin is using 'could have' to mean that he both has the capacity to intend to hole the putt and ordinarily has the capacity to physically hole the putt. As Austin puts it

¹¹⁵ Donald Davidson (1973) "Freedom to Act" reprinted in Davidson (1980) pp. 63-81

'cans' are constitutionally 'iffy.' the point here is that there are two conditionals at work in a counterfactual statement about what one could have done. First, there is an implicit "if I had tried, intended, etc." and second, "if I ordinarily had the capacity." Davidson's account focuses on the iffyness of intention formation and works through the possibility of producing intentions to act in ways that the physical circumstance apparently precludes. This reliance on iffyness leads Austin to reject compatibilism, but others offer an alternative conception focusing on dispositions.

A dispositional account of freedom supports compatibilism by focusing on the second leg of Austin's worrying putter example. Rather than highlight the golfer's intent, dispositionalists focus on the fact that the golfer *ordinarily* holes the putt. The general strategy is to focus on what might have occurred in other nearby possible worlds and to describe possibility in terms of what David Lewis calls "maximally similar" worlds.¹¹⁶ This shift gives the dispositional analysis an edge in explaining difficult cases¹¹⁷ and expands the notion of "alternative possibility" to include alternative actions, not merely alternative intentions about the same action. For example, if Moshe would ordinarily notice the bacon in his quiche, it is possible that he could have acted otherwise and eaten the bagel instead.

In advancing and defending compatibilism on a dispositional account Michael Smith offers the example of failing to say something witty at an appropriate moment in response to an objectionable argument. According to the consequences argument, one should think that if determinism is true, it was literally impossible for

¹¹⁶ David Lewis (1986) "Counterfactual Dependence and Time's Arrow" reprinted in *Philosophical Papers* OUP pp.32-52.

¹¹⁷ I have in mind here McKenna's case of Danielle who must choose between a black and blonde lab.

you to have offered a witty response to the argument. But, Smith maintains, in such cases you are able to offer a witty response just in case “the possible worlds in which you think of the better response are nearby, or very similar to, the actual world in which you don’t.”¹¹⁸ The idea is that if in other possible worlds with highly similar laws and histories, you would have done otherwise, then it is possible that in this world, with these laws you might have done otherwise as well. Smith clarifies, “the crucial point is that we do not need to imagine a massive transformation of your nature in order to imagine you thinking of a better response. We need simply to imagine you, pretty much as you are, but giving a better response.”¹¹⁹

Extending Smith's dispositionalism Jeanette Kennett writes “the reasons so far given for thinking that there is no such thing as weakness of will are crucially flawed.”¹²⁰ Kennett and Smith offer an account of weakness of will focusing on dispositions as a ground for self-control. The idea is that an agent can exhibit freewill by controlling her dispositions about her actions, even when her actions are out of line with those dispositions. In such cases our actions may be pre-empted by deterministic forces, but we control our dispositions with respect to the action. This requires a distinction between our normative judgments, themselves evaluations of our dispositions or “idealized desires”, and our motivations.¹²¹

Consider an example Kennett and Smith adapt from Gary Watson’s *Free Agency*. The mother of a screaming infant is at the end of her rope and finds herself

¹¹⁸ Michael Smith (2004) *Ethics and the A Priori* (Cambridge University Press; New York, NY) p. 91

¹¹⁹ Ibid.

¹²⁰ Jeanette Kennett and Michael Smith (1997) “Philosophy and Commonsense: The case of weakness of will” reprinted in *Ethics and the A Priori* (CUP) p. 59

¹²¹ Ibid. p. 63

wanting to leave her screaming child to drown in the bathwater. Should she do so, we would rightly say that the mother intentionally drowns her child—she acts on motivational reasons she holds, frustrations, exhaustion and the like. However, such an explanation is not a justification and it would not be right to say that she acts on a normative reason. Rather, the mother who out of frustration drowns her infant acts contrary to her normative reasons for action—she does not value her child’s being drowned, she is simply frustrated with his crying.¹²² The point to which Smith and Kennett use this account is to emphasize that our ordinary conception of reasons for action contains an ambiguity between motivating reasons and normative reasons. The exhausted and frustrated mother they suggest has motivational reasons for drowning her child—frustration and exhaustion—and normative reasons to dry and clothe her child and leave it in a safe place while she calms down—the value she places on her child’s life. A distinction between motivational reasons and normative reasons is useful in that it helps to make sense of internal conflict in a commonsense way.

Watson himself extends the example of the deranged mother in a similar way describing the mother as like others in ordinary situations who are “estranged from a rather persistent and pervasive desire.”¹²³ On Watson’s view such desires generate reasons for acting toward “getting rid of the desire” and “hence this kind of reason differs importantly from the reasons based upon the evaluation of the activities in question.”¹²⁴ Here Watson agrees that the source of an agent’s desires is important

¹²² For more on this point see Watson 2004 p. 22

¹²³ Gary Watson (1982) “Free Agency” reprinted in his *Agency and Answerability* (OUP; 2004) p. 19

¹²⁴ *Ibid.* p. 20

in determining the content of the reasons such desires generate. But, Watson, of course, does not think that the possibility of conflict between what we value and what we desire leads to the possibility of strict akratic action. On Watson's view the irrationality (or nonrationality) of acting on what we desire at the expense of what we value represents an impediment to our free agency. On Watson's view those passions, appetites, or acculturated desires which override, blind, or subvert our judgment are compulsive and thus undermine free agency. Using the language of Kennett and Smith, Watson finds motivational reasons compulsive and normative reasons the ground of free agency. As Watson himself puts the point, what makes us responsible agents is normative competence, a minimum responsiveness to reasons. This leads Watson to the view that "God is the only free agent sans phrase." By this Watson means, since God is the only perfectly knowledgeable and perfectly powerful agent, only God has perfect normative competence, and thus only God is perfectly free. People, Watson claims, "are only more or less free agents, typically less."¹²⁵

Kennett and Smith disagree with Watson because they believe that it is misguided to suppose that *acting* in accord with one's normative reasons is the only way to have free agency or self-control. According to Kennett and Smith, "Some exercises of self-control must themselves be, not actions, but rather manifestations of our cognitive dispositions."¹²⁶ Here Kennett and Smith attempt to avoid a regress of putatively self-controlled actions by suggesting that self-control can be manifest in something other than an action. The sort of manifestations that Kennett and Smith

¹²⁵ Ibid. p. 31

¹²⁶ Smith (2004) p. 66

have in mind includes things like thoughts about what one ought to do. The idea is that an agent whose motivational reasons conflict with her normative reasons may express her normative reasons, and thus her freedom, not only through action, but through certain cognitive dispositions. On this view it is possible to be free in virtue of having certain cognitive dispositions, while acting in a way that is contrary to one's own values. That is one's cognitive dispositions might indicate one's freedom, while her actions go against her better judgment.

Imagine, for example, that both Peter and Paul judge it best to give to the hungry, but being less than fully rational desire to buy lattes more than to give to the hungry. Both Peter and Paul have a conflict in reasoning between normative and motivational reasons for acting. Peter and Paul differ however in that Peter's cognitive dispositions are such that he ordinarily thinks through his commitments and these thoughts ordinarily lead him to act on his better judgment, while Paul's cognitive dispositions are such that he rarely thinks about his commitments and so often acts contrary to his better judgment. However, on this occasion both Peter and Paul fail to act as they think best and in the end both buy lattes rather than slip their cash into an envelope for UNICEF. On Kennett and Smith's account we ought to consider Peter free to have done otherwise and Paul unfree. In this case Peter is merely weak-willed, but Paul was literally compelled by his desire. The difference between the two is that Peter has cognitive dispositions which can and ordinarily do lead him to resist his desires and these dispositions make Peter free. Meanwhile Paul has no such dispositions and this lack leaves him unable to resist his desires even when he judges that he should.

Importantly, having dispositions which can and ordinarily do lead one to act on one's better judgment does not entail that one will always act on normative reasons. What such dispositions entail is that one has developed the mechanism required to act on normative judgment in those hard cases where normative reason conflicts with motivational reason. But having the requisite mechanism only entails a capacity to act on normative reasons in hard cases and this capacity is the freedom to act on normative reasons, not a requirement to act on those reasons. It is perfectly possible, on this view, for one to both have the capacity for bring oneself to act on normative reasons and fail to exercise this capacity in a particular instance. It is for this reason that Kennett and Smith argue one can be both free to do otherwise and fail to exercise self-control, and thus they claim the existence of weakness of will is theoretically plausible.

Other compatibilists worry that the notion of "nearby" possible worlds will not withstand much scrutiny and shift the dispositional account to a broader one. Kadri Vihvelin offers the following, "a fragile glass is a glass that is liable to break; that is, it is a glass that *can* break, even if it never does."¹²⁷ Vihvelin's point is that dispositional compatibilists can rely on unmanifested dispositions (tendencies, casual powers, capacities) to act in certain ways to support genuine alternative possibilities. The force of this claim lays in two main points, first, the view that counterfactual dispositions are compatible with the truth of determinism and, second, the view that counterfactual dispositions imply a counterfactual ability. Thus, a

¹²⁷ Kadri Vihvelin (2004) "Freewill Demystified: A Dispositional Account," *Philosophical Topics*, 32: 427-450.

person's unexpressed attributes can make her capable of acting contrary to a casually determined outcome.

Vihvelin goes on to affirm the possibility of weakness of will in the following example:

“Person X lies to Customs officials. When questioned later, X agrees that breaking the law is wrong but candidly explains that he did it because he did not want to spend hundreds of dollars on import duties. X acted contrary to his better judgment about what he ought to do; so he acted against what he most valued. But there is little reason to suppose that he acted unfreely.”¹²⁸

On the compatibilist view acting freely in cases like this simply means that the agent could have done otherwise given her dispositions. Of course, the facts of the situation might determine some other outcome, but the agent herself has features which would allow alternative action. The point is not that the agent could *actually* do otherwise, only that her dispositions allow that she could do otherwise.

The dispositionalist account might be parodied "If it weren't for all these pesky facts I could do X. After all, there is nothing about me that prevents X, it's just the world that won't let me." Fischer and Ravizza report that a person in a locked room might have the disposition to open the door, after all he is capable of opening doors and has often done so. Nevertheless they say, he is not free to open the door, it is locked. However, the point for my purposes here this is that among those who defend compatibilism there is a significant sense that weakness of will is a distinct possibility. Thus, if the dispositionalist account of freedom is eventually vindicated, there is good reason to believe that the truth of weakness of will would follow

¹²⁸ (1994) "Are Drug Addicts Unfree?" In *Drugs, Morality and the Law* Ed. S. Luper-Foy and C. Brown (NY; Garland Press) pp. 51-78.

closely. In what follows I consider another compatibilist view about freedom and responsibility and discuss its implications for the possibility of weakness of will.

Semi-Compatibilism

Semi-compatibilism is the final view of freedom I will consider here. As a working short hand it is helpful to think about semi-compatibilism as maintaining that “causal determinism is compatible with moral responsibility, quite apart from considerations pertaining to “could have done otherwise.”¹²⁹ Semi-compatibilists do not argue that determinism is compatible with both freedom and moral responsibility. They acknowledge that the truth of determinism might eliminate the possibility of freedom understood as having genuine alternatives. However, semi-compatibilists argue that moral responsibility is separable from freedom and defend the compatibility of determinism and moral responsibility.

The separability thesis of semi-compatibilism is based on P.F. Strawson’s view that the final answer to the question of whether humans are free is far less significant than the fact that we believe ourselves to be free. As one writer puts it, “Strawson made a contribution to the free will versus determinism discussions by pointing out that whatever the deep metaphysical truth on these issues, people would not give up talking about and feeling moral responsibility, praise and blame, guilt and pride, crime and punishment, gratitude, resentment, and forgiveness.”¹³⁰ Feelings such as praise, blame, guilt, resentment, gratitude, indignation, love,

¹²⁹ John Fischer *Garden of Forking Paths* Online at: http://gfp.typepad.com/the_garden_of_forking_pat/2004/06/do_you_know_wha.html Accessed: August 25, 2011.

¹³⁰ Peter F. Strawson, *Information Philosopher* Accessed: August 26, 2011, Online at: <http://www.informationphilosopher.com/solutions/philosophers/strawson/>

seemed to Strawson a basic and inalienable part of the human description of the world and thus he thought that these “reactive attitudes” were “facts” of human life independent of the truth of determinism.¹³¹

Strawson saw reactive attitudes like moral praise, blame, indignation, gratitude, etc. as appropriate only in interaction with other human beings. This limit on the appropriateness of reactive attitudes is something Strawson saw as divulged from the use of natural language. For example, though one might jokingly blame a bird for defecating on one’s car, it would be inappropriate to be indignant or to seriously blame a bird for such an action. This is because a bird is not an appropriate candidate for such reactive attitudes. A morally responsible agent on the Strawsonian view is an agent who is an appropriate recipient of reactive attitudes such as praise or blame. It does not follow that reactive attitudes actually must be attributed to morally responsible agents, other considerations may limit actual attribution. The point is simply that morally responsible agents are apt candidates for reactive attitudes. As he puts it, “It matters to us whether the actions of other people...reflect attitudes toward us of good will, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other.”¹³² It is taking or holding such attitudes toward one another that ultimately constitutes our having responsibility for our actions. Watson describes Strawson’s view, “as natural and primitive in human life as friendship and animosity, sympathy and antipathy. It rests

¹³¹ P.F. Strawson (1962) “Freedom and Resentment” *Proceedings of the British Academy* 48, reprinted in Gary Watson (1982) *Free Will* (OUP) pp. 59-80.

¹³² *Ibid.* p. 67

on needs and concerns that are not so much to be justified as acknowledged.”¹³³

This is to acknowledge that a whole range of reactive attitudes constitute responsibility. As Fischer and Ravizza put the point “The Strawsonian approach holds that being morally responsible *just is* being an apt recipient of these attitudes and a participant in the associated practices.”¹³⁴

The “question unanswered” Strawson writes, “is about what it would be rational to do if determinism were true, a question about the rational justification of ordinary inter-personal attitudes in general.”¹³⁵ Most people will agree that we are not responsible for all of our actions. Attributing responsibility to every action is not enriching, it is muddying. One wants to know when and how reactive attitudes are justified. Fischer and Ravizza offer two general conditions for the appropriate attribution of reactive attitudes and thus for responsibility. The first condition is epistemic, “It captures the intuition that an agent is responsible only if he both knows the particular facts surrounding his action, and acts with the proper sort of beliefs and intentions.”¹³⁶ An agent is not an appropriate subject of reactive attitudes when she is ignorant of important facts about her action. That ignorance can excuse moral responsibility is they think indicative of the importance of the epistemic condition for responsibility. Put another way this epistemic condition holds that “a person must know (or be reasonably expected to know) what he is doing—he must

¹³³ Gary Watson (1987) “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme” reprinted in Watson (2004) *Agency and Answerability* (OUP) p. 222-223

¹³⁴ Fischer and Ravizza (1998) *Responsibility and Control* (OUP) fn. 12 p. 10

¹³⁵ Cf. Strawson (1962) p.78

¹³⁶ Cf. Fischer and Ravizza (1998) p. 13

not be deceived or ignorant about the particular circumstances and manner in which he is doing it.”¹³⁷

The second condition Fischer and Ravizza offer for the appropriate attribution of moral responsibility they call the “freedom-relevant” or “control” condition.¹³⁸ The key point regarding control is to identify the sort of control that is worth wanting. Fischer and Ravizza argue that the control worth wanting is the sort that allows for moral responsibility. They describe libertarian and dispositionalist accounts as concerned with alternative possibilities because they assume moral responsibility requires “regulative control.” To the contrary they suggest moral responsibility only requires “guidance control.”¹³⁹

To see this distinction consider Sally, a driver operating a normally functioning vehicle, in normal conditions, without any stipulation of determinism and with typical mechanisms of intentions, desire, etc in place. Sally decides to make a right turn, intends to do so and on the basis of her decision, intention, etc. does in fact turn the car to the right. Semi-compatibilists think that Sally is guiding or controlling the car in the relevant respect. Thus, they say Sally has “guidance control.” And, on their view guidance control is enough to suppose that Sally is morally responsible for turning the car. To have “regulative control” Sally must not only have the power to turn in the direction she actually turns, she must also have the power to engage in a

¹³⁷ Cf. Fisher and Ravizza (1998) p. 13

¹³⁸ Daniel Dennett sums this point up nicely as follows:

“What we fear—or at any rate a very important part of what we fear—in determinism is the prospect that determinism would rule out control, and we very definitely do not want to lose control or be out of control or be controlled by something else—like a marionette or puppet. We want to be *in control*, and to control both ourselves and our destinies. But what is control? Curiously, this obviously important question has been scarcely addressed by philosophers.” Daniel Dennett (1984) *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge, Mass.: MIT Press) p. 51

¹³⁹ Cf. Fisher and Ravizza (1998) p. 31

counterfactual action (including turning left, stopping, etc.) The point is that regulative control requires that Sally have both control over her action and control over an alternative possibility.¹⁴⁰

Fischer and Ravizza defend guidance control as adequate for moral responsibility using a Frankfurt-style example. Again consider Sally, but this time suppose that Sally is behind the wheel of a “driver instruction automobile with dual controls...the instructor is quite happy to allow Sally to steer the car right, but if Sally had shown any inclination to cause the car to go in some other direction, the instructor would have intervened to cause the car to go right (just as it actually goes.)” In such a case Sally has guidance control, but lacks regulative control. The instructor has regulative control, but not guidance control. The point of this example is twofold, first, it makes clear that real world cases exist which support the distinction between guidance control and regulative control. Second, it makes a case for the plausibility of supposing that guidance control is the sort of control that ultimately warrants moral responsibility. If Sally decides to turn right, intends to turn right and in fact does turn right, and in so doing passes her driving test, we suppose that she has earned or deserves her driving license by proving herself responsible driver. But, if she is only responsible for her driving when the instructor is not present, there is no reason to suppose that she deserves responsibility for driving well when the instructor is present. The ground for responsibility in such a case seems to rely explicitly on guidance control and explicitly not on regulative control.

¹⁴⁰ Ibid.

If we think Sally had guidance control in her instructional vehicle and is thus morally responsible, what was it about this action that evidenced her control? Fischer and Ravizza write, "an agent has guidance control when the mechanism that actually issues in the action is his own, reasons-responsive mechanism."¹⁴¹ So, guidance control includes two conditions, first "the mechanism must be *'the agent's own'*."¹⁴² Roughly this means that an agent "takes responsibility" for the spring of her action.¹⁴³ The second condition is that the agent's action was guided in a way that is responsive to reasons for and against the action. Roughly, reasons-responsiveness entails that an "agent" for whom there could be no reason sufficient for her to do otherwise is not morally responsible, while an agent who is highly sensitive to reasons pointing toward contrary action is responsible.

The idea that moral responsibility requires "taking" responsibility is fairly intuitive, but it is far from clear what it means. Fischer and Ravizza offer a three part account of taking responsibility borrowed in significant respect from Thomas Nagel.¹⁴⁴ These elements represent a set of ingredients which taken together account for what it means for an agent to take responsibility for and thereby come to "own" the mechanisms that lead to action. The first ingredient of taking responsibility is that the agent "sees himself as the source of his behavior."¹⁴⁵ This requirement amounts to agent recognition of the fact that his choices and actions effect the real world. The second element of taking responsibility is that "the individual must accept

¹⁴¹ Ibid. p. 39

¹⁴² Ibid. p. 40

¹⁴³ Ibid. Ch. 8 pp. 207-239

¹⁴⁴ Thomas Nagel (1979) "Moral Luck"

¹⁴⁵ Cf. Fisher and Ravizza p. 210

that he is a fair target of the reactive attitudes as a result of how he exercises this agency in certain contexts.”¹⁴⁶ Of course, the agent need not have a full blown theory of moral responsibility, it will be enough that the individual sees praise or blame as appropriate moves in a “social game.” The third element of ownership or taking responsibility is an epistemic requirement related to the first and second conditions. As Fisher and Ravizza put the point, “taking responsibility requires that the individual’s view of himself specified in the first two conditions be based, in an appropriate way, on the evidence.”¹⁴⁷ Although this requirement is left quite vague, the idea seems to be that an agent must have reasons for believing both that her choices effect the world and that she is an appropriate target of praise and blame. On their view taking responsibility is akin to asking to participate in an ongoing moral conversation. Such a request is granted only if it is believed that an individual is a suitable participant in the conversation. One is thought a suitable participant if they ask to engage in the moral discussion at the appropriate time—that is, in response to a morally significant action. This responsiveness to a morally significant moment is in their terminology responsiveness to “evidence.”

Still, it isn’t enough that an agent recognizes the importance of her choices/actions and accepts the appropriateness of responsibility an individual’s ability to respond to reasons contrary to her action is also an important element of moral responsibility. Suppose, an agent has taken responsibility for her actions in the sense that her acts spring from her own mechanism, and, second, suppose that she has sufficient reasons to act in a way other than she has chosen to act. In such

¹⁴⁶ Ibid. p. 211

¹⁴⁷ Ibid. p. 213

circumstances an agent is strongly reasons responsive if she does three things. First, “the agent would *recognize* the sufficient reason to do otherwise.” Second, the agent “would thus *choose* to do otherwise.” And, third, the strongly reasons responsive agent would “*act* in accordance with the choice.”¹⁴⁸ Strong reasons responsiveness is perhaps a laudable goal, but on Fisher and Ravizza’s view it is a more stringent standard than is required for moral responsibility.

In order to explain why they reject strong reasons responsiveness Fischer and Ravizza offer the example of Jennifer, who with an important paper due and a looming deadline decides she has sufficient reason not to go to tonight’s basketball game. But, Jennifer goes to the basketball game anyway, despite realizing that she had sufficient reason not to go. As they put it, “The failure of strong reasons-responsiveness here stems from Jennifer’s disposition toward weakness of will.”¹⁴⁹ On their view it is reasonable to hold Jennifer morally responsible. This leads them to conclude that strong reasons responsiveness is too demanding for moral responsibility. Jennifer recognizes the sufficiency of her reasons to stay home, but chooses to and does otherwise. Two of the three conditions for strong reasons responsiveness are missing and yet Jennifer appears to be morally responsible for going to the game instead of completing her assignment. It should be noted at this point that the semi-compatibilist account of freedom not only accepts the possibility of weakness of will, but makes use of the concept in the process of honing the requirements for moral responsibility. Fischer and Ravizza reject strong reasons-

¹⁴⁸ Ibid. p. 41

¹⁴⁹ Ibid. p. 42

responsiveness *because* they think that an account of moral responsibility must include cases of weak-will.

Weak reasons-responsiveness allows that Jennifer is morally responsible if there is some possible world, no matter how remote, in which she would have responded to her sufficient reasons to stay home. For instance, if the tickets to the basketball game had been a thousand dollars Jennifer would have stayed home, thus acting in the way she had most reason to act. This coming together of reasons and actions in a possible world is enough to warrant reactive attitudes on the weak interpretation of reasons responsiveness. Weak reasons responsiveness is a plausible requirement for moral responsibility and conforms to many people's intuitions. Consider the moral culpability of addicts for using illegal drugs. People often describe the act as morally responsible using the following logic, "If there were a police officer standing next to the addict, she would not use the illegal drug. Therefore, the addict is in control of her behavior." This is, of course, a condensed version of the argument, but logic being used requires that control amounts to weak reasons responsiveness. The argument is not that in the actual case a police officer is present, rather it depends on what would happen in another possible world where the officer is present. The argument then extends the connection between action and reasons in the possible world to represent control in the actual world.

The trouble with weak reasons responsiveness is just that it appears to allow too many cases of apparently unresponsive or unreasonable action count as sufficiently responsive to reasons. As an example Fisher and Ravizza offer the story of a man who boards a ferry with a concealed saber and who intends to wait until the

ferry is underway and then kill all those on board. They further suppose that the man is almost entirely unresponsive to reasons and no matter what reasons are put before him he will proceed to slay all the passengers. The one exception to the lunatic's single minded focus on violent murder is the possibility that a passenger in the lower cabin might be smoking a Gambier pipe. There is nothing especially important about such a man smoking such a pipe, except that the murderous lunatic will recognize such a man with such a pipe as a reason to refrain from producing and using his saber. Further, if the lunatic sees the man smoking a Gambier pipe he will not only recognize this as a reason to hold off on killing, he will respond to that reason and refrain from doing any harm to any of his fellow passengers.

The preceding case is troubling to the view of moral responsibility which requires only weak reasons responsiveness because it seems that the lunatic with a fixation on Gambier pipes is weakly reasons responsive. There is at least one possible world in which he both recognizes a reason to do otherwise and acts on that reason. But the reason is by hypothesis the sort of reason that only a lunatic would find compelling. We could of course fill in the story so that the Gambier pipe smoking man was actually meaningful, but the point of the example is that we need not tell such a story for the killer to be in control of his actions on the weak interpretation of reasons responsiveness. There is nothing about weak reasons responsiveness that prevents attributing responsibility and control to people who respond to outlandish or bizarre "reasons" for acting in one way or the other.

Fischer and Ravizza settle on an account of moderate reasons-responsiveness because it preserves accountability in cases of weakness of will and

excludes accountability in cases of incoherent individuals. The key to moderate reasons responsiveness is an appropriate balance between reasons receptivity and reasons reactivity. Recall that reasons receptivity is the agent's ability to recognize what reasons exist that are relevant to the agent's actions or "the capacity to recognize that reasons exist." Reasons reactivity on the other hand is the agent's ability to choose and act on the basis of reasons or "the capacity to translate reasons into choices and then subsequent behavior."¹⁵⁰ On their view the difficulties of weak and strong interpretations of reasons responsiveness can be overcome by supposing an *asymmetrical* relationship between reasons receptivity and reactivity. They argue that moral responsibility only requires a quite weak form of reasons reactivity, but requires a much stronger receptivity to reasons.

The point of the *asymmetrical* relationship between reasons receptivity and reactivity is twofold. First, a limited or very slight reactivity to reasons is needed to preserve the possibility of weakness of will. And, second, a stronger receptivity to reasons is needed to preclude responsibility for incoherent behavior. The requirement of weak reasons reactivity is met if the agent would act differently in any single other possible world. In other words, if one would act differently in at least one scenario in which the circumstances were different, then one meets the requirement of adequately reacting to contrary reasons. To help explain their view of reasons reactivity Fischer and Ravizza offer the example of Brown "a weak-willed individual with a strong craving for the non-addictive (not literally irresistible) drug

¹⁵⁰ Ibid. p. 69

“Plezu.”¹⁵¹ Plezu is extraordinarily pleasurable and has relatively few short term side effects, but it does tend to leave one unproductive for long periods of time. Brown knows that taking regularly Plezu will damage his job performance, family life and self-respect, and considers these sufficient reasons to refrain, but he is weak-willed and regularly uses Plezu first thing in the morning. The only reason Brown would quit is if he were told that the next time he used Plezu it would kill him. Brown is morally responsible because he is both sufficiently receptive to reasons and in at least one case those reasons would be sufficient for him to translate them into action.

Gary Watson offers an account that is critical of semi-compatibilism based on the condition of weak reactivity. On Watson’s view weak reactivity is an insufficient condition for reasons responsiveness because “it is arguably consistent with motivational compulsion.”¹⁵² Consider Mele’s case of Fred, an agoraphobic whose condition is so severe that he has not left his home in more than ten years and cannot bring himself to leave even to attend his beloved daughter’s wedding in the church next door. However, in a possible alternative scenario Fred would leave his home if it were on fire, and could in such a scenario have fled to the church next door, but only because his fear of fire is more compelling than his fear of leaving the house. Semi-compatibilists seem left thinking that Fred is weakly reasons reactive and so can be held morally responsible for not attending his daughter’s wedding. However, some think, “if Fred’s fear is so debilitating that it takes something as frightening as a raging fire to him to decide to leave his house or to leave it

¹⁵¹ Ibid.

¹⁵² Ibid.

intentionally, then, it seems to me, he is *not* morally responsible for missing the wedding.”¹⁵³ The worry is that Fischer and Ravizza are able to account for weakness of will only by eliminating the capacity for compulsion.

There are however responses available to semi-compatibilists. One might suggest that Fred is not engaged in a reasons responsive activity when he flees the fire, rather he acts on brute instinct. Fred’s overwhelming fear response to both fire and the outdoors might be seen as making clear his lack of control.¹⁵⁴ Alternatively, the condition of reactivity be strengthened to only include reactivity to “weakly sufficient reasons.” The point of such a suggestion is to limit the range of cases which count as relevant to the condition of responsibility to more typical or less radical cases. The intuition at work here is that a person’s responsibility should not depend on what they might have done in a liminal case, but on what is typical for them. This strengthening stops short of reducing to a dispositionalist compatibilism, but is heading in that direction.¹⁵⁵¹⁵⁶

Recall again the case of a weak-willed agent named Brown, for whom taking Plezu is a passion, but *not* an addiction in the sense of being a compulsion. Brown will refrain from taking Plezu if and only if he will die with his next dose. Fischer and Ravizza claim that Brown is responsible for his weak-willed actions, taking the Plezu, in part because Brown uses the same mechanism in order to both refrain from taking Plezu and in order to continue using under ordinary circumstances. The difference

¹⁵³ Cf. Mele (2000) p. 451 Watson uses a similar example from Isaac Marks (1987) *Fears, Phobias, and Rituals* (OUP)

¹⁵⁴ Michael McKenna (2005) “Reasons reactivity and incompatibilist intuitions” *Philosophical Explorations*, 8 (2): 131-143.

¹⁵⁵ Robyn Waller (Under Review) “Revising Reasons Reactivity”

¹⁵⁶ John Fischer (2006) “The free will revolution (continued)” *The Journal of Ethics* 10: 315-345

they say between Brown and Fred is that Fred is using a different mechanism when he misses the wedding than he uses when he flees the fire. Thus, Brown is culpable for taking Plezu, while Fred is not culpable for missing his daughter's wedding?

There is a very real and commonsense difference between Fred and Brown. Fred is a compulsive agoraphobic; he lacks genuine control of his actions and so is not morally responsible. Brown is weak-willed; he has control over his actions, but chooses wrongly.

Conclusion

The accounts of freedom and moral responsibility offered here are among the most philosophically attractive accounts available and they each accommodate the possibility of weakness of will and moral responsibility. Robert Kane's incompatibilist libertarian view is perhaps the most commonsensical account in that it straightforwardly denies the existence of a troubling determinism. It also allows that an agent may be ultimately responsible for actions that run counter to their own better judgment and that such weak-willed actions can coexist with genuine alternatives. Although traditional compatibilists reject a dismissal of determinism they agree that moral responsibility requires that agent's have genuine alternative possibilities. On their view alternative possibilities are compatible with determinism and with weakness of will. Finally, semi-compatibilism, or narrow compatibilism rejects the importance of alternative possibilities but defends the existence of moral responsibility in a universe that may be casually determined. The semi-compatibilist is open to the possibility of weak-will on the grounds that agency

requires does not require acting on our best reasons, only that we act from a reasons responsive mechanism for which we take responsibility.

Attractive as each of these views are, they all also have difficulties yet to face, challenges left unanswered and plausible reasons to believe that they are false. I admit to being skeptical about the possibility of securing genuine ultimate responsibility in the face of the role luck surely plays in our lives. And there is something counter-intuitive about the dispositionalist's view that someone could have done otherwise because they ordinarily do otherwise, despite not having done so this time. Finally, the semi-compatibilists reliance on mechanism individualtion is somewhat murky. But despite these worries I think there is good reason to believe that we are responsible for at least some of our actions. Facing down what Ted Honderich calls the “darkness” of determinism is not a task that we can hope to complete here.

The purpose of this paper is merely to reassure the reader (and author) that we have on offer a range of plausible views about freedom and moral responsibility on which weakness of will is a genuine possibility. We need not dismiss the possibility of weakness of will due to the restrictions of defending moral responsibility more generally. In fact quite to the contrary it seems to me that there is something about the possibility of weakness of will that is important for the existence of genuine moral responsibility.

Chapter 5

Weak-Will Isn't Just Reckless

At the heart of debate about weakness of will is a discussion about practical rationality . Weakness of Will is possible only if there is a gap between reasoning about what one ought to do and what an agent actually does. A person is weak-willed if she believes that she ought to do something, but then proceeds to not do that thing in practice. So to understand if weakness of will is possible, we must understand the relationship between theorizing about action and the mechanisms that lead to action. When we are at our best theory and practice come together and we act on our beliefs. This chapter begins by explaining the distinction between reasoning about action and the mechanisms that lead agents to act and how they come together in strength of will.

The second aim of this chapter is to explain how our beliefs can be effective at guiding action at one moment and impotent at another. This section describes the conceptual space for weakness of will made through claims of mental partitioning advanced by Donald Davidson and then Al Mele. Against these claims Gary Watson argues that purported cases of weak-will devolve into either compulsion or recklessness. Having discussed compulsion in the previous chapter, here I focus on the possibility that weakness of will is reducible to recklessness. I show that weak-willed behavior is distinct from recklessness using a case study strategy illustrating situations in which an agent is both concerned to do the right thing in a way that reckless people are not, and follows something less than her best judgment.

Practical vs. Theoretical Judgment

Traditionally questions of willpower have focused on how the various faculties of will—legislative and executive—cooperate to produce strength of will and come apart to produce weakness of will. Kant described two essential roles that willing plays in guiding action—“*Wille*” and “*Willkür*.”¹⁵⁷ *Wille* describes the legislative function of the will and the role of reason in guiding practical judgment. This is roughly parallel to what I have been calling “theoretical judgment,” what Michael Smith calls “normative reasons” or what Gary Watson calls “practical deliberation.” The main concern of *Wille* is coming to understand the demands of reason, or, alternatively, coming to decide what one has reason or justification for doing or not doing; *Wille* is the activity of making up one’s mind about what course of action one should take. On the other hand, Kant’s *Willkür* is involved with executive issues of practical rationality; that is the reasons which lead to our actual action. Smith calls these sorts of reasons “motivating reasons” because they move us beyond decision and into action. The main concern here is coming to understand how practical rationality operates such that the motivations of *Willkür* stem from the evaluations of *Wille*. A strong will requires that both *Wille* and *Willkür* function appropriately and that they come to function harmoniously together.

Strength of will begins with the identification of one’s reasons for acting in various ways and sound evaluation of these conflicting reasons; which is to say strength of will requires a certain sensitivity or attention to one’s own reasons for action. In part this just means that an agent must be attentive to what Bernard Williams calls the “subjective

¹⁵⁷ Kant, *Introduction to the Metaphysics of Morals*, (6: 213-14, 226; 10-11, 26-7)

motivational set.”¹⁵⁸ The subjective motivational set includes the myriad desires and beliefs that in some way excite the passions of the individual agent. This set of impulses provides a corresponding set of reasons for action. For example, if my motivational set includes a desire for chocolate cake, I have a reason to eat chocolate cake—I desire it. An agent’s sensitivity to her own impulses is a baseline requirement for the legislative function of willing, but there is more to legislating action than recognizing one’s impulses. Evaluating the competing desires that one has at any given moment is a further matter of comparing and contrasting the relative worth of those desires and forming beliefs about how one ought to act in light of competing desires.

The process of comparing competing reasons for action involves the weighting of one’s desires and the development of a conclusion about which course of action satisfies the most or most important desires. The result of this process is what Al Mele calls a “decisive better judgment.”¹⁵⁹ Such judgments are the product of the agent’s own evaluation of her desires in an effort to determine which is most in keeping with her own interests. Several things distinguish simple desires from decisive better judgments. First, desires are necessary conditions of intentional action, but they are often insufficient conditions on their own. This is because our desires are very often in conflict and we cannot simultaneously take conflicting actions. Decisive better judgments adjudicate these internal disputes between our desires and thus produce sufficient conditions for action. Second, decisive better judgments are not strictly speaking desires, rather they are beliefs about how one will best come to satisfy her desires. The formation of such beliefs is an important element of strength of will

¹⁵⁸ Williams, Bernard (1981) “Internal and External Reasons” *Moral Luck* (Cambridge UP) 101-13.

¹⁵⁹ Al Mele (1987) *Irrationality* (OUP).

because will-power requires more than impulse, it requires decision making—that is strength of will requires beliefs about which desires to pursue and which to allow to pass. Without this sort of legislative function of the will impulsivity and conflict would guide our actions.

Still it is not enough to describe the legislative function of willing as strong simply because one moves from desires to beliefs about those desires. Two further elements are required for strong legislative functioning. First, the accuracy of one's evaluation of one's own desires must be reasonably sound. That is to say, a decisive better judgment contributes to strength of will only in cases where the judgment is sound with respect to the agent's actual desires. This is to say that someone who can reliably weigh her own desires will be advantaged in the exercise of will power because she will clearly see the alternatives before her and will not be easily confused about the merits of her actions. For example, consider two college roommates Sam and Pam. Suppose that Sam has some mild damage to his hypothalamus that causes him to only weakly consider his own desire for sleep. Pam on the other hand has no such damage and appropriately weighs the value of sleep against her competing interests. In this scenario Pam is likely to make much better decisions about her own interests. This is not to say that Sam is incapable of regulating his own actions, it is just that this self regulation will be more challenging for someone who does not reliably evaluate his own desires.

Of course, one does need to have any lasting brain damage to compromise the evaluation of one's desires. In many instances much more mundane circumstances cause our evaluations to be out of sync with our actual desires. Consider what happens

when you go the grocery store while hungry. The hunger should not affect your choice of groceries for the week, but perhaps should affect the pace with which you complete your shopping trip. Yet, to the contrary, when shopping while hungry we tend to shop longer, and buy more and worse foods. Things as simple as hunger, tiredness, distraction, etc. can lead to some serious challenges in evaluating the merits of desires we ourselves identify as clear and obviously motivational.

Finally, not every genuine reason one has for acting is both immediately clear and obviously motivational. It is also important that one be able to consider reasons for acting that are more abstract or difficult to identify given one's present context. For example, the moral imperative to respect persons is far too often absent from consideration in our beliefs and desires, but we nevertheless have a reason to respect others. It is not the case that eighteenth century slaveholders had no reason to reject slavery; they had such a reason and it was motivational (though not overridingly so.) However, in some cases they may have been sincerely unable to see the reason and it's motivating force lay dormant. It would be a mistake to call an apparently self-controlled slaveholder strong-willed, because he fails to exercise appropriate control over his own decision making process. It is tempting to see the disciplined military man as the paradigm of self-control, sharp creases on his uniform, fresh shave, crisp physical movements, etc., but if this man is acting without adequately considering how he should act, there is something seriously missing. Some will be tempted to split hairs here and say something like, "Even in the eighteenth century there was a reason to reject slavery" but "It is not the case that slaveholders themselves *had* a reason to reject slavery." This is a slippery move and contributes to a perniciously relativistic morality.

What we need is an account of reasons that is both sufficiently grounded in agency to explain the effectiveness of deliberation and abstract enough to provide some absolute principles—I have in mind here principles which would render decisions such as “Genocide is wrong,” “Sexism, Racism and the like are inappropriate.” These seem to be clear test cases about which an adequate morality could not fail to deliver correct answers.

Willing and The Fact/Value Distinction

Hume famously argued “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.”¹⁶⁰ Modern interpretations of Hume’s thought have developed his proposal into the view that desires (passions, in modern parlance, but not necessarily in Hume’s) are necessary and beliefs (again, reason, in modern parlance) are insufficient motivations to act. In reference to its inspiration this principle is sometimes called the Humean theory of motivation or alternatively an internalist theory of motivation. On the Humean view, if *A* has a reason, *r*, to ϕ , then *A* must have some desire associated with his ϕ -ing. We can contrast this with the idea of there *being* a reason for *A* to ϕ , unassociated with *A*’s desires. In the first case, *A*’s desire purportedly provides the motivation for acting on *r* and thus the link to one’s subjective motivational set explains the possibility of effective deliberation. Humeans argue that the second case will lead to a failure of effective deliberation because the requisite connection to desire is missing. To the contrary, acting on reasons removed from facts about one’s actual desires is possible when those

¹⁶⁰ David Hume *A Treatise of Human Nature*, (New York: Dover, 2003 edition), p. 264

reasons are psychologically connected to an alternative counterfactual version of oneself.

There can be a reason, r , for an agent, A , to act in some specific way ϕ , without A holding r in her actual motivational set, because, if there is an r to ϕ , and r is applicable to A , then A *has* an r to ϕ whether or not r stems from A 's actual motivational set. The trouble with claiming that " A *has* an r to ϕ in case r applies to A and despite A having no desire to ϕ " (or that slaveholders had a reason to release their slaves, despite having no such desire) is that they appear to be in conflict with the source of agent's motivations, i.e. their actual desires.¹⁶¹ While normative claims are not simple facts in the way that desires may be, nevertheless what one ought to do is a type fact about what one has reason to do.

On the Humean view if there is a reason for someone to do something, then she must actually have some desire that would be served by her doing it. However, a number of prominent philosophers have argued against the Humean requirement that reasons correspond with an agent's *actual* desires, in favor of a counter-factual account. Counter-factual accounts come in several varieties but agree in the basic view that agent's have a reason to act both when a reason corresponds to an agent's actual desires and when certain counterfactuals are true of the actual agent. So for instance, one has a reason to eat chocolate cake if one actually desires chocolate cake and one has a reason to eat cake if one would desire chocolate cake were one to know what chocolate cake tastes like. Having a counterfactual reason to act is motivational in the same way that an unrecognized desire can be motivational, upon reflection and with

¹⁶¹ See Michael Smith's (1994) *The Moral Problem* (Blackwell pub.)

better awareness or sensitivity to one's interests one comes to discover a hidden motivation.

The motivational force of counterfactual reasons for action stems from the purported interaction of two simultaneous worlds, the real world in which the reasoning agent actually exists and an imagined world in which the deciding agent is idealized. In this scenario we suppose that an agent confronted with a difficult choice might imaginatively look to her more ideal self for guidance in choosing between the available options. The resulting guidance from what one imagines as one's more ideal self constitutes a reason to act in accord with that guidance. Further, reasons to act stemming from consultation with one's more ideal self are motivational because they connect the subjective motivational set of one's future more ideal self with one's actual self. Although this connection is imagined—one isn't actually consulting a more ideal self—it represents a counterfactual motivating influence.

In a recent series of four studies researchers tested the motivational effect of exposing subjects to imagined versions of their future selves and found that such exposure was significantly motivational. The first study in the series involved outfitting participants with sophisticated virtual reality headgear and exposing them virtually to one of two conditions. In the first condition participants met with and interviewed a three dimensional virtual image of their present selves. In the second condition participants were exposed to age-progressed three dimensional versions of themselves and asked to interview their future selves. Participants in both conditions then engaged in a money allocating task in which participants were told to imagine allocating an unexpected \$1,000 windfall among four options: "Use it to buy something nice for someone special,"

“Invest it in a retirement fund,” “Plan a fun and extravagant occasion,” and “Put it into a checking account.”¹⁶² On average participants exposed to vivid images of their future selves saved more than twice as much as those exposed only to their current selves. Subsequent studies in the series inoculate these results against a priming bias and show that similar results can be obtained with less sophisticated equipment such as the inclusion of age-progressed avatars on a webpage. On such a computer based page participants were asked to allocating savings and spending as before, but in this subsequent study allocation was based on moving a digital “slider.” The slider could be moved right to allocate more money or left to allocate less. Further, the allocation corresponded with the avatar’s perceived emotional response. Thus, subject’s choices about allocating money to present concerns caused present-self avatars to smile or frown. Likewise, choices about savings affected the perceived emotional response of one’s age-progressed avatar. Participants exposed to their age progressed avatar allocated an average of 30% more of their income to savings, then did participants exposed to current-self avatars. This research is aimed at manipulating a phenomenon known as temporal discounting, but it also shows the way that exposure to an imagined, or counterfactual, version of oneself can alter one’s actual behavior.

If being exposed to a virtual image of one’s older self can motivate savings at double the expected rate, then it seems that counter-factual reasons for action can be significantly motivational. Christine Korsgaard offers a version of counterfactual motivating reasons arguing that practical reasons “are best thought of as establishing

¹⁶² Hershfield, H.E., et. al. (2011). Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research*, 48, S23-S27.

ideals of character.”¹⁶³ On this view an agent may come to act on a counterfactual reason when the principles that help to shape the agent’s actual character offer her a reason to act that is not in keeping with her current disposition. This is to say, if one generally guides one’s behavior by a certain set of principles (call this character) but in some specific actual circumstance is inclined to act out of character, those general principles that ordinarily determine one’s actions constitute a set of reasons to act as one ordinarily would. Put another way, even if you are not feeling quite yourself today—and so seem to have no reason to act as you ordinarily would—you still have a reason to maintain your character. Consider an example adapted from Gary Watson’s (1975) “*Free Agency*.” Suppose that after a vigorous game of squash I am so defeated and humiliated that I am overcome with a powerful urge to smash my opponent in the face with my racket. Ordinarily, I would have no such urge as this goes strongly against several principles constituent of my identity and which ordinarily guide my actions. To the contrary, I would ordinarily be inclined to walk over and shake my opponents’ hand and thank him for the match. On Korsgaard’s account I have a reason to walk over and shake my opponents’ hand despite not wanting to do so, it is what I would do if I was acting most like myself.

Michael Smith describes Korsgaard’s view as an “example” theory of counterfactual reasons because the agent has a reason to act in accord with the example of her more idealized self. In the case of the unusually angry squash player the player has a reason to follow the example of his more calm ordinary self and to walk over and shake hands. However, Smith argues that the angry squash player does not

¹⁶³ Christine Korsgaard, (1986) “Skepticism About Practical Reason” *The Journal of Philosophy*, Vol. 83, No. 1, p.18

have a reason to walk over and shake hands; to the contrary he has every reason to smile politely and walk away. This is precisely because the angry squash player is not his ordinary self and so should not follow the example of his ordinary self. On Smith's view trying to emulate one's calm self when overwhelmingly angry is very likely to lead to actually smashing one's opponent in the face—an unwanted outcome indeed.

However, Smith argues that counterfactual reasons for action can be rehabilitated under what he calls an “advice” model. On Smith's preferred version of counterfactual reasons, an agent has reason to act as her fully informed self would advise that she act in the agent's actual situation. Thus, the angry squash player does not have a reason to try to emulate the behavior of his calm rational self. Rather, the angry squash player has a reason to follow the advice that his more fully informed self would give to someone in his actual situation. Of course, the more idealized version of the squash player would have no need to walk away, *he* would be perfectly capable of politely shaking hands. But, because the actual squash player could not do so, he has a reason (based on the counterfactual advice of his idealized self) not to attempt shaking hands. In such a case the agent has a reason to act as his more fully rational self would advise he act, despite the disconnection of walking away from the agent's actual desires. Smith describes this view as an “anti-Humean theory of reasons” because the agent's reasons for action are not relative to her actual subjective motivational set but are identified with an idealized version of the agent.¹⁶⁴

In accounts of counterfactual reasons the idealized agent acts as a bridge between justificatory moral norms and subjective motivational states. On Smith's view

¹⁶⁴ Michael Smith (1995) “Internal Reasons” *Philosophy and Phenomenological Research*, 55 pp.109-31 reprinted in Smith (2004) *Ethics and the A Priori* (Cambridge)

this amounts to reasoning correctly from one's given beliefs and desires to a more "coherent and complete desiderative profile and evaluative outlook."¹⁶⁵ This decision procedure is comparable, Smith thinks, to Rawls method of reflective equilibrium. Rawls notion is that reaching decisions about morality and other important concerns are best thought of as a negotiation, or working out back and forth, between our "considered judgments about specific cases" and more general principles.¹⁶⁶ The upshot of reflective equilibrium is that it conveys the seriousness of conviction about one's individual considered judgments without allowing that the truth of those specific judgments depends only on one's conviction. Rather, the adequacy of specific judgments depends on the degree to which those judgments fit with each other into a coherent set of beliefs. For Smith, the counterfactual desires of our idealized selves bring together considerations about our actual desires with a reflective consideration the result of which is that "our more specific desires are better justified, and so explained."¹⁶⁷

Counterfactual accounts of reasons like Smith's are important here because they help to show how deliberation about general principles can be effective in producing actions. The effectiveness of one's conclusions about general principles of action is a significant part of strength of will because this is the space where the theoretical and practical elements of will meet. At its limit strength of will requires that one be able to move beyond being motivated solely by an antecedently given set of desires and that one come to act on the basis of reasons that one would find compelling in a more ideal

¹⁶⁵ *Ibid.* p. 23

¹⁶⁶ John Rawls, (1951) 'Outline of a Decision Procedure for Ethics', *Philosophical Review*,, 60:2:177–97, reprinted in *Collected Papers*, 1999, pp. 1–19.

¹⁶⁷ Smith p. 23

situation. Thus, in the really hard cases strength of will requires that one be able to act not on one's actual desires, but on the imagined desires (or advice) of one's more ideal self. It is the act of identifying sufficiently with one's imaginative reflections that imbues our idealized selves with motivational force and allows for significant shifts in one's character.

In summary strength of will requires a range of capacities and behaviors associated with the correct balance of legislative and executive components of practical rationality. First, will-power requires a baseline sensitivity to one's impulses, desires, and beliefs. Without this identification with one's self even trivial intentional action may be impossible, nevermind difficult acts of will. Second, it is imperative that one competently evaluate competing elements of her subjective motivational set. Without an evaluation of the relative significance of one's desires, willing becomes a hodgepodge without any focus and without the strength that such an evaluative focus provides. Of course, evaluation with a faulty faculty is nearly as problematic as a general failure to evaluate competing desires in the first place. Finally, it is not enough that one considers the subjective motivations one already possesses. It is important that one also consider the desires and beliefs about how to act that one ought to have. This assuming of a normative stance is crucial to making good decisions about how one acts and is thus an important element of strength of will. Strength of will requires that one be sensitive to all sorts of motivating reasons including desires, beliefs about those desires, principles of sound reasoning and counterfactual reasons one has for acting.¹⁶⁸ Of course, not every exercise of will power will require the same sensitivity to each of

¹⁶⁸ Sensitivity here is meant to emphasize receptivity to one's reasons, but also includes an element of reasons responsiveness.

these reasons for action, but each of these elements is important to some acts of will power.

Strong Internalism: The Logical Impossibility of Weakness of Will

One example of an especially strong connection between practical judgment and action is found in the account offered by R.M. Hare. Hare writes,

“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.”¹⁶⁹

On Hare’s view, a sincere belief in the appropriateness of action and that action are indistinguishable because they define one another. What it means to have a sincere belief in action is to act, and, an action simply is a sincere belief in the appropriateness of that action. Hare’s account is sometimes referred to as “strong internalism,” where ‘internalism’ refers to motivation for action being internal (or necessary) to practical judgment.¹⁷⁰ On this strong formulation, the prescriptions of practical judgment necessarily result in the motivation required to bring about the appropriate action. Note that on this view it is not just that practical rationality entails some motivation in the general direction of a good action. Rather, on the strong internalist view, practical judgment entails a sufficient motivation to ensure the specific action found appropriate. Thus, if I think I ought to do X, then, necessarily, I will be motivated sufficiently to do X. It seems to the strong internalist that if I think I ought to do X and am free to do X, then (*ceteris paribus*) I will do X. This leaves little room for

¹⁶⁹ Hare, R.M., *Freedom and Reason*, (Oxford: Oxford University Press, 1963) p.79

¹⁷⁰ Falk, W. D. (1947) "Ought' and Motivation", *Proceedings of the Aristotelian Society*, 48: 492-510

acting against one's better judgment and Hare is accordingly skeptical of weakness of will.

In an effort to explain the appearance of putatively weak-willed agents, Hare suggests two compelling options. On one account the 'judgment' that the weak-willed agent reports is not a sincerely held prescriptive judgment, but rather a description of external norms or expectations. This is referred to as an "inverted commas" case because the judgment which the agent is unmotivated to follow is not her own. Rather, she agrees that she "should" quit smoking, but is utterly unmotivated to do so. The inverted commas signal the fact that she does not genuinely assent to the judgment. This sort of situation is surely commonplace, but the agent is not weak-willed because she does act on *her* best judgment.¹⁷¹ Hare's other proposed explanation is that at the time of action the agent is overwhelmed by physical or psychological forces beyond her control. In this case, although the prescriptive judgment is sincerely held, it is literally impossible for the agent to act on that judgment. In cases like this one the apparently weak willed agent is really better described as compulsive, or lacking the appropriate faculty of free will. Hare is right to suggest that in both the compulsive and inverted commas case the reportedly weak willed person is really better described in another way. But, the key to Hare's skepticism is not that some prima facie cases of weak will are false alarms, but his view that genuine beliefs are always overridingly motivational.

Moderate Internalism: The possibility of weakness of will

Donald Davidson argues against Hare that weakness of will is possible, on the grounds that practical judgment is a process occurring over time. As Davison sees it,

¹⁷¹ Hare p. 189

the main skeptical concern with weakness of will is that a person's judgment about action and her motivation about action cannot come apart. This is the internalist thesis: practical judgment necessarily issues in motivation to action. Davidson, himself is an internalist, so one might imagine that he would follow Hare and reject the possibility of weakness of will. However, Davidson sees beyond the flat account of internalism and develops a more textured account of practical judgment. As Davidson describes it practical judgment involves a myriad of what he calls *prima facie* or conditional judgments. Such judgments are not themselves necessarily connected to action, that is to say, motivation is not "internal to" conditional judgments. Because *prima facie* or conditional judgments are a vital element of practical rationality on Davidson's view, we might describe his account as a moderate internalism.

The texture in Davidson's account of practical rationality involves three layers each of which contributes to a person's reasoned motivation toward action. The foundational layer of Davidson's account are beliefs and desires, which he calls "primary reasons."¹⁷² The interesting thing about primitive beliefs and desires is that they can and often do conflict with one another. The second layer in Davidson's account of practical reasoning is the formation of conditional or *prima facie* judgments. Conditional judgments are deliberations about a subset of beliefs and desires. Imagine a swirl of beliefs and desires permutable and combinable in many different ways, each of which potentially leads to new conditional judgment. These evaluations working evaluations, not final decisions and are thus *prima facie*. Of course, with a smorgasbord of conditional judgments on the table it is best to prefer those judgments which account

¹⁷² Davidson, Donald "Actions, Reasons and Causes." in *Essays on Actions and Events* pp. 4-19

for more of our beliefs and desires. It is best of all to prefer those conditional judgments which account for all of the relevant considerations—these Davidson calls “All Things Considered” *prima facie* judgments. Although the all things considered judgment is only one of the swirling mass of conditional evaluations, it is also the best evaluation among the bunch. The final layer of Davidson’s account of practical judgment is the “unconditional” or “all out” practical judgment. At this point one conditional judgment becomes a full blown intention and culminates in motivation/action.¹⁷³ Here Davidson’s internalism returns to prominence and in the final step of the reasoning process one takes action. In the best case scenario the final unconditional judgment, the action, is based on an all-things-considered conditional judgment.

Unfortunately, there is no necessary connection between one’s best, all-things-considered, conditional judgment and one’s unconditional judgment or action. It is possible that one’s inferior, or less than all things considered, judgment will become unconditional. There is no sufficient reason why one would choose to on an inferior judgment, but it is nonetheless possible. Thus, on Davidson’s view weakness of will is possible when one’s best judgment remains conditional, while an inferior judgment becomes unconditional. The fact that one has no sufficient reason to act on an inferior judgment makes weakness of will irrational. However, if one’s best judgment is only *prima facie* the logical difficulty Hare saw with weakness of will disappears.

Davidson’s textured conception of what it means to make a practical judgment makes conceptual space for weakness of will, but it is not clear that it is the sort of weakness of will we ordinarily have in mind. Many cases of weak-will involve an agent

¹⁷³ Unhappily Davidson has very little to say about how a judgment moves from being conditional to unconditional except to refer to this as an act of ‘will.’

acting against something stronger than a *prima facie* judgment. For instance, Michael Bratman, a student of Davidson's, offers the account of Sam, a depressed man who, despite recognizing clear better reasons to do otherwise, sits drinking a bottle of wine.¹⁷⁴ Upon being reproached by a concerned friend Sam says,

“I don't think it would be best to drink. Do you think I'm stupid enough to think that given how strong my reasons for abstaining are? I think it would be best to abstain. Still, I'm drinking.”

On Davidson's account, either Sam cannot hold the unconditional judgment, “I think it would be best to abstain,” or his drinking is unintentional. Sam's drinking appears to be intentional. He means to be drinking; he just thinks it stupid. The real trouble is that Sam's drinking also appears to be in violation of an unconditional evaluation. Sam does not think “Well, based on X, Y and Z, I ought to stop drinking.” Rather, he thinks, “All out—I really ought to stop drinking.” But he doesn't. Bratman's point is that it would be odd to suppose that weakness of will like Sam's is based on merely a *prima facie* judgment. It rather seems like people can and do act in ways that violate their stronger unconditional judgments.

Weak Internalism: The possibility of strict weakness of will

What is needed to make good on the commonsense account of weakness of will is something closer to what Al Mele refers to as “strict akratic action.” Strict akratic action is action, “performed intentionally and freely and, at the time which it is performed, the agent holds a judgment to the effect that there is good and sufficient

¹⁷⁴ Bratman, Michael (1979), “Practical Reasoning and Weakness of Will” *Nous* 13: p.156.

reason for his not performing [it].”¹⁷⁵ On Mele’s view beliefs and desires still provide the considerations for an overriding reason for having a motivation, what he calls “decisive better judgment.” But actually having the motivation to act involves the formation of an intention. Intentions on Mele’s view amount to the motivation to act. Mele writes, “just as Davidson maintains that there is a step between all things considered judgments and unconditional judgments, it is open to me to postulate a step between decisive better judgments and intentions.”¹⁷⁶

A step between decisive better judgment and intention makes intuitive sense of weakness of will because it provides an explanatory locus for the break between judgment and motivation. Weakness of will is possible on this account because the bulk of the motivational force contained in beliefs and desires circumvent judgment and act directly on intention. Distinguishing intention from judgment does not eliminate the motivational strength of judgment, but it does allow that motivational force may follow a path around, rather than through judgment. Decisive better judgment maintains a strong motivational force drawn from the beliefs and desires from which judgment is formed. This necessary connection between judgment and motivation maintains a weak internalism about practical judgment. The connection is weak in the sense that judgment and motivation may not necessarily agree in strength. A strong judgment, like a decisive better judgment, may be subverted motivationally when beliefs and desires act directly on intention.¹⁷⁷

¹⁷⁵ Mele, Alfred R., *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control* (Oxford: Oxford University Press, 1987). p. 7

¹⁷⁶ *Ibid.* p. 20

¹⁷⁷ Consider Mele’s account of John, a biology student who is assigned the task of drawing a drop of his own blood and determining its type under a microscope. Despite having an aversion to drawing his own

If motivations can effectively circumvent judgment acting directly on intention, then Davidson is wrong when he writes “If an agent judges that it would be better to do x than to do y, then he wants to do x more than he wants to do y.” However, Mele can maintain a weaker commitment, “If an agent judges that it would be better to do x than to do y, then he wants to do x, though not necessarily more than he wants to do y.” Recall the case of Sam, the depressed drinker. Sam’s reasons, all things considered, lead him to judge it better not to drink. Yet he has a reason for being motivated to drink. He thinks rightly that drinking the bottle of wine will ease his depression for a little while. Mele contends that the trouble in cases like this is not trying to find a reasons-explanation for akratic action. Rather, the trouble is supposing that a reasons-explanation alone is sufficient to account for all actions. Clearly if the strength of reasons was perfectly related to motivational strength, Sam would act on his better judgment. But the failure here is not failure to act on a reason; it is the failure to act on the best reason.¹⁷⁸

Watson’s Skepticism About Weakness of Will

Gary Watson offers a subtle account of weakness of will, maintaining that if there is no sufficient reason someone would act on desires against her better judgment, “we are entitled to be skeptical about the common view, and to conclude that the person

blood, John makes the practical judgment that the thing to do is immediately prick his finger with a pin and to do so with sufficient force to draw blood. In short, John has formed a decisive better judgment to prick his finger. Suppose, John even forms an intention based on his judgment and moves the pin toward his finger. However, when the needle touches his finger John stops. John consistently maintains the judgment that he ought to prick his finger, yet his intention to do so has been altered. *Ibid* pp.35-6

¹⁷⁸ A complete etiology of weakness of will requires an account of *why* someone would act against her best judgment and *how* someone could act against her best judgment. The present essay addresses only how someone could act against her better judgment. A complete answer to the ‘why’ question seems unlikely, but would almost certainly include some discussion of the proximity of an action’s desired consequence, the agent’s focus of attention and her strategies of self-control. See George Ainslie’s *Breakdown of Will*.

was unable to resist.”¹⁷⁹ On Watson’s account there are two possible answers to how an agent’s desires may entice her to -act against her better judgment: either she, *chooses* not to resist those desires, or, her *effort* to resist is insufficient. The first possibility presents a straightforward objection to akratic action. If an agent chooses not to resist her desire, this choice is itself a shift in the original judgment and thus the action performed does not violate the agent’s better judgment.¹⁸⁰ Further, Watson claims that if an agent chooses not to resist a desire she believes ought to be resisted, this is not a case of weakness, but recklessness.

On the second possibility, an agent’s effort to resist desires contrary to her better judgment might be insufficient; Watson considers three reasons this might be so. First, the agent might judge that the effort is not worth it. But, again, this suggests that the agent’s better judgment has changed or was insincere in the first place. Second, the agent might misjudge the amount of effort required to resist her recalcitrant desire. Here Watson seems to find that the action’s being caused by misjudgment implies that it is unintentional and thus not akratic.¹⁸¹ The final, and by elimination, leading possibility is that the agent makes an insufficient effort to resist her desires, because for her those desires are literally irresistible.

Weakness of will is distinguishable from compulsion *simpliciter* by normative considerations about what is normally thought to be a resistible desire. As Watson puts

¹⁷⁹ Watson, Gary, “Skepticism About Weakness of Will” in *Agency and Answerability* (Oxford: Oxford University Press, 2004) p.57

¹⁸⁰ This shift in judgment might however be derivatively akratic, if the original judgment at issue was meant to resist change. It is this failure of a judgment to resist alteration, or, to be too easily revised, that Holton (1999) refers to as the commonsense understanding of weakness of will. I disagree finding that the common view includes both strict and derivative akasia.

¹⁸¹ Watson also rejects the possibility of the misjudgment itself being intentional, but here Watson’s claim does not rely on the impossibility of doxastic akasia. Rather, the relevant point is that if misjudgment causes an apparently akratic action, the action seems to be unintentional.

it, “individuals we describe as weakly giving in to temptation are those who reasonably could be expected to have resisted or to have developed the capacities to resist.”¹⁸² Thus, weakness of will is the culpable failure to develop appropriate capacities of resistance, and thereby become susceptible to compulsion. Compulsion *simpliciter* is being, “subject to such strong desires that it is unreasonable to expect even a strong willed person to hold out.”¹⁸³

In Defense of Commonsense

Mele focuses on refuting the claim that misjudging the *effort* involved in resisting a desire cannot lead to intentional akrasia. He suggests, “misjudgment of the amount or kind of effort required to resist a pertinent desire quite properly enter into explanations of strict akratic actions.”¹⁸⁴ This is right in that misjudgment might accompany an akratic act and thus help to explain the act as a whole. For instance, I might misjudge the distance I need to keep from an apple pie to avoid being tempted. Suppose that as a result of my misjudgment I walk within smelling distance and forgetting my diet eat a slice of pie. By eating the pie against my better judgment I show myself to be weak-willed and it was my misjudgment that led me into temptation.

In general I support Mele’s effort, but it seems to me that the best strategy for defending the commonsense view of akrasia is to address Watson’s concerns about the weak-willed agent’s *choice* not to resist recalcitrant desires. Recall Watson’s claim that in choosing not to resist a desire, the putatively akratic agent makes a new judgment in keeping with her action. Watson’s argument depends on an exceptionally strong

¹⁸² Watson, Gary, “Disordered Appetites: Addiction, Compulsion, and Dependence” in *Agency and Answerability* (Oxford: Oxford University Press, 2004) p. 72

¹⁸³ *Ibid.* p. 73

¹⁸⁴ Mele, p. 27

internalism identifying choice with better judgment entirely. If, on the contrary, it is possible for better judgment to be circumvented by motivations, the choice not to resist need not reflect a modification of judgment. Rather, choice may be identified with intention, which is distinguishable from judgment. On this model the choice (intention) not to resist need not reflect a new judgment, only the indulgence of a contrary desire. Certainly some motivational connection to judgment must exist, but this does not seem to imply that best judgment is always overriding.

Consider for example the case of Sarah, a congenial and conscientious Kantian who works long hours and believes she has an imperfect duty to help others. One day on her way home from work she notices a man slumped on the sidewalk. She pauses and thinks to herself, “I should help this man.” Perhaps she also has a desire to go home at once and put her feet up. It seems possible that Sarah chooses not to help, indulging her desire to put her feet up, against her better judgment. The failure of her better judgment to tip the balance of her motivations is consistent with better judgment having some necessary motivational weight—just not with its always being overriding.

At this point the skeptic is likely to recall that Watson also warns of collapsing akratic action into recklessness. This is an important concern, but Sarah’s failure hardly seems reckless. On one account recklessness involves an agent taking “insufficient care making a judgment...but, having made her judgment, her desires match her beliefs perfectly.”¹⁸⁵ Clearly on this account Sarah is not reckless, her desires do not match her beliefs and there is no evidence that her judgment is formed with insufficient care. Watson himself seems to have in mind something more like the preanalytic account of

¹⁸⁵ Smith, Michael and Kennett, Jeanette, “Philosophy and Commonsense: The Case of Weakness of Will” in *Ethics and the A Priori* (Cambridge: Cambridge University Press, 2004) p. 70

recklessness as freely choosing poor consequences. But even on this account it is unclear that Sarah was reckless.

What if instead of going home to put her feet up, Sarah akratically goes to volunteer at a soup kitchen? She judges that the consequences of helping the man are likely to be better, since the soup kitchen will likely run without her, but nevertheless she is moved by her enjoyment of the kitchen's camaraderie. It seems odd to call volunteering at a soup kitchen rather than helping the man on the sidewalk reckless. After all she has not decided to make sloppy judgments, or utterly disregard her considerate judgments, she only occasionally gives in to somewhat lesser desires. This makes her weak-willed, but not reckless.

Sarah's case demonstrates the absurdity of collapsing all cases of intentional, non-compulsive choice of a lesser option into recklessness. If all such cases are reckless, then Sarah is reckless and this conclusion seems so implausible that we ought to reconsider the reduction. However, the strength of this argument against Watson's view lays in the conviction that calling Sarah reckless would be absurd. It certainly seems absurd to me, but others may have different views. If one finds that Sarah recklessly disregards her judgment by walking past the man slumped on a sidewalk then the argument offered here will be unconvincing. However, this seems unlikely precisely because Sarah clearly is at least moderately responsive to the reasons she has for acting one way or the other.

Sarah's choice to act in a socially responsible way, i.e. working in soup kitchen, suggests that she is not reckless, but this is not the main force of the example. Rather, it is the reasoning on which Sarah's actions depend that indicates most strongly that

she is not reckless. Sarah forms a broad evaluation about her situation (perhaps even an ATC judgment) the conclusion of which is that she should help the man slumped on the sidewalk. The act of consciously forming such a judgment indicates that Sarah values reasoning about her actions. Further, the fact that Sarah responds motivationally to this and other judgments indicates that she takes her judgments seriously. The paradigm of reckless behavior is the person who refuses to evaluate reasons for acting one way or another, the person who acts from caprice is reckless. It is not reckless to treat ones judgments seriously, quite the opposite, and some weak-willed people seem to take their judgments quite seriously.

One might still think that Sarah is reckless to the extent that she discounts her own judgment in intentionally choosing the lesser of her options. If Watson means to take this route it does indeed become impossible to distinguish between weakness of will and recklessness. However, on this view it also becomes necessary to distinguish between an ordinary form of recklessness as acting from caprice and a special case of recklessness as an intentional choice of lesser options. Watson's move to collapse weakness of will into recklessness only succeeds at the expense of disintegrating the idea of recklessness. It seems more reasonable to say that reckless behavior involves acting from caprice, while unduly limited responsiveness to judgment is better understood as weakness of will.

Conclusion

Despite strong objections I believe that it is possible to maintain a commonsense account of weakness of will. Of course, I have not here made an affirmative case for the existence of any actual cases of weakness of will. What I have attempted to do is

outline the prominent arguments for the possibility of akratic action and contribute a small supplementary account. My aim was to sketch a defense of the possibility of weakness of will against collapse into recklessness. The suggestion was that it is neither impossible nor necessarily reckless to freely and knowingly act in ways that are somewhat less than best.

In my view it is not impossible for one to freely and knowingly act in ways that are less than best because, as Mele suggests, one's judgment can be circumvented by one's motivation. This circumvention is possible if one supposes the existence of a motivational faculty both weakly connected to and partially independent of judgment. Weak connection of the motivational faculty allows that motivations are often supported by judgments, and weak independence allows that motivations may not correspond to judgments in strength. It is notable that this is a structural view of how mental events might occur and therefore not a claim about how the mind is structured, but about the conceptual possibility of akratic action.

In summary the possibility of akratic action can be defended against a collapse into recklessness on two grounds. First, there is at least one case, Sarah's, in which it is absurd to think that purportedly akratic behavior could be appropriately described as reckless. Second, as Sarah's case reveals, akratic behavior is often deeply rooted in the agent's conscientious judgments and calling such behavior reckless threatens to cause a fault in the idea of recklessness. Together I take these to be convincing reasons to believe that weakness of will is both conceptually possible and distinguishable from recklessness. Therefore, there is good reason to endorse the commonsense model of weakness of will.

Chapter 6

The Value of Willpower and Weakness of Will

We all know that willpower is a valuable thing, but we often fail to consider why. Willpower is valuable because we want to do things that are hard—in particular we want to do things that go against our given desires by producing motivations to act that run contrary to our natural inclinations. Strong-willed people are more able to act in ways opposed to their own natural inclinations than are weak-willed people. This chapter is a discussion of the ways that matters of moral value intersect with the phenomenon of willpower, including instances of weakness of will. To be clear, this is not a chapter on moral issues, rather it is an analysis of how morally significant cases help to explain willpower and weakness of will.

The chapter begins with a brief account of the mental events that are necessary for willpower and weakness of will to be possible. This section contrasts the elements of motivation that are given such as what we find pleasant or desirable with what we find good or important. In cases where the two conflict willpower is the capacity to act based on the latter instead of the former. Of course, we do not always find the right things good or important and willpower can be used in the service of evil. In the second section of this chapter I argue that in cases where willpower is used in the service of ill, there is good reason to think that it is bad on balance. On the contrary, there are cases where weakness of will works in the service of the good. cases of so-called "inverse akrasia" can be good on balance

and begin to articulate several of the ways that this might be true. In the third section I explore how failures of willpower are practical necessities in so much as willpower is a skill which must be developed over time. Finally I conclude by exploring the implications of the foregoing account for the experience of regret for moral failures.

Why we need willpower

Being self-aware entails that one has preferences, concerns, or interests in how things go--this amounts to having a perspective. The mirror self-recognition test gives us reason to suspect that a range of animals including Chimpanzees, Magpie's and Dolphins, are self-aware.¹⁸⁶ Current views of the self in psychology position the self as playing an integral part in human motivation, cognition, affect, and social identity.¹⁸⁷ Beyond merely having a sense of self, a perspective, humans also take a metacognitive stance by evaluating and regulating our perspective. This means that we engage in a process of self-criticism whereby we endeavor to change the way that we think and behave. Self-criticism is an essential step beyond our natural selves, beyond the preferences and desires that we discover in ourselves and toward the preferences we create for ourselves. Willpower is what we need to transcend our natural preferences in an effort to act in ways that we think are preferable.

Some will worry that self-criticism means self-denigration, or, that rejecting our "natural" selves is a recipe for disaster. These are legitimate concerns, but are misplaced here. By self-criticism I mean something more specific than a generalized

¹⁸⁶ http://en.wikipedia.org/wiki/Mirror_test

¹⁸⁷ Sedikides, C. & Spencer, S.J. (Eds.) (2007). *The Self*. New York: Psychology Press

rejection of oneself; I do not mean to suggest that self-control requires, or is benefited by, self-loathing. Quite to the contrary, self-loathing is detrimental to self-control.¹⁸⁸ However, self-criticism, the identification of troublesome thoughts or behaviors in one's life, can be a constructive process. Identifying one's own challenges and weaknesses is a valuable practice in the process of overcoming those obstacles. In short, self-criticism is not antithetical to self-compassion.

Consider that as infants we all learn rules of sleeping and eating, and eventually speaking and even "pooping" in the right ways and at the right times. We are trained to be like others in our community. Hubert Dreyfus describes this as the challenging but important first step in skill acquisition; it represents the beginnings of self-control.¹⁸⁹ This early foundation in self-control is of course directed by others, but following these rules is our earliest training in the exercise of willpower. It is by learning to follow our biological urges in socially approved ways that we begin to develop the skill of self-control. But here, as latter in life, self-control is only necessary because there is a conflict between one urge and another. On the one hand the young person has the urge to scream here and now, on the other she wants the approval and understanding of her family. Balancing these concerns teaches us the skill of self-regulation and is an early step in becoming a full member of the moral community.

¹⁸⁸ Leslie Becker-Phelps (2013) "Self-Criticism Can Sabotage Your Happiness; and Productivity" Psychology Today, <http://webcache.googleusercontent.com/search?q=cache:http://www.psychologytoday.com/blog/making-change/201309/self-criticism-can-sabotage-your-happiness-and-productivity>

¹⁸⁹ Hubert Dreyfus & Stuart Dreyfus (1988) Mind Over Machine, Free Press

It takes an act of courage to train a critical gaze at one's own biological or social predispositions. To do so risks self-hatred, but it is not the only alternative. One can courageously risk looking at themselves reflectively with an eye toward acting on what they find. For many of us, in many ways, we are unable or unwilling to even begin the process of self-examination—this choice of remaining whomever we are, as dictated to us by nature and society, is the safer choice in terms of avoiding internal conflict and weakness of will. But, surely it is preferable to be self-constituting and risk failure, then to meekly accept whomsoever one happens to be. Weakness of will is not possible for the blindly self-accepting, but it is a price worth paying for the promise of intentional identity formation.

The courage to engage in honest self-examination is a vital evaluative component of a larger project: becoming our own people, taking responsibility for ourselves. Reams of data indicate that we are, each of us, mostly directed by situational and biological forces that we do not create or endorse.¹⁹⁰ This view has led some philosophers to conclude that we must accept determinism. Derk Pereboom argues that the practical consequence of determinism is that social arrangements must reject moral responsibility.¹⁹¹ But, if we believe that we each have freewill, we can also hold out hope that we can choose our own lives despite the forces that push us around. This is part of the activity of intentional self-creation.

Libertarian accounts of freewill allow that people have an ultimate responsibility for at least some of their actions. According to Robert Kane the source of this

¹⁹⁰ Smith, Kerri (2011). "Neuroscience vs philosophy: Taking aim at free will". *Nature* **477** (7362): 23–5

¹⁹¹ Derk Pereboom (1995) "Determinism al Dente" *Nous* Vol 9 No. 1: 21-45
<http://www.class.uh.edu/faculty/tsommers/Freewill%2012/determinism%20al%20dente.pdf>

responsibility is that we are the "arche" of our actions.¹⁹² Arche is a Greek term which means cause, motivation, or, more technically, a sufficient condition. On this view we are responsible for those actions that occur because we will them to occur. It is this choosing of something as a path that makes us responsible for it. Freedom and responsibility are worth wanting because there are genuine choices to be made about one's life. We do not want to have freedom just to say that we are free. It is not a matter of fiat that freedom is better than compulsion. Rather, freedom is valuable and worth wanting because we have genuine choices to make about who we shall become. And, making genuine choices involves struggling with our own internal conflicts, struggling with weakness of will.¹⁹³

Willpower Gone Wrong

The relationship between value and willpower is commonly thought to be straight-forward, having willpower is good and weak-will is bad. But this account is mistaken in a number of ways. The first way that this account is mistaken is that having willpower can be, on balance, a very bad thing. It is also possible that failures of willpower can be, on balance, quite good things. In its practical effect the exercise of willpower is by itself neither good, nor bad. Willpower is a skill which by itself doesn't have a moral valence; rather willpower is a requirement of adopting a normative perspective. Although making moral judgments requires at first that one believe the practical question of how to act is not determined by our desires, not all resistance of desire is noble and not all concession to desire ignoble.

¹⁹² Robert Kane, *The Significance of Free Will* (OUP 1996) p. 224

¹⁹³ Frankfurt style examples make clear that making genuine choices cannot be reduced to having alternative possibilities.

Consider first the possibility of resisting one's natural inclination and desire successfully, but immorally. Imagine, for example, a dedicated assassin who stalks and eventually assassinates the President. The successful assassin will have dedicated huge amounts of time to planning, preparing for and eventually executing his assignment. This persistence of attention over time and avoidance of distraction require the kind of willpower that is unusual in people. Clearly this sort of action requires great strength of will. However, it is also clear that this is an immoral action. The fact that an action is difficult, that one must resist impulses of distraction, exercise great skill, etc. to perform an action, does not make it valuable.

It is tempting to say that we only exercise great effort on actions that we think are valuable (morally, prudentially, or otherwise.) But that is almost certainly untrue. For example, I have spent countless hours learning to juggle balls, rings, and clubs with some degree of proficiency. I am under no delusion that this is a prudent or practical way to spend time, let alone that this practice has any moral value. Nevertheless, the activity requires deferred gratification, concentration, impulse suppression and in general seems to take what we ordinarily call willpower. The fact that I am able to muster these skills of willpower together in order to juggle does not mean that juggling is morally valuable. It certainly is not valuable, but it does require willpower.

Consider an alternative type of case, one of so-called "inverse akrasia." The traditional account of the inverse akratic involves Huck Finn helping to hide his friend the runaway slave "nigger Jim." From the narrator's explicit perspective it was plainly the case that Jim should be arrested and returned to his owner. However,

when the chance to turn the runaway slave over to authorities arose Huck did something else. Jim and Huck were friends and it would have been hard for Huck to inform on Jim even though he clearly thought it was the right thing to do. In this moment Huck is weak-willed, he chooses to do what comes easily instead of resisting that natural impulse. Of course Huck was right not to turn over Jim, not simply because he was Jim's friend, because Jim was a person and people should not be held in slavery. In this case Huck's weakness of will serves the interests of more moral decision making.

Allison McIntyre and Nomy Arpaly have both contributed to discussion of inverse akrasia with incredibly interesting and novel approaches. Arpaly claims that when weakness of will is based on unconscious reasons, it can be more rational than acting on conscious deliberation. McIntyre's claim is that weakness of will may, in some instances, be more rational than revising one's better judgment without sufficient reason. On my view it never makes sense to argue that weakness of will is rational. Rather, weakness of will is always less than reasonable, but may be seen as important for other reasons.

Arpaly's view of weakness of will as "rational non-deliberative action" is plausible because deliberation can overlook overwhelming reasons to which the unconscious attends. As Arpaly puts it, "There are cases in which people do not deliberate but still act rationally, and there are cases in which people act as a result of deliberation but are acting irrationally."¹⁹⁴ It is true that people often do what they have overwhelming reason to do without conscious deliberation. In fact conscious

¹⁹⁴ Nomy Arpaly (2003), *Unprincipled Virtue*, (Oxford: Oxford University Press). P. 59

deliberation sometimes gets in the way. This leads Arpaly to conclude that actions against our better judgment, but in keeping with our interests may be rational. This shades toward a consequentialist account of reason and fails to acknowledge that weakness of will always involves the procedural error of failing to act as one believes one should. The correct process of practical reasoning involves maintaining coherence between beliefs and actions. The trouble is Arpaly's account of rationality cannot explain why the failure to do so is a failure at all.¹⁹⁵ This sort of radical externalism severs judgment from motivation and mischaracterizes the value in weak-will.

McIntyre's account of rational weakness of will is closer to the mark and relies on the tension produced between procedural concerns and substantive concerns. On her view we are sometimes faced with a choice between sacrificing coherence or sacrificing our best judgment. In such cases, she argues, it may be rational to act against our better judgment. Of course, this counts against the action's rationality, but it may count less than sacrificing better judgment. Consider the example of Dixon a houseguest who believes that he should confess to accidentally burning the rug in his room, but also believes that he will be unable to muster the courage to confess. If Dixon forms the intention to confess, he does so knowing that he cannot follow through. On the other hand, if he fails to form the intention to confess, he does not act as he believes he should. It is clear that Dixon has an incoherent

¹⁹⁵ Despite her claim to the contrary, "I still agree that every agent who acts against her best judgment is...less than perfectly rational as the schism indicates a failure of coherence in her mind." (Arpaly p.36) It seems that a failure of coherence only tips us off to the real source of irrationality, the faulty deliberation.

mental state and we should be inclined to pity poor Dixon, but what should he do? Can he escape being irrational in such a condition?

One option, a slippery one, is that Dixon might relinquish his judgment that he should confess. This seems wrong, Dixon should confess, he just can't. McIntyre argues, "lucid akrasia may well be Dixon's best move in the circumstances." (2006 p.307) This view is based on the premise that Dixon will not be confessing under any condition. McIntyre seems to think that it is more rational for Dixon to know that he is failing, than to rationalize away his failure. There is something important and correct about the view it can be a poor decision to rationalize failures to behave as we believe we should. To make an exception in a personal policy can open the floodgates of exceptions by shifting our attention to the exceptions and away from the policy. It is a much better strategy to recognize that one has had a momentary failure, to recognize that this is perfectly normal and to seek out an opportunity to recommit to one's goal. That being said, it does not follow from a case of weakness being better than revision, that weak will is the most rational option in that case.

The Practical Value of Weak-Will

McIntyre's account goes a long way toward showing that in some specialized instances weakness of will can be more rational than some alternatives, but it does not show that weakness of will is ever a rational choice. The central concern of Arpaly and McIntyre (with respect to WOW) is the question, "Is weak-willed action ever an instance of sound decision making?" I think the answer is clearly "no." However, I do think that weak-willed action is valuable even when the agent is not mistaken about the correct course of action as is the case with the assassin or Huck.

In the first instance there is an advantage to weak-willed action in so much as it is a practical requirement of coming to be strong-willed.

As I argued earlier, especially in chapter one, willpower is best thought of as a skill and skills are developed over time through repetition. Just as there is no person who is born writing a philosophy paper well or hitting a baseball well, there is also no person born controlling themselves well. Self-control and the willpower it requires is something that we develop through an iterative process and this process includes failure. Imagine a person who practiced hitting baseballs for years, but never missed a pitch. Or a student who wrote regularly for years without ever putting together a terrible thought. Is it hard, perhaps impossible, to imagine such a person? I presume such a person is impossible to imagine. At first blush this seems ridiculous because we have never encountered anyone who practiced without failure at anything.

Beyond the sense that an infallible person would be somehow inhuman, perhaps an angel or a god, there is something about failure that is practically important to how people learn. Failures teach us the ways that an action can be thwarted and in turn teach us how to improve our technique in an action. Consider the analogy to a great hitter like Henry Aaron, who, as a matter of fact, missed countless balls. There is nothing necessarily true about the fact that Aaron missed lots of pitches. It is technically possible that Aaron would be just as he is without ever having missed a pitch. But, in the real world this is not how things work. Aaron had to miss lots of pitches because he had to practice a lot before he became a great hitter. But, beyond that, missing lots of pitches helped Aaron become a better

hitter because it showed him the ways that hitting can be confounded by pitching, wind, lighting, etc. With willpower, like with hitting, our failures provide feedback on the ways that we can create successes.

The formation of willpower is a process that requires time and effort. There are instances of willpower that resemble the muscular account in that they occur once and then are over. Some people also hit a baseball once and are done. But, when we discuss hitters we are talking about a skill someone has developed in an iterative process over time and when we talk about strong-willed people, we are talking about people who have been through a similar iterative process. As a practical matter developing a characteristic strength of will is something that requires failure. In a practical sense we need failure both because we are not angels or gods and failures help to show us the ways that we can be strong.

The A Priori Value of Weak-Will

There is a second sense in which weakness of will is valuable and it is also practical in the sense that it involves the practical question of how one should act. When we take up the practical question "what should I do?" we are already presuming that the answer to that question might be different from the inclinations and desires that we already possess. It is this distance between our inclinations and our judgments that makes willpower important and weakness of will possible. Therefore, taking up the practical question "what should I do?" presumes the existence of willpower and the possibility of weakness of will. As R.J. Wallace puts the point, "We inevitably assume that we ourselves have this capacity when we undertake to deliberate about what we should do...the idea that we have it in our

power to determine what we shall do in ways independent of our given desires provides the natural context for our own deliberative activity."¹⁹⁶

Elsewhere Gary Watson writes, "If free will means the capacity to act irrationally, some will say, it is not a power so much as a liability, and it is far from clear that we should want it; it seems a weakness that we would be well rid of."¹⁹⁷ From Watson's perspective, God, or, a perfectly virtuous man, would not need to be free. This is because without any internal conflict about how to act or who to be there is no value in freedom. And presumably neither God, nor, the perfectly virtuous man experiences any internal conflict about doing the right thing. It is the internal conflict which makes weak-will possible that also makes freedom worth having. As Kane puts this point, "To be sure, weakness in the presence of avowed purpose or better judgment *is* motivationally perverse and irrational from one point of view. But the possibility of such "motivational perversity"—of having powerful motives to act against better judgments or to fail in sustaining purposes already formed—is, I think, the price to be paid for free will."¹⁹⁸ Of course, freedom is not enough for us to be the authors of our own identity, we must also take up the pen—that is we must exercise our capacity for affirming and choosing who we are to be.

R.J. Wallace describes our capacity for self-creation writing, "Human agents have the capacity for a sophisticated kind of rational agency, insofar as they can reach independent normative conclusions about what they have reason to do, and

¹⁹⁶ R.J. Wallace "Moral Responsibility and the Practical Point of View" in *Normativity and the Will* (OUP2006).

¹⁹⁷ Gary Watson, "Free Action and Free Will" in *Agency and Answerability* (OUP 2004) p. 188

¹⁹⁸ Robert Kane, *The Significance of Free Will* (OUP 1996) p. 155

then choose in accordance with such normative conclusions. This capacity presupposes that we are equipped with the power to choose independently of the desires to which we are subject.”¹⁹⁹ The point Wallace is making is twofold, first, we are not committed to act on antecedently given desires as the Humean tradition sometimes purports. Second, Wallace argues that we are also not committed to every exercise of instrumental rationality being in support of our normative commitments as the Kantian tradition suggests. Wallace’s view is that volitional attitudes are importantly distinct from both desires and normative commitment. This is not to say that desires and normative commitments do not necessarily have an impact on choice; it is only to say that neither desires nor normative commitments are necessarily decisive in producing action.

By distinguishing normative commitment from volition, Wallace aligns himself with a view much like Al Mele’s view that there is “a step between decisive better judgment and intention.”²⁰⁰ For both Mele and Wallace it is this step between judgment and intention or volition which makes weakness of will possible. Wallace goes on to make this point explicitly saying, “once we have this power it can be put to use in ways that are at odds with our own practical judgments about what we have reason to do. That is, we can treat our dispositions to do what we ought as a further desire from which we set ourselves apart, choosing to act in a way that is at

¹⁹⁹ R. J. Wallace (2001) “Normativity, Commitment, Instrumental Reason” *Philosopher’s Imprint* Vol. 1, No. 3

²⁰⁰ Mele, Alfred R., *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control* (Oxford: Oxford University Press, 1987). p. 20

variance with our reflective better judgment.”²⁰¹ Here we see Wallace essentially making the earlier point that weakness of will can be understood as the result of an internal conflict.

Recall the self-sabotaging behavior of dieters, shoppers and addicts. When they see themselves as successful at dieting, thrift, or kicking a habit, they are significantly more likely to “reward” themselves with the very thing they wish to avoid. The only reason that this makes any sense is because the agents treat their own normative dispositions as distinct from themselves in the same way that they earlier treated their own desires as distinct from themselves. It is the same capacity to distance oneself from desire in making normative judgments that produces weakness of will in those struggling with self-control. Weakness of will is valuable because it is made possible by an important element of agency, internal conflict. In discussions of freewill it also seems that this same internal conflict is what makes freedom worth having. In the end then freedom is worth having because we have genuine choices about our internal conflicts and weakness of will is a byproduct of such conflicts that serves to signify our having choices.

Weakness of Will as a Signifier

There is something extraordinarily valuable about our capacity to transcend mere impulse, habituation and desire in the effort to make rational judgments about our own actions. We want to be able to consider reasons for not doing what we ordinarily would do and this capacity is crucial to our moral agency. But the capacity to engage in this sort of self-evaluation also means that we sometimes distance

²⁰¹ R. J. Wallace (2001) “Normativity, Commitment, Instrumental Reason” *Philosopher’s Imprint* Vol. 1, No. 3 p. 10

ourselves from our own evaluations. When this cognitive distance between our motivation and our judgments is resolved in favor of some inferior motivation, we have weakness of will. Of course, weakness of will is not a good thing, but it is something that we should want to be possible for ourselves. The possibility of weakness of will indicates the capacity for self-determination and actual instances of weak-will indicate that an agent is making an effort at self-determination.

Weakness of will only becomes an issue for those individuals who take an interest in evaluating themselves and their commitments. It is this self-evaluation that leads to making changes in our natural behavior and it is trying to make such a change that creates inner conflict. For this reason weakness of will need not be seen as an exclusively troubling event in one's life; rather, it is valuable to see weakness of will as a marker of one's willingness to confront one's own cultural and personal baggage. Weakness of will is only possible for those individuals who choose to courageously risk looking at themselves reflectively with an eye toward acting on what they find. For many of us, in many ways, we are unable or unwilling to even begin the process of self-examination—this choice of remaining whomever we are, as dictated to us by nature and society, is the safer choice in terms of avoiding internal conflict and weakness of will. But, surely it is preferable to be self-constituting and risk failure, then to meekly accept whomsoever one happens to be. Weakness of will is not possible for the blindly self-accepting, but it is a price worth paying for the promise of intentional identity formation.

A tidy summary of the point here is that "the capacity for rational self-guidance necessarily contains the potential for self-alienation, because it requires a

capacity of motivational transcendence that can be turned against the disposition to be guided by normative conclusions.”²⁰² Because weak-will *and* rational self-guidance both require "motivational transcendence" we can use one as an indicator of the other. Of course it doesn't necessarily follow that "when these enabling capacities are misused their manifestations are themselves instances of self-determination.”²⁰³ Some instances of purported weak-will may be coerced; but, so long as we accept the capacity for self-determination and the capacity for self-alienation there is little reason to suppose that they only work together in ways we would prefer. To the contrary, evidence suggests that in many cases it seems that our efforts to assess our own values have disappointing results. These disappointments do not suggest a lack of self-determination. The much stronger explanation is that changing oneself is hard. Further, specific instances of weakness of will work as an indication of more than just a capacity for self-determination; weakness of will is an indication that one is making an effort to utilize the capacity for self-determination.

In summary, the link between self-determination and weakness of will is fairly well established and agreed to by many prominent and diverse thinkers. Wallace establishes the connection especially clearly, but Gary Watson agrees that “The capacity for rational self-guidance necessarily contains the potential for self-alienation.” Of course, Watson and Wallace disagree about the implications of this connection, but that they agree on the connection is notable. Likewise, this is the same point that Robert Kane makes with respect to freewill. The same basic point

²⁰² Gary Watson, “The Work of the Will” in *Agency and Answerability* (OUP 2004) p. 136 fn. 34

²⁰³ Ibid.

has also been made across popular culture from fortune cookies to the autobiographies of actresses. These observations are accurate and when we have a point of agreement from such diverse perspectives we should look into its value.

The importance of connecting weakness of will to self-determination has been overlooked for far too long. It is possible to extend this important observation and to see that there is something valuable we can salvage from weakness of will because of its connection to the capacity for self-determination. When we are weak-willed we very often take our own failures as a sign that we cannot be trusted, that our judgment is poor, or that we are in some other way deeply flawed people. These self-assessments may be in some ways accurate but they are also self-defeating, feeding into a cycle of failure and regret. Self-criticism is an important element in self-determination, but ample empirical evidence reveals that we are most successful in improving ourselves when we see failure as a part of the process of self-improvement. Before reviewing that evidence and showing how motivational failures can be harnessed to improve willpower I want to cover some alternative accounts of how weak-will might be valuable.

Weakness and Self-Compassion

Weakness of will is not good per se, but it is a pitfall associated with our capacity for self-determination. As a practical matter, if we make an effort to improve ourselves, we will fail sometimes, and, it is likely that we will fail more often than we succeed. So how can we increase our odds of success? And, what are we to make of our failures? An earlier chapter some data on the successful exercise of willpower. Some successful strategies focused on strengthening the brute force of

will by getting adequate rest and maintaining a healthy blood sugar level. Other strategies were more, well, strategic, including predicting potential sources of difficulty and making arrangements to avoid temptation. In this section we look at a difficulty that inhibits the successful exercise of willpower across a range of situations. Specifically, we will address how failure to succeed in early attempts at self-control (weak-will) can negatively affect future attempts and what strategies have been associated with overcoming this common hurdle.

We know that self-control relies on the ability to look critically at one's own desires and judgments, and but we can go too far in employing self criticism. At a certain point one's efforts at self-assessment can turn into feelings of regret, guilt, or remorse. The evidence on such negative self-assessments indicates that they have a deleterious effect on future acts of willpower and avoiding future collapse of will power. This is an unusual finding, but it suggests that all willpower challenges are effectively like dieting in one important respect. Dieting is a particularly difficult challenge in part because we cannot simply swear off food, one must eat. Similarly, in willpower challenges one cannot simply swear off self-criticism; rather, we must moderate our self-criticism and temper it with self-forgiveness and acceptance. Once we know that some specific behavior is in need of change, we are best served by not harping on our own failures in this regard. A new strategy that I hope this essay can provide is to see willpower failures as reflective of some broader commitment to self-determination. If we can think about our failures in terms of our overall positive commitment to the difficult challenge of taking responsibility for our own characters, we may be more successful in avoiding future failures of willpower.

A range of data supports the claim that feelings of anxiety, guilt, remorse associated with weakness of will tend to prime future failures of self-control. For instance, one marketing study found that when exposed to anxiety and sorrow producing reports of death on the news people were more likely to respond favorably to advertisements for luxury goods.²⁰⁴ Likewise, a study on drinking behavior asked people to record the number of drinks they anticipated having and then to record the number of drinks they actually consumed. Those who most regretted drinking heavily the night before, not those who drank most heavily, were the ones who were most likely to exceed their limits the following evening.²⁰⁵ A follow-up study also showed that those with more rigid limits tended to feel worse about consumption behaviors and accordingly drank more the following evening.²⁰⁶ Procrastination is another willpower failure about which we often feel guilty, and another case in which regret appears to lead to more procrastination, while self-forgiveness mitigates the problem.²⁰⁷ This effect has also been demonstrated in studies on gambling behavior and cigarette smoking.²⁰⁸ It seems that the harder you are on yourself after a will power failure, the more likely you are to have a relapse, psychologists call this a “disinhibition effect.”

²⁰⁴ Mandel, N. and S.J. Heine (1999) “Terror Management and Marketing: He Who Dies with the Most Toys Wins” *Advances in Consumer Research* 26 pp.527-532

²⁰⁵ Muraven, et. al. (2005a) “One too many: Predicting future alcohol consumption following heavy drinking.” *Experimental and Clinical Psychopharmacology*, 13, 127–136.

²⁰⁶ Mauren Muraven et. al. (2005b) “The Morning After: Limit Violations and the Self-Regulation of Alcohol Consumption.” *Psychology of Addictive Behaviors* 19 pp. 253-262

²⁰⁷ M.J.A. Wohl et. al. (2010) “I Forgive Myself, Now I can Study: How Self-Forgiveness for Procrastinating Can Reduce Future Procrastination.” *Personality and Individual Differences* Vol. 48 pp. 803-808

²⁰⁸ American Psychological Association Study “Stress in America” APA Washington D.C.: 2007

So why it is that being critical of our own failures doesn't help to correct those failures in the future? Isn't chastising oneself after failure meant to ensure that such failures do not reoccur? Doesn't saying nice things to oneself after failure license the failure? To better understand how self criticism can lead to disinhibition and future failure it is helpful to return to the biology we discussed earlier. Feelings of guilt and shame induce stress on the body with effects that are similar to low blood sugar or sleep deprivation. When we stress ourselves out, we weaken the efficiency of the prefrontal cortex responsible for impulse control and strengthen the midbrain responsible for basic impulses. Essentially, guilt and anxiety evoke a stress response in us that triggers increased desire and weakens self-control. Baumeister et. al. attribute this link between emotional distress and poor impulse control to an effort to avoid negative self-assessment or other ego threats by replacing self-awareness with awareness of more concrete elements of the environment.²⁰⁹ Unfortunately, this is exactly the opposite of what we need for success in high willpower activities. What we need in order to accomplish our most challenging goals is an active prefrontal cortex and a relatively quiet impulse system.

In an effort to test the effectiveness of different willpower strategies Claire Adams and Mark Leary arranged a study in which dieters would be asked to break their diets temporarily and then either primed with self-forgiving thoughts or given no such priming. The experimenters began by asking the subjects to pre-load by eating either a chocolate or glazed doughnut. The subjects were also asked to drink at

²⁰⁹ Tice, D. M., Bratslavsky, E., & Baumeister, R. F. (2001). Emotional distress regulation takes precedence over impulse control: If you feel bad, do it! *Journal of Personality and Social Psychology*, 80, 53–67.

least one entire glass of water to encourage a feeling of fullness. Then, subjects were asked to participate in what they were told was a separate “taste test” in which they would evaluate the flavor of various candies. In reality the candy was provided to test the dieters willpower following indulgence. Before the taste test began subjects in the “self-compassion” condition were told the following by a researcher.

“You might wonder why we picked doughnuts to use in the study. It’s because people sometimes eat unhealthy, sweet foods while they watch TV. We thought it would be more like the “real world” to have people eat a dessert or junk food. But several people have told me that they feel bad about eating doughnuts in this study, so I hope you won’t be hard on yourself. Everyone eats unhealthily sometimes, and everyone in this study eats this stuff, so I don’t think there’s any reason to feel really bad about it. This little amount of food doesn’t really matter anyway. Just wait a second and I’ll bring you the questionnaire.”²¹⁰

This passage was recited for each participant and was designed to elicit self-compassion in three distinct ways articulated by earlier research, specifically, self-kindness, mindfulness, and common humanity.²¹¹ The control and non-self-compassion groups were simply told “Just wait a second and I’ll bring you the questionnaire.” Then each person was given three large bowls of unwrapped candies and told to eat as much as they needed to complete the evaluation. Participants in the self-compassion group ate an average of 28g of candy, whereas those without the self-compassionate priming ate an average of 80g of the candy.²¹²

²¹⁰ Adams & Leary (2007) “Promoting Self-Compassionate Attitudes toward Eating Restrictive and Guilty Eaters.” *Journal of Social and Clinical Psychology*. p. 1129

²¹¹ Neff, K. (2003). “Self-compassion: An alternative conceptualization of a healthy attitude toward oneself.” *Self and Identity*, 2, 85–101.

²¹² Adams & Leary (2007) “Promoting Self-Compassionate Attitudes toward Eating Restrictive and Guilty Eaters.” *Journal of Social and Clinical Psychology*.

This suggests that guilt or remorse may exacerbate weakness of will, while self-compassion is a promising motivational strategy.

In a series of studies researchers at UC Berkeley tested the effects of self-compassion on subjects motivation "to improve personal weaknesses, moral transgressions, and test performance. Participants in a self-compassion condition, compared to a self-esteem control condition and either no intervention or a positive distraction control condition, expressed greater incremental beliefs about a personal weakness (Experiment 1); reported greater motivation to make amends and avoid repeating a recent moral transgression (Experiment 2); spent more time studying for a difficult test following an initial failure (Experiment 3); exhibited a preference for upward social comparison after reflecting on a personal weakness (Experiment 4); and reported greater motivation to change the weakness (Experiment 4). These findings suggest that, somewhat paradoxically, taking an accepting approach to personal failure may make people more motivated to improve themselves."²¹³

When we focus on chastising ourselves we fail to recognize the central element of our own failure, namely that it could never have happened if we weren't trying to do something difficult. By emphasizing the act of willing, as opposed to its failure or success we reorient ourselves toward the image of our better selves. In some ways the problem is like missing a turn while driving. Say that you know you need to make a left on to main street, but get distracted and miss the turn. You can look back toward the street you just passed, but this will be a further distraction and is likely to cause you to miss the next turn as well. In most cases this looking

²¹³ Breines JG, Chen S. Self-Compassion Increases Self-Improvement Motivation. Pers Soc Psychol Bull. 2012 May 29.

backwards is also quite unnecessary, after all you know that you've missed the turn. A better strategy is to keep looking forward toward a new turn that may eventually get you where you need to go.

The importance of being weak-willed is that it reflects something valuable about us, that we are trying to engage our own failures. It is counterproductive to dwell on those failures once they have been identified. However, if we take our failures as a signal that we are making an effort to something hard and worthwhile we are more likely to succeed. Louis C.K. has made a point of highlighting that he is an ordinary person who tries and fails at all sorts of things. For instance, when he sees a soldier getting on plane he thinks about giving the soldier his first class seat. Louis says, "The soldier "giving his life for his country (he thinks)...it's scary but he's doing it. I should trade with him--I never have, let me make that perfectly clear...--but I was actually proud of myself for having thought of it."²¹⁴ Louis is pointing out the absurdity of feeling good about *not* doing what he thinks he should. The experience of weak-will is commonplace and Louis is saying that it can feel good. Of course, he is right, and I argue that this experience of feeling good about our judgments despite our weak-wills is important in our becoming better people.

Weakness of Will and Moral Licensing

Visualizing the process of reaching one's goals is an important element of willpower, without setting goals for ourselves there would be no need for willpower. But, there are a number of ways that willpower can be weakened, we set unrealistic goals, are depleted by other challenges, "reward" ourselves by giving in to

²¹⁴ <https://www.youtube.com/watch?v=F-JFPePh6yM>

temptation, sometimes we just get insufficient rest or food. These reasons and more like them help to explain why we sometimes fail to maintain our resolutions or to act in the ways that we think we really ought to act. We need willpower precisely because our actions are not what they should be, but resolving the tension between who we are and who we should be is a difficult balancing act. On one hand self-evaluation requires that we imagine ways in which we might be more ideal, perhaps even trying to imagine how our more ideal selves would advise that we should behave in our less-than-ideal circumstances. On the other hand, an idealized self image is unattainable, imagining such a self threatens to generate an impossible goal, and for a variety of reasons imagining an idealized self can contribute to weakness of will.

In one way the image of an idealized self can present us with a goal, something to strive for, a set of improvements on the way that we cope with life. However, one potential pitfall of noticing our own flaws and imagining ourselves differently is that the image we create may be so vastly different from ourselves that we can scarcely see how to get from where we are to where we want to be. Recall the people who thought that they would exercise more, despite the evidence that they had unrealistic expectations. Those individuals, like us, ran into the problem of thinking about their future selves in ideal circumstances or with ideal powers of self-control and judgment. Studies like that one show clearly that a strategy of realistically imagining obstacles and challenges is vital to the exercise of willpower. Without identifying the specific hurdles one must face in order to change, the idealized self image is merely fantasy. This kind of fantasy can be problematic

because it leaves one with the belief that one should act in some way quite removed from the ways in which one does in fact act; and, without a clear path from the present self to the imagined future self the agent is left in the akratic position of believing she should act differently, but not being able to scratch out exactly how to do so. In this context goals are valuable conceived of as a series of means. Otherwise, goals tend toward fantasy and rely on magic to produce change--this is a recipe for weakness of will.²¹⁵

Consider an example, a couch potato throughout college John has decided that he should be fit, in fact he thinks it important to be able to run a marathon. John believes that he should run a marathon, not today of course, but eventually. John sees that his present self is not able to run a marathon, but believes that in the future he should be able to do so. In effect, John sees something in himself which needs improvement and has set a goal which will constitute successful self-improvement. The problem is that if John stops here he will always think that he should be able to run a marathon, but will never actually be able to do so. Setting a distant goal is not going to bring results without a recognizing the intermediate steps to reaching that goal. As the proverb says, "a journey of one thousand miles begins with a single step."²¹⁶ One problem of weakness of will is very often that we imagine the desired result, but not the process by which it comes about.

A related difficulty is that imagining an idealized self is often enough to make people believe that they are acting in accord with their goals of improvement and

²¹⁵ See Dewey John Dewey. "Habits and will," *Human Nature and Conduct: An Introduction to Social Psychology*. New York: Modern Library (1922): 15-42

²¹⁶ Lao Tzu (600BC?) Tao Te Ching, sect. 64

seeing progress toward one's goals is strongly associated with weakness of will. The association between perceived progress toward a goal of self-improvement and the failure of willpower regarding that goal is one of the genuine oddities of human psychology. It seems that we are willing to engage in what Kelly McGonigal calls "willpower hypocrisy" succeeding in self control only to see those successes license or justify future failures.²¹⁷ McGonigal cites research by Monin and Miller who surveyed and then tested support for the use gender stereotypes in a group of students at Princeton University. They first asked the students to answer one of two questionnaires. The first asked questions about women generally, such as, "Most women are better suited to stay at home taking care of the children than to work." The second, asked more limited questions, such as, "Some women are better suited to stay at home taking care of children." Predictably, the students with the first group of statements tended to strongly disagree, while responses in the second group were more neutral. Less predictably, when participants were then asked to rate candidates for a job, male and female, subject performance was associated with an anti-sexist bias. Those who had strongly disagreed with sexist statements were more likely to discriminate against women in hiring than were those who responded more neutrally to more neutral statements.²¹⁸ It seems that the earlier rejection of sexism granted permission to engage in subsequent sexist behavior, psychologists call this effect "moral license."

²¹⁷ Kelly McGonigal (2012) *The Willpower Instinct* (Avery Press) p. 83

²¹⁸ Benoit Monin and Dale Miller (2001) "Moral Credentials and the Expression of Prejudice" *Journal of Personality and Social Psychology* Vol 81, pp. 33-43

What is especially notable about the Princeton sexism study is that the subjects were not licensed by *doing* anything to act in defense of women's rights. Rather, the subjects apparently excused their own sexism because they disagreed with sexist statements. A related study focuses on how considering donating to a charitable cause licenses consumer choices. The study showed that when given the option between a utilitarian purchase, like a vacuum cleaner or strictly functional sunglasses, and a hedonic purchase, like designer jeans or posh sunglasses, a control group only made the self-indulgent purchase about twenty six percent of the time. However, when subjects were asked to suppose that they had volunteered to work four hours a week doing community service and asked to decide between spending that time either educating children or improving the environment, the rate of self-indulgent purchase shot up to above sixty percent. Likewise, rate of around fifty six percent self-indulgent purchases were obtained by telling participants to imagine having agreed to donate \$100 to a charitable organization and then to determine which organization should get the money. Part of why this is startling is that the participants did not actually donate any time or money to a charitable cause. Participants were twice as likely to indulge their own interest in frivolous or extravagant purchases because they had merely thought about charitable giving.²¹⁹

If thinking about doing something good leads to a failure of self-control we should have some concerns about the efficacy of imagining the ways that our future self could be better than our present self. In part this might be addressed by simply focusing moral progress on the deficiencies we notice in ourselves at the present.

²¹⁹ Kahn, Uzma. and Ravi Dhar (2006) "Licensing Effect in Consumer Choice" *Journal of Marketing Research* Vol. 43 pp. 259-66

This has the advantage of focusing on obstacles to be overcome which we have seen tends to willpower enhancing. The strategy of focusing on the present self also has the advantage of not presenting an idealized future self to license indulgent behavior. But, there are two limitations to the strategy of focusing on the present self. First it makes long term planning quite a challenge—this is important, because we tend to live quite a long time and focusing only on immediate goals can have detrimental long term effects. Recall for example the research on current rates of saving for retirement. Second, focusing on one’s real successes also appears create a licensing effect, so it seems that the willpower challenge is not that we idealize ourselves but that we irrationally reward ourselves. To see how this might be so, let’s begin by discussing the effect of reflecting on real success.

In one study Ayelet Fishbach and Ravi Dhar surveyed a group of women and asked them to report, among other things, their desire for weight loss. They were then asked to measure progress on this goal based in two different ways. One group was asked to color in their progress on a wide chart another asked to color in progress on a narrow chart. Earlier studies had revealed that coloring the wide chart makes people feel as if they have made greater progress as compared to the narrow chart. Then, after the survey, participants were asked to choose a parting gift of either an apple or a chocolate bar. Eighty-five percent of those primed by reflecting on their high-progress chose the chocolate bar, while only fifty-eight percent in the low progress condition chose the chocolate.²²⁰ When we actively consider our successes toward a goal we are much more likely not exactly to lose control, but to

²²⁰Dhar, Ravi & Ayelet Fishbach (2005) “Goals as Excuses or Guides: The Liberating Effect of Perceived Goal Progress on Choice” *Journal of Consumer Research* Vol. 32:3 pp. 370-377

choose in a way that is opposed to our goals. This is a very similar effect to the licensing effect of merely imagining a charitable gift. It is also related to an established “halo effect” in consumer behavior.

When consumers associate something perceived as healthy with something unhealthy they are much more likely to purchase the unhealthy indulgence than they are when the healthy option is removed—this is called the halo effect because the healthy option seems to cast a halo around the unhealthy choice. For example, when McDonald’s began to offer salads on their menu the sales of their least healthy option, the big mac, rose dramatically. Some of this could be attributed to higher general sales, but sales did not account for all of the increases. To test this effect a group of researchers set up a faux fast food restaurant and had people choose from menus with and without healthy items. On one menu the subjects could choose between a bacon cheese burger, fried chicken sandwich or fried fish sandwich; on the other menu a veggie burger was included as a healthy option. Subjects given the more healthy option tended to choose the least healthy option, the bacon cheese burger, at increased rates.²²¹ A similar study confirms that picturing healthy items, like strawberries, arranged to complement unhealthy items, like a can of cola, dramatically raises the perceived value of the unhealthy item. In fact, when pictured together the cola appears more valuable than the strawberries, while the opposite is true when they are pictured apart.²²² This research shows that is that when we see ourselves as making progress toward our goals we tend to wrap

²²¹ Wilcox, et.al (2009) “Vicarious goal fulfillment: When the Mere Presence of a Healthy Option Leads to an Ironically Indulgent Decision” *Journal of Consumer Research* Vol. 32 pp. 380-393

²²² Fishbach, Aeyet & Y. Zhang (2008) “Together or Apart: When Goals and Temptations Complement Versus Compete” *Journal of Personality and Social Psychology* Vol. 94 pp. 547-559

our decisions in a halo which ironically licenses behavior antithetical to our better judgment.

It should be clear at this point that reflecting on ones progress or self-improvement can have the effect of licensing self-indulgent behavior, but in concluding this section it is worth taking a pause to examine the sort of self-sabotage that such licensing represents. It is straightforwardly counterproductive for a dieter to “reward” her hard work with a chocolate bar. The reward undoes a great deal of the hard work that went into dieting in the first place. The sort of reasoning a play here borders on the irrational. Consider how we would respond to a heroin addict who thought it reasonable to reward her sobriety over the course of two days by having a fix on the third. Of course in many cases there is no conscious deliberation licensing the weak-willed behavior, but it is intentional nonetheless. In some cases perhaps the dieter thinks “I have really shown tremendous resolve in sticking by my diet. I deserve a reward for being so strong.” So far this is reasonable enough. I say reasonable enough because showing resolve in one’s commitments is really its own reward, but invoking a strategy of external reward can be motivational. The wild leap in logic occurs when the dieter then supposes, “As a reward for being strong about my diet I will stop sticking by my diet.” If being on a diet, or more likely losing weight, is what the dieter really wants, then a reward which undermines that goal is no reward at all. This is like saying to a boy scout “Great, now that you have earned all of those merit badges your reward is that you get to throw them all away!” This suggestion would be met with horror by many, perhaps most, Boy Scouts. They worked hard for those badges and usually want to keep

them. It is tempting to write off this sort of licensing behavior as a sort of weak-willed irrationality, but that would be hasty.

Consider the example of “rewarding” the merit badge collecting Boy Scout by encouraging him to destroy his merit badges. I have said that most would balk, but surely some boy scouts do this very thing. What kind of Boy Scout would want to destroy his merit badges? Perhaps the scout who secretly always detested wearing a uniform, tying knots, and meeting with other scouts? It is not difficult to imagine a young man who under great pressure to be good scout completes many merit badges and finds at the end great satisfaction in leaving scouting altogether. Such a person might relish destroying his merit badges as an act of rebellion or an expression of his true feelings toward the effort that he put in toward a goal that was not his own. And this is the point, we are often tempted to act in ways that sabotage our own efforts because we do not particularly identify with those efforts. Like the disgruntled Boy Scout, many dieters, addicts, and others struggle with tasks that tax the capacity for self-control because they identify with the temptation to quit putting forth the effort or to stop resisting the temptation. Rewarding ourselves for working hard on a diet or addiction recovery by sabotaging that effort only makes sense if we see ourselves as basically committed to the unhealthy lifestyle and not to recovery. Kelly McGonigal captures this challenge of self-control when she writes, “Moral licensing turns out to be, at its core, an identity crisis.”²²³

²²³ Kelly McGonigal (2012) *The Willpower Instinct* (Avery Publishing) p. 104

Conclusion

We all have the capacity to direct our own lives for the better, but it is a scary thing to do. In order to guide our own lives we must first take an honest look at ourselves with an eye toward improvement. This step of self-awareness opens us up to the possibility that our self-esteem is based on a mistake--we want to like ourselves and this is much easier to do without self-examination. For those who are courageous enough to try to improve things may get worse before they get better. Once we have identified ways to improve we are in the position of having to regress, to accept ourselves as less than we want to be, or, we must start the difficult work of change. Change very often means failure, and again this can be painful. The upside of failure is that it reflects something valuable in ourselves. Those who try and fail are the courageous ones who are making an effort. We depend for our improvement on two sorts of people, angels who display perfection and those mere mortals who try and fail and try again. Since angels are in short supply mortal failures are tremendously important. If we can come to see that our weakness of will indicates a core strength we move closer to success. Weakness of will is not the foundation of strength, but it is only possible because of our most important strength--the capacity for self-determination--and recognizing this connection when we fail makes us stronger.

Conclusion

Weakness of will is a perennial philosophical puzzle, a practical problem and something about which new empirical literature is available. This dissertation uses the emerging empirical literature to help make sense of long standing questions about practical rationality and intentional action. Philosophers often think that science should depend on philosophy for the identification of questions and determinations about what would count as good answers and methodologies. I agree that these are proper roles for philosophy, but I do not engage science in that way here. Rather, I rely on the scientific literature to help inform my conception of practical rationality.

Empirical literature on self-control or will power begins with Walter Mischel's marshmallow experiment. The interesting thing is that the children then tried to wait using various strategies. Some stared at the marshmallow, others, played in a corner, one even fell asleep. These children were tracked into adulthood and the study revealed that children who were able to wait were also likely to have higher SAT scores, higher educational achievement, higher sense of self-worth, better ability to cope with stress and less drug use. The most effective resisters were those who distracted themselves. This result has been replicated in other contexts and it seems that willpower is a strategic faculty which depends on some physiological resources.

Compare willpower with another skill like hitting a baseball or writing a philosophy paper. These activities depend on physiological resources, they are taxing in the short run and we can develop stamina at them. People who are hungry

or tired do not perform especially well. However, these are activities that reward practice for reasons beyond physiology. We learn to anticipate the curve when a pitcher is ahead or to recognize the structure of philosophical arguments. That there is a physiological component to willpower is not a surprise, but reducing willpower to the physiological misses the point that some willpower strategies effectively cope with limitations on physiology.

Beyond offering an account of willpower as a skill I have tried to extend this model in making sense of the traditional problem of weakness of will. On my view it is best to see weakness of will as on a continuum between compulsion on the one hand and recklessness on the other. On the side of coercion I have argued that there are at least three plausible theories about freedom of will which are compatible with the possibility of weakness of will. I have not taken a position on which account of freedom is best. However, there is a convincing case to be made that we have freewill in a sense robust enough to support morality and that there is still room for the commonsense view that we can be weak-willed. If none of the options outlined are robust enough to ground an adequate account of human freedom, the problem is much more widespread than any concern about willpower especially.

Against the other end of skeptical accounts I have offered a description of the mind and practical rationality that makes room for the commonsense conception of weak-will. Strong internalism along the lines of Plato or Hare requires that purported cases of weak-will must be based on an error in understanding because evaluation and motivation are necessarily connected in terms of direction and strength. The seed of internalism is the very important insight that one's evaluation has no

importance or meaning unless the evaluation can lead to action—not responding to our judgment is recklessness. The trouble raised by weakness of will is that we need an account of evaluation that is not epiphenomenal and that is not physicalist in a strong enough sense to necessarily determine outcomes. I have argued that there is room for multiple judgments with opposing directions for action. Each judgment on my account must have some connection to action—each judgment represents a pathway to intention. However, there is no reason to suppose that our best judgment from a rational perspective is the most motivational. In sort, we can be motivated by lesser judgments without being reckless.

The value of motivational failure, weakness of will, is that it is only possible because of our attempts to do something difficult but worthwhile. If we never try to do something difficult, than we might go without failure. But, of course, we should try to do difficult things. Weakness of will is not possible for those who merely drift along through life on whatever course they happen to be thrown. Weakness of will is made possible when we make an effort to not simply follow along with our urges; we expose ourselves to motivational failure when we try to actively decide who we will become. If intentional self creation is valuable (and it seems to be), weakness of will is also valuable. Weakness of will is not a good thing *per se* (after all it is failure), but it does reflect something good about us. As R.J. Wallace puts the point, weakness of will is "a hazardous byproduct of the capacity for self-determination."²²⁴ Weak-willed people aim to take responsibility for their lives, to engage in intentional self-creation, and this is why they *can* fail. As a practical matter it is important to

²²⁴ R. J. Wallace (2001) "Normativity, Commitment, Instrumental Reason" *Philosopher's Imprint* Vol. 1, No. 3 p. 10

keep this in mind when we fail as a means of combating the self-doubt, guilt and regret which can often stand in the way of progress.

The end of this dissertation begins to touch on some themes of practical moral importance but for the most part this is an effort to engage with and clarify issues in metaethics and moral psychology. On my view the work done here should be seen as laying ground work for some future projects of a more practical nature. Specifically, the conception of willpower as a skill which can be developed over time and dependant on strategies of implementation has a direct bearing on the ways that we conceive of addiction in the medical and political spheres. Of special interest to me are issues of triage, addiction treatment, punishment and consent to treatment with addictive substances.

Bibliography

Ainslie, G. (2001). *Breakdown of Will*. Cambridge, MA: Cambridge University Press.

Ainslie, G. (1999). The dangers of willpower. In O.-J. S. Ed. Jon Elster, *Getting Hooked: Rationality and Addiction* (pp. 65-92). Cambridge, MA: Cambridge University Press.

Arpaly, N. (2003). *Unprincipled Virtue*. New York, NY: Oxford University Press.

Ayer, A. J. (1954). Freedom and Necessity. In A. Ayer, *Philosophical Essays* (pp. 271-84). London: Macmillan.

Banaji, M. &. (2010). Attitudes. In D. G. S. Fiske, *Handbook of Social Psychology* (pp. 348-388). New York: Wiley.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist* , 1175.

Basten, U. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences* , 217-267.

Baumeister, et. al. (2012). Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology* , 1318-1335.

Baumeister, et. al. (1998). Self-Control as Limited Resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology* , 777.

Baumeister, R. e. (1998). Self-Control as Limited Resource: Regulatory Depletion Patterns. *Journal of Personality and Social Psychology* , 774-789.

Baumeister, R. (2007). The physiology of willpower: linking blood glucose to self-control. *Personality and Social Psychology* , 325-356.

Becker, H. (1960). Notes on the concept of commitment. . *American Journal of Sociology* , 32-44.

Becker-Phelps, L. (2013). Self-Criticism can sabotage your happiness and productivity. *Psychology Today* .

Beckner, M. (1959). *The Biological Way of Thought*. New York, NY: Columbia University Press.

Bowen, S. &. (2009). Surfing the Urge: Brief Mindfulness-Based Intervention for College Student Smokers. *Psychology of Addictive Behaviors Vol. 23* , 666-671.

Bowen, S. e. (2009). Mindfulness-Based Relapse Prevention for Substance Use Disorders. *Journal of Consulting and Clinical Psychology, Vol. 78* , 295-305.

Brandsma, e. a. (1980). *The Outpatient Treatment of Alcoholism: A Review and Comparative Study*. Baltimore, MD: University Park Press.

Bratman, M. (1979). Practical Reasoning and Weakness of Will. *Nous* , 156.

Breines, J. G. (2012). Self-Compassion Increases Self-Improvement Motivation. *Personality and Social Psychology Bulletin* .

Britton, e. a. (2010). The Contributuon of Mindfulness Practice to a Multicomponent Behavioral Sleep Intervention Following Substance Abuse Treatment. *Substance Abuse* .

Brown, C. (1994). Are Drug Addicts Unfree? In E. S.-F. Brown, *Drugs, Morality and the Law* (pp. 51-78). New York, NY: Garland Press.

Bufe, C. (1998). *Alcoholics Anonymous: Cult of Cure?* Tuscon, AZ: Sharp Press.

Carlson, R. J. (2009). Unrealistically Optimistic Consumers: A Selective Hypothesis Testing Account for Optimism in Predictions of Future Behavior. *Journal of Consumer Research*, Vol. 35, No. 5 , 810-822.

Carnap, R. (1947). *Meaning and Necessity*. Chicago, Il: University of Chicago Press.

Carroll, M. F. (2003). Severe hypoglycemia in adults. *Reviews in Endocrine and Metabolic Disorders* 4(2) , 149-157.

Chalmers, D. &. (1998). The extended mind. *Analysis* , 7-19.

Couchman, J. J., Coutinho, M., & Smith, J. (2010). "Rules and Resemblance: Their Changing Balance in the Category Learning of Humans (*Homo sapiens*) and Monkeys (*Macaca mulatta*)". . *Journal of Experimental Psychology: Animal Behavior Processes* , 1287.

Davidon, D. (1980). Freedom to Act. In D. Davidson, *Essays on Actions and Events* (pp. 63-81). New York: Oxford University Press.

Davidson, D. (1988). Actions, Reasons, Causes . In D. Davidson, *Essays on Actions and Events* (pp. 4-19). New York, NY: Oxford University Press.

Dennett, D. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting* . Cambridge, Mass.: MIT Press.

Dewey, J. (1922). *Human Nature and Conduct: An Introduction to Social Psychology*. New York, NY: Modern Library.

Dhar, U. K. (2006). Licensing Effect in Consumer Choice. *Journal of Marketing Research* , 259-66.

Ditman et. al. (1967). A controlled study on the use of court probation for drunk arrests. *American Journal of Psychiatry* , 160-163.

Doris, J. (2002). *Lack of Character*. New York, NY: Cambridge Press.

Dreyfus, H. (1988). *Mind Over Machine*. New York, NY: Free Press.

Everitt, B. J. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature and Neuroscience* , 1481-1489.

Everitt, B. J. (1996). Neurobehavioral mechanisms of reward and motivation. *Current Opinion in Neurobiology* , 228-236.

Falk, W. D. (1947). Ought and Motivation. *Proceedings of the Aristotelian Society* , 492-510.

Fedorikhin, S. a. (1999). Heart and Mind in Conflict: The interplay of affect and cognition in consumer decision making. *Journal of COConsumer Research* , 278-292.

Feinberg, J. (1970). What is so special about mental illness. In J. Feinberg, *Doing and Deserving*. Princeton, NJ: Princeton University Press.

Fischer, J. (n.d.). *Garden of Forking Paths*. Retrieved August 25, 2011, from http://gfp.typepad.com/the_garden_of_forking_pat/2004/06/do_you_know_wha.html

Fischer, J. (2006). The free will revolution continued. *The Journal of Ethics* , 315-345.

Fishbach, A. &. (2008). Together or Apart: When goals and temptations complement versus compete. *Journal of Personality and Social Psychology* , 547-559.

Fishbach, R. D. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *Journal of Consumer Research* , 370-377.

Forman, E. M. (2007). A comparison of acceptance and control-based strategies for coping with food cravings. *Behavior Research and Therapy*, Vol. 45 , 2372-2386.

Frankfurt, H. (1988). *The Importance of What We Care About* . New York, NY: Cambridge University Press.

Gailliot, M. T. (2007). Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. *Journal of Personality and Social Psychology* 92 (2) , 325.

Garcia-Portilla, M. P.-B. (2014). Long term outcomes of pharmacological treatments for opiod dependence: does methadone still lead the pack? *British Journal of Clinical Pharmacology* , 272-284.

Giligan, C. (1977). In a Different Voice: Women's conceptions of self and morality. *Harvard Educational Review* , 481-517.

Gold, J. I. (2006). Effectiveness of virtual reality for pediatric pain distraction during IV placement. *Cyber Psychology & Behavior* , 207-212.

Grandin, T. (1980). Livestock behavior as related to handling facility design. *International Journal of the Study of Animal Problems* , 33-52.

Greenwald, A. M. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology* 74 (6) , 1464.

Haney, C. B. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology* , 69-97.

Hanson, A. &. (2011). Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *Journal of Urban Economics* , 99-114.

Hare, R. M. (1963). *Freedom and Reason*. New York, NY: Oxford University Press.

Harlow, J. (1868). Recovery from the passage of a bar through the head. *Publications of the Massachusetts Medical Society* , 327-347.

Hershfield, H. E. (2011). Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research* 48 , 23-27.

Hinojosa, M. &. (2009). Challenging colorblind education: A descriptive analysis of teacher racial attitudes. *Research and Practice in Social Sciences* 4:2 , 27-45.

Holton, R. (2009). *Willing, Wanting, Waiting*. Oxford: Clarendon Press.

Hume, D. (2003). *A Treatise of Human Nature*. New York, NY: Dover.

Hume, D. (1748). *Enquiry Concerning Human Understanding*. New York, NY: Prentice Hall 1995.

Inwagen, P. V. (1983). *An Essay on Free Will*. New York, NY: Oxford University Press.

James, W. (1890). *The Principles of Psychology, Vol. 1*. New York: Henry Holt and Co.

James, W. (1884). *The Will to Believe*. New York, NY: Dover 1956.

K., L. C. (n.d.). Retrieved November 10, 2014, from <https://www.youtube.com/watch?v=F-JFPePh6yM>

Kane, R. (1996). *The Significance of Free Will*. New York, NY: Oxford University Press.

Keim, J. (1997). A Compatibalist theory of alternative possibilites. *Philosophical Studies* 8(3) , 319-330.

Kennette, J. &. (1997). Philosophy and Commonsense. In M. Smith, *Ethics and the A Priori* (p. 59). New York, NY: Cambrige University Press.

Kleiber, C. &. (1999). Effects of distraction on children's pain and distress during medical procedures:a meta analysis. *Nursing Research* , 44-49.

Korsgaard, C. (1986). Skepticism about practical reason. *The Journal of Philosophy* Vol. 83 , 50-62.

Laplace, P. S. (1814 (Translation 1902)). *A Philosophical Essay on Probabilities*. London: John Wiley & Sons.

Leary, A. a. (2007). Promoting Self-Compassionate Attitudes toward Eating Restrictive and Guilty Eaters. *Journal of Social and Clinical Psychology* .

Levy, N. (2011). Resisting Weakness of the Will. *Philosophy and Phenomenological Research* , 134-155.

Lewis, D. (1986). Counterfactual Dependence and Time's Arrow. In D. Lewis, *Philosophical Papers* (pp. 32-52). New York: Oxford University Press.

Mandel, N. a. (1999). Terror Management and Marketing: He who dies with the most toys wins. *Advances in Consumer Research* , 527-532.

Manning, R. L. (2007). The Kitty Genovese Murder and the social psychology of helping: the parable of the 38 witnesses. *American Psychologist* , 555.

Marks, I. (1987). *Fears, Phobias, and Rituals*. New York, NY: Oxford University Press.

Mauren, e. a. (2005). The Morning After: Limit Violations and the Self-Regulation of Alcohol Consumption. *Psychology of Addictive Behaviors* , 253-262.

McCaul, K. D. (1984). Distraction and coping with pain. *Psychological Bulletin* , 516.

McGonigal, K. (2012). *The Willpower Instinct* . Avery Publishing.

McKenna, M. (2005). Reasons reactivity and incompatibilist intuitions. *Philosophical Explorations* 8 (2) , 131-143.

Mele, A. (2012). *Backsliding*. London: Oxford University Press.

Mele, A. (1987). *Irrationality*. New York, NY: Oxford University Press.

Mele, A. (2000). Reactive Attitudes, Reactivity, and Omissions. *Philosophy and Phenomenological Research* , 447-452.

Milgram, S. (1974). *Obedience to Authority*. New York: Harper.

Mill, J. S. (1882). *A System of Logic, Ratiocinative and Inductive, Being A Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. New York, NY: Harper & Brothers, Publishers.

Miller, B. M. (2001). Moral Credentials and the Expression of Prejudice. *Journal of Personality and Social Psychology* , 33-43.

Miller, J. (2009). Minding the animals: Ethology and the Obsolescence of Left Humanism. *American Chronicle* .

Mischel, W. (1970). Attention in Delay of Gratification. *Journal of Personality and Social Psychology* .

Mischel, W. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology* , 204-218.

Moore, G. (1903). *Principia Ethica*. <http://fair-use.org/g-e-moore/principia-ethica>.

Muraven, e. a. (2005). One too many: Predicting future alcohol consumption following heavy drinking. *Experimental and Clinical Psychopharmacology* , 127-136.

Myerson, R. B. (1991). *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.

Nagel, T. (1979). *Moral Luck*. Retrieved March 21, 2015, from https://people.brandeis.edu/~teuber/THOMAS_NAGEL_Moral_Luck.pdf

Neff, K. (2003). Self-Compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity* , 85-101.

Nowlis, S. a. (2004). The effect of distraction while tasting a food sample. *Journal of Consumer Research* , 599-608.

Oaten, M. &. (2006). Improved self-control: The benefits of a regular program of academic study. . *Basic and Applied Social Psychology* , 1-16.

Oaten, M. &. (2007). Improvements in self-control from financial monitoring. *Journal of Economic Psychology* , 487-501.

Oaten, M. &. (2006). Longitudinal gains in self-control from regular physical exercise. *British Journal of Health Psychology* , 717-733.

Pereboom, D. (1995). Determinism al Dente. *Nous* , 21-45.

Ravizza, F. a. (198). *Responsibility and Control*. New York, NY: Oxford University Press.

Rawls, J. (1951). Outline of a decision procedure for ethics. *Philosophical Review* , 1-19.

Rodgers, T. &. (2013). Situationism versus Situationism. *Ethical Theory and Moral Practice* , 1-18.

Rorty, A. O. (1980). Akrasia and Conflict. *Inquiry* , 193-212.

Rosch E. and Mervis, C. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* , 573-605.

Rosch, E. (1975). Cognitive Representation of Semantic Categories. *Journal of Experimental Psychology* , 192-233.

Schelling, T. C. (1978). Economics, or the art of self-management. *The American Economic Review* , 290-294.

Schelling, T. C. (1980). *The Strategy of Conflict*. Cambridge : Harvard University Press.

Schelling, T. C. (1980). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Sedikides, C. &. (2007). *The Self*. New York, NY: Psychology Press.

Seligman, M. E. (2000). Positive Psychology: An Introduction. *American Psychological Association, Vol. 55, No. 1* , 5-25.

Sellars, W. (1962). Philosophy and the Scientific Image of Man. In W. Sellars, *Empiricism and the Philosophy of Mind* (pp. 1-40). London: Routledge.

Smith, K. (2011). Neuroscience vs philosophy: Taking aim at free will. *Nature vol. 477* .

Smith, M. (2004). *Ethics and the A Priori*. New York, NY: Cambridge University Press.

Smith, M. (1995). Internal Reasons. *Philosophy and Phenomenological Research* , 55-68.

Smith, M. (1994). *The Moral Problem*. New York, NY: Blackwell Publishers.

Sterelny, K. (2010). "Minds: extended or scaffolded?" . *Phenomenology and the Cognitive Sciences* , 465-481.

Strawson, P. F. (1982). Freedom and Resentment. In G. Watson, *Free Will* (pp. 59-80). New York, NY: Oxford University Press.

Strawson, P. F. (n.d.). *Information Philosopher*. Retrieved August 26, 2011, from <http://www.informationphilosopher.com/solutions/philosophers/strawson/>

Tice, D. M. (2001). Emotional distress regulation takes precedence over impulse control: If you feel bad, do it! *Journal of Personality and Social Psychology* , 53-67.

Tierney, B. a. (2011). *Willpower*. New York: Penguin Press.

Tzu, L. (1995). *Tao Te Ching*. New York, NY: Griffen.

Uhlmann, E. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. . *Organizational Behavior and Human Decision Processes*, 104 (2) , 207-223.

Vaillant, G. (1995). *The Natural History of Alcoholism Revisited*. Cambridge, MA: Harvard University Press.

Verhoeven, e. a. (2012). Pain catastrophizing influences the use and the effectiveness of distraction in school children. *Eropean Journal of Pain*, Vol. 16 , 256-267.

Vihvelin, K. (2004). Freewill Demystified: A Dispositional Account. *Philosophical Topics* 32 , 427-450.

Vohs, K. D. (2008). Making choices impairs subsequent self-control: A limited resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology*, 94, , 883-898.

Wallace, R. J. (2006). Moral responsibility and the practical point of view. In R. J. Wallace, *Normativity and the Will*. New York, NY: Oxford University Press.

Wallace, R. J. (2001). Normativity, Commitment, Instrumental Reason. *Philosophers Imprint* Vol. 1, No. 3 , 10-36.

Waller, R. (2014). Revising Reasons-Reactivity: Weakly and Strongly Sufficient Reasons for Acting. *Ethical Theory and Moral Practice* 17 (3) , 529-543.

Watson, G. (2004). Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. In G. Watson, *Agency and Answerability* (pp. 222-223). New York, NY: Oxford University Press.

Watson, G. (2004). Disordered Appetites. In G. Watson, *Agency and Answerability* (pp. 59-87). New York: Oxford University Press.

Watson, G. (2004). Free Action and Free Will. In G. Watson, *Agency and Answerability* . New York, NY: Oxford University Press.

Watson, G. (1982). Free Agency. In G. Watson, *Agency and Answerability* (p. 19). New York, NY: Oxford University Press.

Watson, G. (2004). Skepticism about weakness of will. In G. Watson, *Agency and Answerability*. New York, NY: Oxford University Press.

Watson, G. (2004). The Work of the Will. In G. Watson, *Agency and Answerability*. New York, NY: Oxford University Press.

Weinberg, R. &. (2007). Introduction to psychological skills training. In W. a. Eds., *Foundations of Sport and Exercise Psychology*. Champaign, IL: Human Kinetics

Wilcox, e. a. (2009). Vicarious goal fulfillment: When the mere presence of a healthy option leads to an ironically indulgent decision. *Journal of Consumer Research* , 380-393.

Williams, B. (1981). Internal and External Reasons. In B. Willaims, *Moral Luck* (pp. 101-13). New York, NY: Cambridge University Press.

Witkiewitz, B. a. (2010). Depression, Craving, Substance Use Following a Randomized Trial of Mindfulness-Based Relapse Prevention. *Journal of Consulting and Clinical Psychology, Vol. 78* , 362-374.

Wohl, M. J. (2010). I forgive myself, now I can study: How self-forgiveness for procrastinating can reduce future procrastination. *Personality and Individual Differences* , 803-808.

Xu, K. N. (2015, February 25). *Race IAT 2002-2013*.

Yoo SS, G. N. (2007). The human emotional brain without sleep: A prefrontal-amygdala disconnect. *Current Biology* , 877-878.