

8-1-2013

# Willpower and Ego-Depletion: How I Do What I Don't Want to Do, and Why It's Not (Completely) My Fault When I Don't

Samuel C. Sims  
*Georgia State University*

Follow this and additional works at: [http://scholarworks.gsu.edu/philosophy\\_theses](http://scholarworks.gsu.edu/philosophy_theses)

---

## Recommended Citation

Sims, Samuel C., "Willpower and Ego-Depletion: How I Do What I Don't Want to Do, and Why It's Not (Completely) My Fault When I Don't." Thesis, Georgia State University, 2013.  
[http://scholarworks.gsu.edu/philosophy\\_theses/140](http://scholarworks.gsu.edu/philosophy_theses/140)

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

Willpower and Ego-Depletion:

How I Do What I Don't Want to Do, and Why It's Not (Completely) My Fault When I Don't

by

Sam Sims

Under the Direction of Eddy Nahmias

ABSTRACT:

Experimental studies on willpower confirm the Strength Model of Self-Control, which claims that willpower depends on limited physiological resources. Exercising willpower depletes these resources, which impairs further exercises of willpower. This phenomenon is called “ego-depletion.” As a result, depleting these resources impairs further exercises of executive control. My thesis argues that this phenomenon has two important philosophical consequences: First, ego-depletion provides evidence against the Humean approach to motivation, according to which people always act according to their strongest desires. Second, people suffering from ego-depletion are not fully responsible for failures of self-control.

INDEX WORDS: Ego-Depletion, Willpower, Self-Control, Executive Control, Motivation, Desire, Humean Theory of Motivation, Responsibility;

Willpower and Ego-Depletion:  
How I Do What I Don't Want to Do, and Why It's Not (Completely) My Fault When I Don't

by

Sam Sims

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of  
Master of Arts in Philosophy  
in the College of Arts and Sciences  
Georgia State University  
2012





Copyright by  
Samuel Casey Sims  
2012

Willpower and Ego-Depletion:

How I Do What I Don't Want to Do, and Why It's Not (Completely) My Fault When I Don't

by

Sam Sims

Committee Chair: Eddy Nahmias

Committee: Neil Van Leeuwen

Andrea Scarantino

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2012



## ACKNOWLEDGEMENTS

I thank all of the members of my committee, Eddy Nahmias, Neil Van Leeuwen, and Andrea Scarantino for reading my work and providing lots of helpful feedback and important challenges, for their letters of recommendation when I applied to PhD programs, and for being excellent philosophical role models. I thank Eddy for being supportive a thesis director, and, since my thesis is partly about responsibility, for being just hands off enough to let me be responsible for my own work and career choices while still pushing me to become more responsible as a graduate student. I thank Neil for giving me the idea to write the paper that eventually became my writing sample and the second chapter of my thesis, and for other very good ideas that went into this paper and other parts of the thesis. I thank Andrea for sharp critical feedback early on in grad school that helped me improve the way I write and do philosophy, and for being a great teaching role model when I was preparing to teach intro to philosophy.

I am also grateful to Gwen Frishkoff for helpful feedback on the paper that eventually became the last chapter of my thesis, for helping me develop my interests in empirical psychology and cognitive neuroscience, for letting me have the experience of working in her lab, and for lots of encouragement.

I would like to thank Sandy Dwyer, Tim O'Keefe, and George Rainbolt for their help and support with various teaching and administrative issues, and for teaching me how to deal with those issues. In particular, I am grateful to Sandy for being so generous with her time and energy, and looking out for us grad students. We are very lucky to have you in the department.

I am grateful to my parents, Beth and Steve Sims, for their continued support and help as I am still growing up, and my aunt and uncle, Sandy and Pat Dickson, for acting as catalysts for my personal and intellectual growth, especially during my teen years. I would not be where I am or who I am now if not for all of them.

I thank Cami Koepke for stimulating philosophical conversations, for good advice, and for being a wonderful friend as well as someone I can look up to. I miss you.

Finally, I thank Toni Adleberg for all of the wonderful times we've had together at Georgia State and for the time we spent traveling together before going to our PhD programs. I miss being near you already.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS .....</b>	<b>v</b>
<b>1 THE LIMITATIONS OF WILLPOWER.....</b>	<b>1</b>
<b>1.1 What are Self-Control and Willpower? .....</b>	<b>2</b>
<b>1.2 Evidence for the Strength Model .....</b>	<b>6</b>
<b>1.3 Other Possible Explanations .....</b>	<b>15</b>
<b>2 WILLPOWER AND MOTIVATION:</b>	
<b>HOW WE DO WHAT WE DON'T WANT TO DO.....</b>	<b>27</b>
<b>2.1 The Humean Approach .....</b>	<b>28</b>
<b>2.2 Is Fatigue Generality Consistent with the Strength Principle? .....</b>	<b>34</b>
<b>2.3 Conclusion.....</b>	<b>43</b>
<b>3 SELF-CONTROL FAILURE:</b>	
<b>WHY IT'S NOT (COMPLETELY) MY FAULT .....</b>	<b>45</b>
<b>4 PHILOSOPHICAL CONCLUSIONS AND FURTHER PROJECTS .....</b>	<b>56</b>

## 1 EGO-DEPLETION AND THE LIMITATIONS OF WILLPOWER

Willpower is something you exercise to stop yourself from doing something you very much want to do, or to make yourself do something you very much want to avoid doing. As such, willpower raises some philosophical questions. For instance, can you try not to do something you want to do more than anything else? Are you morally responsible for doing something you want to do, even if you tried very hard not to do it? Traditionally, philosophers attempt to answer such questions with philosophical methods like conceptual analysis, phenomenology, and philosophical thought-experiments. However, a growing number of philosophers also rely on empirical data to look for answers to philosophical questions like these. In this thesis, I rely primarily on this last method in order to address these two philosophical questions about willpower.

The following chapters discuss the philosophical implications of some experimental studies on willpower, and what these results suggest about the two philosophical questions just mentioned. According to the Strength Model of Self-Control, willpower is like a muscle in that using willpower requires certain physiological resources, which means that exercising willpower produces a kind of fatigue that temporarily weakens one's willpower, a phenomenon called "ego-depletion" (Baumeister et al. 1998, Baumeister, Vohs, and Tice 2007).

Since the Strength Model implies that willpower has certain limitations, I will argue that the Strength Model can help answer the two philosophical questions above. In Chapter 2, I argue against the Humean approach to motivation, according to which we only try to do what we want to do more than anything else. I argue that this approach predicts that ego-depletion should not affect certain self-control tasks. However, there are studies which suggest that ego-depletion

does affect performance on these vary tasks. Therefore, these studies provide evidence against the Humean approach.

In Chapter 3, I argue that people suffering from ego-depletion are not fully responsible for failures of self-control. Ego-depletion makes self-control more difficult, and greater difficulty makes one less responsible. Therefore, ego-depletion makes a person less responsible for a failure of self-control than she would be if she were not ego-depleted.

Since neither of these conclusions follow if the Strength Model is false, I begin in this chapter by arguing that the Strength Model is correct about the way depleting certain physiological resources impairs willpower. First, I discuss evidence for the Strength Model, and then I argue against alternative explanations of this evidence. Since the Strength Model is a better explanation for this evidence than any of these alternatives, the Strength Model is probably correct. Having provided support for the Strength Model, I rely on this model to challenge the Humean approach to motivation in Chapter 2 and develop an important consideration about moral responsibility in Chapter 3.

### **1.1 What are Self-Control and Willpower?**

Self-control is often very difficult. Consider the difficulty of not eating something tempting but unhealthy, not laughing at something funny but inappropriate, exercising or continuing to work when tired, or doing something unpleasant, uncomfortable, or even physically painful. Sometimes we successfully resist temptation, stifle laughter, exercise, continue working, or overcome pain with a kind of mental effort that we call “willpower.” Sometimes, however, we have failures of self-control in spite of trying to exercise willpower, which raises the following question: does willpower have limitations? In this chapter, I argue

that willpower is limited, but first I need to define what I mean by “self-control” and “willpower.”

Since self-control is something we exercise to prevent ourselves from acting out of weakness of will, one possible definition of “self-control” is just not acting out of weakness of will. However, there are two problems with this definition. First, this definition is not helpful without also defining “weakness of will,” and philosophers do not agree about what weakness of will is. Since I want the arguments in this thesis to convince people with different conceptions of weakness of will, I should offer a definition of “self-control” that is independent of any particular conception of weakness of will.

Second, there is more to self-control than simply not acting out of weakness of will, which is evident from cases in which a person does not act out of weakness of will without exercising self-control. Suppose I order a meal that comes with either a Caesar or house salad, and I want to house salad more than the Caesar salad, so I simply take the house salad without having to do anything else to prevent myself from taking the Caesar salad. Even if this case is not an example of any particular conception of weakness of will, it does not seem like a case of self-control.

Self-control would involve doing something more than just taking the house rather than the Caesar salad. If I am trying to follow a vegan diet, for instance, I might prevent myself from taking the Caesar salad by trying not to think about it or trying to control how much I want it. This case seems like an example of self-control because, in order to take the house salad, I do something to prevent myself from taking the Caesar salad.

Since self-control involves doing something to prevent yourself from doing something else, perhaps self-control is just what Al Mele calls “resisting” a desire (1990, p. 459). According to Mele, resisting a desire means attempting to “manipulate [one’s] motivational

condition, [one's] environment, or both in such a way as to bring it about that [one] does not act on the desire being resisted." However, there are two reasons for why self-control is not limited to resisting desires.

First, rather than resisting a desire, self-control may involve preventing, suppressing, producing or sustaining mental states and processes other than desires, like attention, thoughts, beliefs, and emotions. Focusing attention in spite of distracting stimuli, suppressing thoughts, avoiding and dismissing beliefs that are intuitive but unjustified, and controlling emotions like fear, anxiety, and anger all require self-control.

Second, resisting a desire involves preventing an action based on that desire, but self-control may involve preventing a behavior that is not intentional and therefore not an action. Such behaviors include crying, laughing, coughing, sneezing, and blinking. Preventing yourself from crying in public when you are upset, laughing at something funny but inappropriate, coughing during a talk when you have a tickle in your throat, sneezing when you do not have a tissue, or blinking in response to a camera flash all require self-control.

Self-control, then, involves doing something that affects attention, thoughts, beliefs, desires, emotions, actions or behavior. However, nearly all of what I do affects my mental states and behavior in some way or another. For instance, taking the house salad causes me to believe that I have taken the house salad, but this effect on my beliefs does not make taking the house salad an act of self-control. Rather, doing something is an instance of self-control only if it affects my attention, thoughts, beliefs, desires, emotions, actions or behavior in a way that allows me to accomplish something I am trying to do. Here is my definition of self-control:

**Self-Control**

A person S exercises self-control by doing  $\phi$  if and only if S is trying to do  $\psi$ , and doing  $\phi$  affects S's attention, thoughts, beliefs, desires, emotions, actions or behavior in a way that allows S to do  $\psi$ .

This definition allows for three kinds of self-control. First, you can exercise self-control by arranging your environment in certain ways, by putting yourself or yourself in certain situations and by avoiding certain other situations. Examples of this kind of self-control include disabling your internet connection to keep yourself from getting distracted while you work, placing your alarm clock on the other side of the room to help yourself get up in the morning, and avoiding situations that lead to temptation.

Second, you can perform certain mental exercises that affect your mental state or behavior in a way that allows to do something or avoid doing something. Examples of these exercises include imagining something tempting in an unattractive way in order to make yourself want it less (Mele, 1990, p. 459), counting to ten in your head when you are angry in order to calm down, or thinking about why you care about your goals when you are trying to get yourself to do something unpleasant.

Third, sometimes you can exercise self-control even though you do not know of any way to change your environment or of any mental exercises that would affect your mental state or behavior in the necessary way. In these situations, sometimes you can exercise self-control just by trying to alter or sustain your attention, thoughts, beliefs, desires, emotions, actions or behavior directly with just effort alone. Examples of this kind of self-control include trying to attend to the person speaking to you rather other conversations nearby, trying not to think while attempting to meditate or fall asleep, trying not to believe someone is cheating on you when you feel unjustifiably jealous, trying not to let yourself want something you believe is wrong even to want, trying not to let yourself feel angry, trying not to eat something that would break your diet, and trying not to cough during a talk. People often exercise self-control without changing their

environment or performing mental exercises, and this capacity for altering and sustaining one's mental states and behavior is what we commonly call "willpower:"

**Willpower**

A mental capacity for altering or sustaining attention, thoughts, beliefs, desires, emotions, actions and behaviors in a way that allows one to accomplish something one is trying to do.

Willpower is a "mental capacity," because it is a capacity that one can exercise with one's mind alone, unlike the capacity for lifting a heavy objects or the capacity for running long distances. Other mental capacities include visual, auditory, and tactile perception, language acquisition, comprehension, and production, emotional responses, working memory, thinking, reasoning, planning, and decision making.

Since willpower is a mental capacity, a natural question is whether willpower has limitations like other capacities. For instance, my capacity for lifting heavy objects is limited by how strong my muscles are, my capacity for running long distances is limited by my cardiovascular endurance, my capacity for working memory is limited by how many chunks of information it can hold, and my capacity for reasoning is limited by the complexity of the problems it can solve. In the next two sections, I argue that willpower has similar limitations.

**1.2 Evidence for the Strength Model**

In the 1990s, scientists began conducting experiments testing the Strength Model of Self-Control, which claims that the mental capacity of willpower has limitations that are similar to those of a muscle:

**The Strength Model**

Willpower depends on limited, physiological resources in the following way: (a) exercising willpower depletes these resources, and (b) depleting these resources impairs the capacity for willpower, making self-control more difficult.

In other words, exercising willpower produces a kind of fatigue that temporarily weakens one's willpower, just like using a muscle (Baumeister, Heatherton, and Tice 1994, Muraven, Tice, and Baumeister 1998, and Baumeister et al. 1998, Baumeister, Vohs, and Tice 2007).. This kind of fatigue is called "ego-depletion:"

### **Ego-depletion**

A state of fatigue in which the physiological resources necessary for willpower are depleted.

The Strength Model makes several testable predictions about willpower, and scientists have conducted lots of experiments that test these predictions. This section focuses on what I think are the three most important predictions of the Strength Model: Fatigue Generality, Reduced Prefrontal Activity, and the Physical Substance Prediction.

The first prediction is about performance on different tasks:

### **Fatigue Generality (of Willpower)**

Performing one willpower task impairs performance on another willpower task, even if these tasks have nothing in common other than requiring willpower.

The Strength Model predicts this phenomenon because performing the first task should drain some limited resource needed for the second task. If every self-control task requires the same mental capacity of willpower, and exercising willpower always draws from the same "reservoir" of physiological resources, then the first task should impair performance on the second task even if these tasks have nothing in common other than requiring willpower.

Lots of studies have tested this prediction and confirmed it. For instance, Baumeister et al. (1998) conducted a study in which they brought participants into a room filled with the smell of baking chocolate and presented them with two food options: delicious looking and smelling chocolate chip cookies on one side of a table, and a bowl of radishes on the other. Experimenters told participants that this task was part of a study on taste perception. The experimenters

instructed some participants to eat radishes but not cookies, and other participants to eat cookies but not radishes. Then experimenters would leave the participant alone with the two food options to watch the participant from behind a one-way mirror. A third group of participants, who would serve as a control group, skipped this task altogether, never encountering the table with the two food options.

Assuming that eating radishes rather than cookies requires willpower while eating cookies does not, this task should require more effort for the people who ate radishes than for people in the other two groups. In fact, eating radishes rather than cookies was apparently quite difficult. While all participants followed the instructions, participants in the radish condition would often look longingly at the cookies, some even going so far as to smell one of the cookies before setting it down and having some radishes. On a later questionnaire, participants who ate radishes rated this task as significantly more difficult than participants who ate cookies.

After the first task, experimenters presented participants in all three conditions with a difficult geometrical puzzle, which, unknown to participants, was actually unsolvable. To measure self-control, experimenters recorded how long and how many times participants would attempt to solve this puzzle before giving up. Perseverance on a frustrating task should require willpower, so amount of time and number of attempts should reflect the “strength” of people’s willpower.

On average, the participants who had eaten radishes spent less than half the amount of time and made less than two-thirds as many attempts at solving the puzzle as participants in the other two groups, a statistically significant difference. There was no significant difference between participants who had eaten cookies and those who skipped the first task, which makes sense because neither of these groups should have exercised willpower before the puzzle task.

These results suggest that exercising willpower on the first task impairs willpower on the second task, making perseverance on a frustrating task more difficult. If willpower depends on some limited physiological resources, then depleting these resources on the taste-testing task would explain why the participants who ate radishes would quit more easily on the puzzle task.

Someone could argue that the participants who ate radishes gave up more quickly not because of impaired willpower, but because they resented the experimenter for not letting them have cookies, and therefore did not want to try as hard on the puzzle task. However, this explanation does not cover all cases of Fatigue Generality. Some self-control tasks would not lead participants to resent the experimenter because these tasks do not involve denying participants something that they want, like cookies. Even though some self-control tasks would not cause resentment, these tasks still cause participants to perform worse on another self-control task.

For instance, Muraven, Tice, and Baumeister (1998) asked some participants to suppress their emotional responses to a disturbing video, and asked others to watch the same video without trying to suppress their emotional responses. Assuming that suppressing emotional responses should require more effort than watching the video without suppressing these emotional responses, participants in the first group should deplete more of the resources upon which willpower depends than the second group. On the other hand, there is no reason to think that suppressing emotional responses should cause participants to resent the experimenters in the same way as not having cookies, because suppressing emotional responses does not involve denying participants something that they want.

After the video, experimenters measured self-control by asking participants to squeeze a handgrip exerciser continuously without releasing. Continuing to squeeze should require

willpower because of the increasingly uncomfortable feeling of muscular fatigue that results from squeezing without releasing. Sure enough, people who suppressed their emotions on the video released the handgrip exerciser significantly sooner than people who did not suppress their emotions, again confirming the Strength Model's prediction of Fatigue Generality.

There are now an abundance of studies which confirm the phenomenon of Fatigue Generality, but I should mention one caveat. Many studies also suggest that the effects of ego-depletion are not confined to self-control and willpower. According to these studies, willpower is not the only way to cause ego-depletion or the only mental capacity impaired by ego-depletion.

Willpower belongs to a family of mental capacities involving what scientists call "executive control." These capacities depend on different parts of prefrontal cortex and share the following characteristics: First, people can exercise these capacities consciously, so are aware of and can report what they are doing when they exercise these capacities. Second, people can exercise these capacities voluntarily, which implies, for instance, that they can exercise these capacities in response to a verbal request or verbal incentive. Third, exercising these capacities can be effortful, so exercising these capacities can involve a conscious feeling of effort as well as physiological indicators of effort such as increased skin conductance responses (Elliot 2003, Fuster 2008, Norman and Shallice 1980, Jacoby 1991, Naccache *et al.* 2005). By contrast, non-executive mental capacities, like syntactic processing, implicit learning, and emotional responses, are automatic, involuntary, effortless, and often occur without conscious awareness.

In addition to willpower, mental capacities involving executive control include conscious reasoning, planning, and decision making (Elliot 2003). Studies suggest that these other capacities are affected by ego-depletion. For instance, Schmeichel *et al.* (2003) found that

exercising willpower on the emotion suppression task impairs performance on a test with complex problems that require conscious reasoning, but not on a general knowledge and vocabulary test that requires little conscious reasoning. They asked participants to watch an excerpt from a documentary with upsetting footage of animal suffering, and told some participants to suppress emotional responses to the video and other participants to react naturally to the video.

After the video, participants from both groups took part of the GMAT test, which includes simple vocabulary and general knowledge questions that require remembering words and facts, but not very much conscious reasoning. After the GMAT, participants took the CET test, which consists of 20 difficult questions with unclear answers that require lots of conscious reasoning to answer accurately. Participants who suppressed their emotional responses while watching the video performed significantly worse on the CET, but not the GMAT. These results suggest that suppressing emotional responses impairs conscious reasoning, but not remembering words and facts, which does not involve much conscious reasoning.

Studies also suggest that exercising these other capacities involving executive control can also cause ego-depletion. For instance, Vohs, et al. (2008) found that making conscious choices on one task impairs willpower on a task that requires resisting physical pain. They asked one group of participants to make a long series of choices between different products, which involves executive control, and they asked another group of participants, as a control group, to write down thoughts, feelings, and opinions about a series of advertisements.

To measure self-control, experimenters used the cold pressor task of pain tolerance. They asked participants to submerge the non-dominant arm up to the elbow in a tank full of painfully cold, almost freezing water, and hold it there as long as possible. On average,

participants who made choices between products quit after a little less than half a minute, whereas other participants quit after just over a minute, a statistically significant difference. Assuming that holding one's arm under painfully cold water requires willpower, these results confirm that making a long series of conscious choices causes ego-depletion and impairs willpower.

Since studies suggest that Fatigue Generality applies to other capacities involving executive control as well as willpower, here is a more accurate description of this phenomenon:

**Fatigue Generality (of Executive Control)**

Performing one executive control task impairs performance on a second executive control task, even if these tasks have nothing in common other than executive control.

Allowing that the Strength Model applies, not just to willpower, but to all mental capacities involving executive control, Fatigue Generality is still a prediction of the Strength Model, and the experimental results discussed so far confirm this prediction.

The second prediction of the Strength Model is the neural correlates of Fatigue Generality, or what happens in the brain when the first task impairs performance on the second task:

**Reduced Prefrontal Activity**

Performing one executive control task should cause reduced activity in the relevant part of the prefrontal cortex during a second executive control task.

The Strength Model predicts this phenomenon because performing the first task should impair certain mental capacities on the second task, and these capacities depend on different parts of the prefrontal cortex. Since exercising these capacities involves activity in these different parts of the prefrontal cortex, impairments in these capacities should accompany reduced activity in these areas. Measuring prefrontal activity is a non-behavioral way of measuring performance on the second task.

A couple of studies confirm this prediction. Hedgcock, Vohs and Rao (2012) used fMRI to look at activity in the middle frontal gyrus, part of the prefrontal cortex, while participants performed the consumer choices task. Before the consumer choices task, participants performed the attention control task, in which participants watched a silent video of a woman talking while words appear for brief periods of time in the corner of the screen. Some participants watched the video while trying not to look at the words, while other participants watched without making any such effort. Trying not to look at the words should require willpower, while watching the video without making any such effort should not. Trying not to look at the words during the attention control task resulted in slower reaction times during the consumer choices task, which confirms the prediction of Fatigue Generality, as well as reduced activity in the middle frontal gyrus, which confirms the prediction of Reduced Prefrontal Activity.

Inzlicht and Gutzell (2007) used EEG to look at electrical brain activity in the anterior cingulate cortex (ACC), another part of the prefrontal cortex, while participants performed the Stroop task, which requires willpower because people must stop themselves from reading color words in order to respond to the color in which the word printed. First, participants watched an upsetting video. Participants who suppressed their emotional responses during the video task performed worse on the Stroop task, which confirms the prediction of Fatigue Generality, and ACC activity was weaker during the Stroop task for these participants, which confirms the prediction of Reduced Prefrontal Activity.

The third prediction of the Strength Model is about the physical substance upon which executive control depends:

**The Physical Substance Prediction**

There is some physical substance X that decreases in concentration when performing an executive control task, and restoring this substance before a second executive control task prevents impaired performance on this second task.

The Strength Model predicts the existence of substance X because this substance would be the physiological resource upon which executive control depends. If depletion of this substance accounts for the effect of one executive control task on another, then this substance must decrease after the first task, and restoring this substance should eliminate the effect of the first task on the second.

Some studies suggest that this substance is blood glucose. Gailliot et al. (2007) found that the attention control task (avoiding looking at words appear during a movie) caused a significant reduction in the concentration of blood glucose. Experimenters measured blood glucose before and after the task, and found that blood glucose declined significantly for participants who tried not to attend to the words, while there was no significant decline for participants who watched the video without making any effort to avoid attending to the words. These results suggest that exercising willpower on the attention control task caused a decrease in blood glucose.

Gailliot et al. (2007) also found that providing participants with glucose can eliminate the effects of ego-depletion. After performing the attention control task, participants received either a glass of lemonade sweetened with glucose or with a non-glucose artificial sweetener. Then all participants performed the Stroop Task.

For participants who drank the non-glucose lemonade, performance was typical of other studies on willpower: those who exercised willpower on the attention control task performed worse on the Stroop task than those who did not exercise willpower. For participants who drank lemonade sweetened with glucose, however, the attention control task had no effect on the Stroop task. Drinking glucose seems to eliminate the effect of ego-depletion, which suggests

that ego-depletion involves a lack of blood glucose, confirming the Physical Substance Prediction.

The experimental results discussed so far support the Strength Model because they confirm its three most important predictions: Fatigue Generality, Reduced Prefrontal Activity, and the Physical Substance Prediction. Unless there is some better explanation for these experimental results, they provide good evidence for the Strength Model. The next section considers some alternative explanations of these experimental results, but argues against these alternatives explanations.

### **1.3 Other Possible Explanations**

The studies discussed so far evince the phenomenon of Fatigue Generality: performing one executive control task results in worse performance on another executive control task. The Strength Model is committed to a particular explanation of this phenomenon: exercising executive control produces a kind of fatigue by depleting certain physiological resources, and this state of fatigue impairs mental capacities involving executive control. Since certain tasks require these capacities, impairing these capacities makes these tasks more difficult, which is why people in this state of fatigue perform worse on these tasks.

Notice that a crucial part of the Strength Model's explanation is that impaired mental capacities make certain tasks more difficult. This section argues against some alternative explanations of Fatigue Generality. To distinguish the explanation just given from these alternative explanations, I call it the "Difficulty Explanation:"

#### **The Difficulty Explanation**

People suffering from ego-depletion have impaired mental capacities that make executive control tasks more difficult.

This explanation of fatigue generality is plausible because similar explanations apply to other cases of poor task performance. For example, sleep deprivation makes people worse at certain tasks like mental arithmetic, and the explanation for this effect is that sleep deprivation impairs certain mental capacities, which makes certain tasks like mental arithmetic more difficult. Similarly, the Difficulty Explanation claims that people suffering from ego-depletion quit more quickly on the cold pressor task, for instance, because ego-depletion impairs willpower, which makes it more difficult to endure the painfully cold water.

While the Strength Model is committed to the Difficulty Explanation, there are other explanations of Fatigue Generality that do not involve impaired mental capacities or greater difficulty of a task. I discuss three of these alternative explanations.

The first of these three explanations is that willpower and other capacities involving executive control are not actually limited in the way the Strength Model claims, but rather people believe that willpower and these other capacities are limited in this way. This belief leads them to perform worse on their second task either because they believe they are not able to perform at the same level, or because they believe that they need to conserve their willpower. The belief in limited willpower works as a self-fulfilling prophesy. People believe they must perform worse, and then they do.

### **The Belief Explanation**

People believe that willpower and other capacities are limited and that these capacities are impaired from their first task, causing them not to try as hard on other tasks.

This explanation is distinct from the Difficulty Explanation because people might believe in limited willpower even though willpower is not actually limited, people might not believe in limited willpower even though their willpower is actually limited. These two possibilities provide ways of testing the Belief Explanation against the Difficulty Explanation.

Job, Dweck, and Walton (2010) conducted such a test by using questionnaires designed to influence one group of participants to believe in limited willpower and another group of participants to believe in unlimited willpower. Then participants performed one of two versions of a task, designed to require willpower or not. Finally, all participants performed another task that requires willpower, and experimenters measured their performance. For participants primed to believe in limited willpower, exercising willpower on the first task resulted in significantly worse performance on the second task, just like in most of the studies discussed thus far. For participants primed to believe in unlimited willpower, exercising willpower on the first task had a smaller but opposite effect. The participants who exercised willpower on the first task actually performed slightly better on the second task. These results seem more consistent with the Belief Explanation than the Difficulty Explanation.

However, perhaps the belief in limited willpower is actually correct, and the belief in unlimited willpower only reverses the effect of ego-depletion when ego-depletion is mild. In this case, severe enough ego-depletion might still impair performance even for people who believe in unlimited willpower.

Vohs, Baumeister, and Schmeichel (2012) produced a conceptual replication of the results of the previous study, and tested the hypothesis that the belief in unlimited willpower would not reverse the effects of ego-depletion when ego-depletion is severe enough. They measured executive control using a delayed gratification task and a logical reasoning task, preceded by either four, one, or no tasks designed to require executive control.

For participants who believed in limited willpower, performing one or four executive control tasks prior to the measurement phase of the experiment resulted in worse performance on the delayed gratification and logical reasoning tasks, as in most of the studies described thus far.

For participants who believed in unlimited willpower, performing one executive control task prior to the measurement phase of the experiment did not result in worse performance in delayed gratification and logical reasoning, a conceptual replication of the previous study.

However, for participants who believed in unlimited willpower, performing four executive control tasks did result in significantly worse performance on delayed gratification and logical reasoning, suggesting that this belief does not reverse the effect of ego-depletion when ego-depletion is severe enough. These results suggest that the belief in limited willpower is actually correct, and the belief in unlimited willpower only reverses the effect of ego-depletion when ego-depletion is mild, which is more consistent with the Difficulty Explanation than the Belief Explanation.

The second alternative explanation of Fatigue Generality is that executive control tasks have a negative effect on people's mood, and this negative effect on mood causes people to perform worse on another executive control task. For instance, previously I considered the possibility that participants perform worse on the frustrating puzzle because they resent the experimenter for not allowing them to have any cookies. This resentment involves a negative mood.

I dismissed this possibility because, in many cases, the first executive control task should not lead participants to resent the experimenter. Often, this first task does not involve denying participants something they want.

However, suppose the first executive control task affects performance on the second task, not because of resentment for not getting something, but simply because participants dislike the first task or find it unpleasant. In this case, the first executive control task would have a negative

effect on mood, which might result in worse performance on another executive control task.

Since this explanation is an alternative to the Difficulty Explanation, it also deserves a name:

**The Mood Explanation**

Performing one executive control task has a negative effect on mood, which causes people not to try as hard on other tasks.

This explanation is distinct from the Difficulty Explanation because people's mood could get worse without impairing any of their mental capacities, and their mental capacities could become impaired without any change in their mood. These two possibilities provide ways of testing the Mood Explanation against the Difficulty Explanation.

Researchers have tested this explanation in a couple of ways. In many studies, including the study with the radishes and cookies, experimenters check for changes in mood using a questionnaire, such as the Positive Affect Negative Affect Scale (PANAS) or the Brief Mood Introspection Scale (BMIS), designed to detect differences in mood between different groups of participants, (Baumeister et al. 1998). For instance, if eating radishes rather than cookies has a negative effect on mood, then people who ate radishes, on average, should score differently on these questionnaires than people who ate cookies. In most of these studies, including the study with the radishes and cookies, people who exercised executive control on their first task did not score significantly differently from people who did not have to exercise executive control, suggesting that exercising executive control does not have such a negative effect on mood. If there is no negative effect on mood, then mood does not explain why people who exercise executive control on the first task perform worse on the second task.

There are also more direct ways of testing the mood explanation: two of the studies I mentioned above involve a video task that elicits negative emotions by requiring participants to watch an upsetting video. A similar task elicits positive emotions by requiring participants to

watch a funny video. Comparing the effects of these two tasks provides experimenters with a more direct way of testing for an effect of mood on executive control. In one study, for instance, experimenters asked some participants to watch a funny video, and other participants to watch a distressing video. For participants watching both videos, experimenters asked some participants to suppress emotional responses such as laughter and distress, while other participants received no such instructions. If negative mood explains why depleted participants perform worse on the second task, then participants who watched the funny video should perform better than those who watch the distressing video. However, there was no significant relationship between the content of the video and performance on the anagram task. Since the content of the video had no significant effect on task performance, these results suggest that negative mood does not explain why depleted participants perform worse. These results are not consistent with the Mood Explanation.

The Difficulty Explanation, on the other hand, predicts that participants who suppress their emotional responses, positive or negative, will suffer from ego-depletion and perform worse on the second task. Just as the Difficulty Explanation predicts, people who suppressed both positive and negative emotional responses performed significantly worse on an anagram task than people who did not try to suppress their positive or negative emotional responses. These results confirm the Difficulty Explanation.

Since the Mood Explanation relies on the claim that negative mood reduces motivation to perform well on the second task, the Mood Explanation is really a special version of a more general kind of explanation:

**The Motivation Only Explanation**

Exercising executive control on one task reduces motivation to perform well on a second executive control task, which causes people not to try as hard on the second task.

Importantly, the Difficulty and Motivation Only Explanations are quite distinct. Suppose Charlie, a surly teenager, is getting C's and D's in high school pre-calculus. Dad thinks that Charlie is performing poorly because math is difficult for Charlie. Charlie is just not as good at math as other kids. Mom thinks Charlie is performing poorly because Charlie does not care about math and does not want to do better.

Perhaps Mom and Dad are both partially correct about why Charlie is not getting better grades, but they are not offering the same explanation for his performance. If Mom is correct, then Charlie's performance might improve if they offer him incentives to perform better. If Dad is correct and Mom is completely wrong, then incentives should have no such effect.

If some people perform worse on their second task simply because of a lack of motivation, then a sufficient increase in their motivation should cause them to perform better. If so, then offering participants an incentive to perform better might compensate for the effects of ego-depletion. Muraven and Slessareva (2003) conducted three studies in which they found that manipulating motivation does influence performance on the second task, in a way that can compensate for the effects of ego-depletion. They manipulated motivation in different ways for the three studies: by telling participants that good performance on the second task would provide valuable information for developing treatments that would benefit people with Alzheimer's (the first study), by telling participants that practicing on a task would or would not greatly improve their skill (the second study), and by offering participants a small or large monetary reward for task performance (third study). Questionnaires used in the first and third studies confirmed that these manipulations did affect motivation.

Task performance reflected these differences in reported motivation. In all three studies, ego-depletion had the usual effect of impairing performance on the second task, but only for

participants who received the smaller incentive or no incentive at all. For participants who were offered a large enough incentive, executive control on the first task did not result in significantly worse performance on the second task. Motivation compensates for the effect of ego-depletion.

Do these results show that differences in motivation account entirely for differences in performance in all of the studies discussed thus far? Muraven and Slessareva acknowledge that their results are compatible with a couple of different interpretations. Perhaps ego-depletion reduces people's motivation to perform well on the second task, and offering people an incentive restores this motivation. This interpretation is consistent with the Motivation Only Explanation. Another interpretation, however, is that ego-depletion impairs mental capacities involving executive control, making the second task more difficult, and offering people a strong enough incentive provides them with some extra motivation, causing them to try harder on the second task, improving their performance in spite of the increased difficulty of this task.

By analogy, consider Charlie's classmate Rachel, who is getting A's in pre-calculus. Perhaps Rachel's grades are better than Charlie's because Rachel has more motivation than Charlie, or perhaps because Rachel is better at math than Charlie, or perhaps both explanations are partially correct. Suppose Mom offers Charlie an incentive to get an A on his next math test by threatening to take away his Xbox if he gets anything worse than an A. Charlie finds this incentive to be highly motivating, studies very hard, and earns an A on the test. One interpretation of the results of this "experiment" is that Charlie's grades were worse than Rachel's because Charlie had less motivation than Rachel, and that the prospect of having his Xbox taken away increased his motivation enough to cause him to get an A, perhaps even enough to have as much motivation as Rachel. Another interpretation, however, is that, while Rachel and Charlie have equal amounts of motivation, Rachel's grades are better simply because

she is better at math. In this case, the prospect of losing his Xbox increases Charlie's motivation to a level much greater than Rachel's, causing him to try much harder (by studying harder) causing him to perform as well as she does on the exam even though she is better at math than he is.

One dissimilarity between this example and the studies in Muraven and Slessareva (2003) is that, in the case of Charlie and Rachel, we can actually determine which interpretation is correct by comparing how hard Charlie and Rachel try. Charlie tries as hard as Rachel before receiving an incentive, and tries much harder than her after receiving one, then the second explanation is probably correct. Rachel's grades have been better than Charlie's up until this point probably because she is better at math than he is, not because she has more motivation. Otherwise, maybe the first interpretation is correct. We can determine which interpretation is correct by measuring how hard Charlie tries before and after receiving an incentive, and comparing both levels of effort to Rachel's. We can measure both Charlie's and Rachel's efforts by looking at the way they study, and how much time they spend studying. In the studies in Muraven and Slessareva (2003), there is no analogous way to measure how hard people in different groups are trying on the second task. All we can measure is their performance on this task, but differences in performance are what we are trying to explain in the first place. To rely on task performance as an indicator of how hard people try would beg the question against the difficulty explanation.

We need some other way of determining which interpretation is correct that does not require directly measuring effort. Fortunately, there are other differences between these two interpretations that might suggest a way of testing which is correct. For instance, if worse performance on the second task is only due to lack of motivation, then providing enough extra

motivation should always improve performance on this task. However, if worse performance on the second task is due to an impaired capacity, then providing some extra motivation might or might not improve performance on the second task by causing a person to try hard enough to compensate for this impairment.

When the capacity in question is not impaired too severely, perhaps one can compensate for the impairment by trying hard enough. Sometimes, however, the capacity in question might be impaired so severely that one cannot compensate for this impairment no matter how hard one tries. In this case, some extra motivation should not reverse the effects of ego-depletion.

By analogy, suppose Charlie has a severe concussion a few days before his exam, resulting in brain damage that significantly lowers his IQ. If the loss is severe enough, then he will not get an A on the test no matter how hard he tries, and threatening to take his Xbox away will not cause him to get an A.

If exercising executive control depletes a limited resource that makes tasks requiring executive control more difficult, then perhaps for severe enough cases of ego-depletion some incentives will not affect performance on the second task. Vohs, Baumeister, and Schmeichel (2012) produced a conceptual replication of the results of Muraven and Slessareva's results, and tested the hypothesis that extra motivation would not compensate for the effects of ego-depletion when ego-depletion is severe enough.

To obtain varying degrees of ego-depletion, they required some participants to perform four executive control tasks in a row (severe depletion), some participants to perform one executive control tasks (mild depletion), and some participants to perform no executive control tasks. Participants from all three groups would then perform two tasks for measuring delay of gratification and logical reasoning. Some participants were told that good performance on these

tasks would provide valuable information that would benefit consumer welfare, happiness, and health, while other participants were not told about any benefits. To ensure that this manipulation actually did provide extra motivation, experimenters administered a questionnaire at the end of the experiment. Participants who were told about the benefits of good performance reported significantly greater motivation to perform well than participants who were not told about any benefits. Also, there was no significant relationship between the number of tasks people performed prior to the measurement phase and reported motivation, suggesting that performing four tasks, one task, or no tasks before the measurement phase did not affect motivation.

For participants who did not receive extra motivation, executive control prior to the measurement phase resulted in significantly worse performance on both the delayed gratification and logical reasoning tasks, just like in all the studies discussed thus far. For participants who did receive extra motivation, performing one executive control task prior to the measurement phrase had no significant effect on delayed gratification and logical reasoning, which suggests that the extra motivation compensated for the effect of the first task just as Muraven and Slessareva had found. However, for participants who performed four tasks prior to the measurement phase, those with some extra motivation did not perform significantly better than those who did not receive the extra motivation, suggesting that four tasks depleted people severely enough to make the extra motivation ineffective. Since severe enough ego-depletion makes extra motivation ineffective, these results suggested that differences in performance on the delayed gratification and logical reasoning tasks are not entirely due to differences in motivation. These results are more consistent with the Difficulty Explanation than with the Motivation Explanation.

Since some experimental results are not consistent with the Mood, Motivation Only, and Belief explanations, the Difficulty Explanation is the best explanation for the entire body of experimental results discussed so far. Exercising executive control depletes some limited physiological resource, probably blood glucose, and tasks that require executive control are more difficult when this resource is depleted. People suffering from ego-depletion perform worse on these tasks because these tasks are more difficult for them, the same way that running several miles is more difficult for someone suffering from cardiovascular exhaustion. In the next two chapters, I will argue that this conclusion has some important philosophical consequences for the relationship between desire and motivation, and also for moral responsibility.

## 2 WILLPOWER AND MOTIVATION: HOW WE DO WHAT WE DON'T WANT TO DO

One project of action theory is to develop a theory of motivation that explains why people intentionally act one way rather than another. For instance, why does someone go out for a drink after work rather than going home to grade papers? Philosophers who advocate the Humean approach to motivation attempt to explain people's actions entirely in terms of the strengths of their desires. Someone intentionally goes out for a drink rather than going home because the desire to go out is stronger than the desire to go home. In other words, we only act on a desire that is stronger than any conflicting desire. In this chapter, I argue that the Humean approach makes testable predictions, and that experimental studies on self-control disconfirm these predictions and provide evidence against the Humean approach.

In Section 2.1, I introduce the Humean approach and explain why this approach has a strange implication about self-control: Self-control often involves intentionally trying to resist a desire, but the Humean approach is committed to the claim that we never intentionally try to resist a desire that is stronger than any conflicting desire. For instance, I will never intentionally try to resist my desire to stay in bed when this desire is stronger than my desire to get up.

In Section 2.2, I argue that the Humean approach predicts that certain self-control tasks are exceptions to the phenomenon of Fatigue Generality discussed in the previous chapter. However, experimental studies find that these tasks are not exceptions to Fatigue Generality. Since these studies disconfirm the Humean approach's prediction, they provide evidence against the Humean approach.

## 2.1 The Humean Approach

Hume famously argued that “reason” is the “slave of the passions,” meaning that reason alone cannot motivate someone to act in a certain way because passions are necessary for motivation (Hume, *Treatise*, 2.3.3). In contemporary action theory, Hume’s position on motivation inspired the “Humean” approach to motivation, according to which beliefs are not sufficient for the motivation to act in a certain way, because desires are necessary for motivation (Sinhbabu 2009). For instance, every semester I have students who never come to class even though they know they cannot pass without attending. According to the Humean approach, the belief that one cannot pass without attending is not sufficient to motivate people to come to class. A desire, such as the desire to avoid failing, is necessary for motivation.

While philosophers disagree about what desires are, they tend to agree about certain characteristics of desire. First, desires are mental states, like beliefs and emotions.

Second, desires have content. In other words, like beliefs and perhaps emotions, a desire is about something. We often say that one has a desire for something, a desire to do something, or a desire that something happen or that something is the case.

Third, desires motivate. Specifically, desires motivate us to act in ways that we believe will accomplish what our desires are about. If I have a desire for a margarita that is sufficiently sour and not too sweet, and I believe mixing one part triple sec, two parts lime juice, and three parts tequila makes a margarita that is sufficiently sour and not too sweet, then my desire for such a margarita will motivate me to mix these ingredients together in this way.

Finally, desires are non-truth-apt. In other words, unlike beliefs, desires cannot be true or false. For beliefs, what a belief is about determines whether that belief is true or false. For instance, the belief that the earth is flat is false, while the belief that water is H<sub>2</sub>O is true. For

desires, on the other hand, what a desire is about determines what that desire motivates a person to do, but does not make that desire true or false. For instance, my desire for a margarita can motivate me to do what I believe will make one, but this desire is not true or false.

These characteristics provide a fairly uncontroversial, working definition for “desire:”

**Desire**

A non-truth mental state that motivates one to act in a way that one believes will accomplish what that mental state is about.

The Humean approach is committed to explaining everything people try to do in terms of this kind of mental state.

Sometimes a desire motivates a person to act in a certain way while another desire motivates her not to act in that way. When two desires conflict, how does the Humean approach explain why a person acts in one way rather than another? A natural way for the Humean approach to answer this question is to claim that people act according to the following principle:

**The Strength Principle<sup>1</sup>**

Whenever someone tries to do  $\phi$ , her desire to do  $\phi$  is stronger than any conflicting desire.

With this principle, the Humean approach provides a “hydraulic” picture of motivation in which conflicting desires “push” a person in opposing directions with a kind of motivational “pressure,” and stronger desires “overpower” weaker ones.<sup>2</sup>

The Humean approach is appealing because this principle makes it flexible enough to provide a simple explanation for a variety of intentional behavior: someone tries to do  $\phi$  because her desire to  $\phi$  is stronger than any conflicting desire. Social and behavioral scientists seem to

---

<sup>1</sup> Advocates of the Humean approach offer different formulations of this principle such as “Agents do whatever maximizes expected desire satisfaction” (Sinhababu 2011) or “whenever we try to do something, we want to do that thing more than we want to do anything else we can do” (Kennett and Smith 1996). The first formulation entails that we always act in a way that is instrumentally rational, while the second formulation entails that we never try to watch TV and exercise at the same time if one desire is stronger than the other, unless “jointly satisfiable contemporaneous desires agglomerate” (Mele 1998). My formulation of the Strength Principle is compatible with these commitments, but does not rely on them.

<sup>2</sup> The hydraulic metaphor originates in McDowell (1981).

rely on this principle when they use decision theory to model human and animal behavior. These models assign a value called “utility” to various behavioral options, where a person or organism always takes the option with the greatest utility.<sup>3</sup> Assuming that utility measures desire strength, advocates of the Humean approach could argue that the best explanation for the success of these decision theoretic models is that humans and perhaps some non-human animals always act according to the Strength Principle.

Someone might think the Strength Principle does not allow the Humean approach to explain why people sometimes resist their desires. If I only act on the strongest of two or more conflicting desires, then how do I intentionally resist my desire to eat something delicious but unhealthy, to waste money on unnecessary purchases, to stay in bed, or to stop working on my paper? An advocate of the Humean approach could account for at least some cases of in which someone resists a desire. In order to explain how, I need to clarify what it means to “resist” a desire.

Mele (1990, p. 459) defines resisting a desire in the following way: “[a]n agent’s attempt to resist a desire is an attempt to manipulate his motivational condition, his environment, or both in such a way as to bring it about that he does not act on the desire being resisted.” By “motivational condition,” I believe Mele means the facts about someone’s mental state that affect what she has motivation to do, including all of the facts about someone’s desires and their relative strengths.

Mele discusses two ways of resisting a desire: “skilled strategies” and “brute resistance.” Skilled strategies include rearranging one’s environment in order to make oneself unable to act on a desire at a later time, avoiding tempting stimuli that cause the strength of a desire to

---

<sup>3</sup> Sometimes the Strength Principle is formulated in terms of decision theory, such as in Sinhababu (2011).

increase, and performing mental exercises, such as imagining the object of a desire as something undesirable.

However, there are some cases in which a person cannot resist a desire with one of these strategies, because she is not aware of an effective strategy for the situation. Even so, someone might still be able to resist a desire with what Mele calls “brute resistance,” which involves resisting a desire with “sheer effort of will,” what we ordinarily call “willpower.” When someone tries to resist a desire by exercising willpower, she makes a direct, mental effort to control the strengths of her desires. For instance, exercising willpower to resist my desire to stay in bed would involve making a direct, mental effort to weaken my desire to stay in bed.

However, examples like this are problematic for the Humean approach. According to the Strength Principle, I would not do anything to try to get myself up unless my desire to get up were stronger than any conflicting desire. But if my desire to get up were stronger than my desire to stay in bed, then I could get up without trying to resist the latter desire. If people always act according the Strength Principle, then do they ever try to do anything that involves resisting a desire?

Kennett and Smith (1996) argue that the Strength Principle is compatible with resisting a desire in cases of “diachronic” self-control in which one does something at an earlier time in order to prevent oneself from acting on a desire at a later time. Diachronic self-control is consistent with the Strength Principle because I can intentionally do something to resist a desire before it becomes stronger than any conflicting desire.

For instance, placing my alarm across the room before I go to bed is an example of diachronic self-control. I place my alarm across the room because I anticipate that when I wake up the next morning, my desire to stay in bed will be stronger than my desire to get up unless I

cannot silence my alarm without getting up and walking across the room. The Humean approach can explain why I place my alarm across the room as long as, before I go to bed, my desire to get up early is stronger than my current desire to sleep in. Resisting the desire to sleep in is consistent with the Strength Principle because my desire to get up is stronger when I place my alarm across the room.

Sometimes, however, people cannot anticipate changes in their desires far enough in advance to change their environment accordingly. Can we still try to prevent ourselves from acting in a certain way by using some kind of mental effort or “willpower” to control the strengths of our desires? The Strength Principle does allow the Humean approach to explain some cases in which people control their desires with willpower. As I am grading papers, for instance, my desire to stop grading gradually becomes stronger. Exercising willpower to prevent this desire from becoming stronger than my desire to continue grading is compatible with the Strength Principle as long as my desire to continue is still stronger than my desire to stop. Therefore, I could resist my desire to stop grading by exercising willpower before the desire to stop exceeds the strength of the desire to continue.

Perhaps the Strength Principle is compatible with diachronic self-control, but trying to weaken the desire to stay in bed when one is currently lying in bed would be an example of what Kennett and Smith call “synchronic” self-control, in which one does something to stop oneself from acting on a desire in the immediate present.

Synchronic self-control presents a difficulty for the Humean approach: if I am intentionally lying in bed, then my desire to lie in bed is stronger than any conflicting desire, including my desire to get up. However, I would not try to resist my desire to stay in bed unless some conflicting desire, such as the desire to get up, were stronger than the desire to lie in bed.

Therefore, if I am intentionally lying in bed, then I will not, at the same time, try to exercise willpower or do anything else in order to stop myself from lying in bed. In general, I never intentionally do anything to exercise synchronic self-control.

This consequence of the Strength Principle is a problem for the Humean approach because synchronic self-control seems very common. For instance, most people would claim to have experience of resisting the desire to stay in bed. Do these cases violate the Strength Principle?

Jeanette Kennett and Michael Smith (1996) discuss a way for people to exercise synchronic self-control without violating the Strength Principle: sometimes we do exercise synchronic self-control, but not by trying to do anything or by doing anything intentionally. In their words, synchronic self-control is “non-actional,” where an action is something that someone does intentionally—something one tries to do.

Not everything people do is intentional. For instance, people have many thoughts and other mental processes without trying to have them, because these thoughts and processes occur spontaneously and automatically. Kennett and Smith argue that some of these thoughts and processes affect desires.

For instance, thoughts about deadlines might strengthen the desire to get out of bed.<sup>4</sup> If I have a disposition to have these thoughts when I wake up in the morning, then these thoughts can cause my desire to get up to become stronger than my desire to stay in bed. Having these thoughts could be a way of exercising synchronic self-control.

Since I can have these thoughts without trying to have them, the Strength Principle does not require that my desire to get up is stronger than any conflicting desire to explain how I exercise self-control, change the strengths of my desires, and get up. Instead, my disposition to

---

<sup>4</sup> Again, Kennett and Smith make the same kind of argument using a different example.

have these thoughts explains how my desire to get up becomes strong enough to overcome my desire to stay in bed. Since exercising synchronic self-control in this way would not violate the Strength Principle, the Humean approach can allow for synchronic self-control as long as one does not exercise synchronic self-control intentionally.

However, this explanation of synchronic self-control is not consistent with many of our experiences of resisting desires, such as the desire to stay in bed. The experience of resisting the desire to stay in bed often seems to involve a feeling of effort, suggesting that resisting this desire is something I try to do. Perhaps this experience is an illusion and I am not intentionally trying to exercise willpower, but in the next section I present some experimental results suggesting that people do intentionally try to resist a desire when they exercise synchronic self-control.

## **2.2 What does Humean Approach entail for Fatigue Generality?**

Since the Humean approach puts certain constraints on how we exercise self-control, perhaps this approach makes predictions that can be tested by studies like those in Chapter 1. Recall the phenomenon of Fatigue Generality, in which performing one task weakens performance on a second task, even if these tasks have nothing in common other than self-control or executive control. For some of these tasks, like the puzzle, handgrip, and cold pressor, performance depends on resisting a desire, like the desire to give up on the puzzle, the desire to release the handgrip, or the desire to pull one's arm out of the cold water.

There is a simpler way and a more complicated way for the Humean approach to explain why performing an executive control task makes people perform worse on these self-control tasks. The simpler way has a couple of problems: it is committed to the Motivation Only

Explanation from Chapter 1, and it cannot account for the feeling of effort involved exercising self-control. The more complicated way for the Humean approach to account for Fatigue Generality avoids both of these problems: it is committed to the Difficulty Explanation from Chapter 1, and it does account for the feeling of effort involved in self-control.

Because of the Humean approach's commitment about synchronic self-control, however, I argue that this second way of accounting for Fatigue Generality predicts that tasks involving synchronic self-control are exceptions to Fatigue Generality. In other words, performing an executive control task should not affect performance on tasks requiring synchronic self-control. Since this prediction is testable, I mention some studies with tasks requiring synchronic self-control. These studies disconfirm the prediction of the Humean approach, so I argue that they provide evidence against this approach.

The simpler way for the Humean approach to account for Fatigue Generality is that performing an executive control task weakens some desires relative to others. Consider the cold pressor task, in which the participants must hold an arm under painfully cold water for as long as they can. Perhaps fatigue weakens certain desires, like the desire to satisfy an obligation to the experimenter or the desire to perform well on the task. Suppose a participant continues to hold her arm in the cold water because of a desire to perform well on the task. If the desire to relieve pain continuously gets stronger while her arm remains in the water, then, for people suffering from this kind of fatigue, this desire would not take as much time to become stronger than the desire to perform well. According to the Strength Principle, people pull out of the cold water when the desire to relieve the pain becomes stronger than any conflicting desire, which would happen sooner for people with a weaker desire to perform well on the task. If the fatigue of ego-

depletion weakens the desire to perform well, then people suffering from ego-depletion should pull out of the cold water sooner than people who are not suffering from ego-depletion.

There are two problems with this way of accounting for Fatigue Generality. The first problem is that accounting for Fatigue Generality in terms of changes in the relative strengths of people's desires entails the Motivation Only Explanation from Chapter 1. Weakening some desires relative to others reduces motivation to perform well on the second task. Recall from Chapter 1 that the Motivation Only Explanation claims that performing one executive control task reduces motivation to perform well on a second executive control task, which causes people not to try as hard on the second task.

In Chapter 1, I argued that experimental results from Vohs, Baumeister, and Schmeichel (2012) provide evidence against Motivation Only Explanation. The Motivation Only Explanation predicts that motivational incentives should cause some improvement on task performance for any level of ego-depletion, but Vohs, Baumeister, and Schmeichel found that motivational incentives only have this effect for milder levels of depletion, and that this effect disappears for severe levels of depletion. These results disconfirm the prediction of the Motivation Only Explanation, which is a problem for accounting for Fatigue Generality only in terms of changes in relative desire strengths.

The second problem for the simpler way of accounting for Fatigue Generality is one aspect of the phenomenology of self-control: the feeling of effort involved in acting contrary to some desires but not others. Intuitively, acting contrary to stronger desires should involve a greater feeling of effort than acting contrary to weaker desires, but an example will illustrate why the Strength Principle is not consistent with such a straightforward relationship between the motivational strength of a desire and this feeling of effort.

Imagine what it feels like to hold your arm in the tank of ice water. At first, the cold water would feel unpleasant, but keeping your arm in the tank should not require much effort. As the cold water begins to hurt, continuing to hold your arm in the tank should require more and more effort until you pull it out. Unlike holding your arm in the water, pulling it out of the water is easy and lacks the same feeling of effort.

This example suggests that, if the Strength Principle is correct, then the motivational strength of a desire does not determine the intensity of this feeling of effort. According to the Strength Principle, you hold your arm under the ice water as long as the desire to continue is stronger than the desire to remove your arm, and you remove your arm when the desire to remove your arm becomes stronger than the desire to continue. Therefore, holding your arm in the ice water is difficult even when the desire to continue is stronger than the desire to remove your arm. When the desire to relieve the pain becomes stronger than the desire to perform well, pulling your arm out is easy. The strengths of these desires do not explain why continuing feels difficult or why stopping feels easy.

Fortunately for the Humean approach, there is a more complicated way of accounting for Fatigue Generality that is not committed to the Motivation Only Explanation and can also account for the feeling of effort involved in self-control. Recall from Section 2.1 that the Humean approach can explain why people would intentionally exercise willpower in order to resist their desires in cases of diachronic self-control. For instance, someone could intentionally exercise willpower to resist her desire to remove her arm if she begins to exercise willpower before this desire becomes stronger than the desire to perform well. As long as the desire to perform well is stronger, then intentionally trying to resist the desire to remove her arm does not violate the Strength Principle.

If ego-depletion weakens the effectiveness of willpower, then the Humean approach could explain why people suffering from ego-depletion pull out of the cold water twice as soon. Perhaps when participants put one arm into the tank, the desire to perform well on the task is stronger than the desire to relieve the pain. As the pain increases, so does the desire to relieve the pain. Perhaps participants are aware of the increasing strength of this desire, and, since the desire to perform well on the task is still stronger than the desire to relieve the pain, they intentionally exercise willpower to resist the desire to relieve the pain. If willpower is less effective for participants suffering from ego-depletion, then they will not be able to resist the desire to relieve the pain as well as those who are not suffering from ego-depletion, and the desire to relieve the pain will become stronger than the desire to perform well more quickly than if willpower were more effective.

This way of accounting for Fatigue Generality is not committed to the Motivation Only Explanation. Accounting for Fatigue Generality in terms of an impaired capacity for willpower is more consistent with the Difficulty Explanation, because the impaired capacity for willpower makes certain desires more difficult to resist.

Accounting for Fatigue Generality in terms of the difficulty of controlling one's desires also provides a way to explain why holding one's arm in the water is difficult while pulling out is not. Perhaps holding an arm in the cold water involves a greater feeling of effort because holding one's arm in the water requires intentionally controlling the strength of one's desires, whereas pulling out of the ice water does not. The feeling of effort comes from intentionally controlling one's desires.

While this more complicated way of accounting for Fatigue Generality allows the Humean approach to avoid the two problems raised by the simpler way of accounting for Fatigue

Generality, the more complicated way suffers from a different problem: because of the Humean approach's commitments about synchronic self-control mentioned in Section 2.1, this more complicated way of accounting for Fatigue Generality makes a testable prediction: tasks requiring synchronic self-control should be exceptions to Fatigue Generality. In other words, performing an executive control task should not affect a second task requiring synchronic self-control.

Recall from Section 2.1 that the Humean approach is committed to claiming that we never intentionally exercise synchronic self-control. Instead, we can only exercise synchronic self-control by unintentionally having thoughts or other mental processes that affect our desires. Since we must have these thoughts or other processes without trying to have them, they must occur automatically and effortlessly, which means they should not involve executive control the way intentionally exercising willpower in order to resist a desire does.

Recall from Chapter 1 that ego-depletion only effects mental capacities involving executive control, and these capacities are effortful and voluntary. For instance, complex reasoning is more effortful and voluntary than recalling words and facts, and one study mentioned in Chapter 1 found that ego-depletion impairs complex reasoning more than recalling words and facts (Schmeichel et al., 2003). Ego-depletion affects willpower because willpower involves executive control. It is a kind of mental effort that people exercise voluntarily.

The Humean approach predicts that tasks requiring synchronic self-control are not affected by ego-depletion because the Humean approach is committed to explaining synchronic self-control only with thoughts and mental processes that occur automatically and effortlessly. Since thoughts and mental processes that occur automatically and effortlessly do not involve

executive control, they should not be affected by ego-depletion. Therefore, ego-depletion should not affect synchronic self-control.

However, I argue that some studies disconfirm this prediction. In these studies, experimenters measure performance on tasks that seem to require synchronic self-control, and ego-depletion does affect performance on these tasks. For instance, Vohs and Heatherton (2000) recruited female college students who scored highly on a scale indicating that they were chronic dieters. Participants began with the video task described in Chapter 1, with one group of participants suppressing their emotional responses and the other groups allowing themselves to react naturally to the video. After the video task, experimenters would present the participant with three large containers of ice cream, and ask the participant to taste and rate the three different flavors. Before leaving the room, the experimenter would tell the participant to “help yourself to any ice cream you want; we have tons in the freezer.”

On the ice cream task, self-control is measured in the quantity of ice cream that participants scoop into their bowls to eat. Resisting the desire for more ice cream would be an example of synchronic self-control because, if the participant has already begun scooping, she must resist the desire for more ice-cream while she is intentionally scooping more ice cream. While she is intentionally scooping more ice cream, the Strength Principle entails that her desire for more ice cream is stronger than any conflicting desire. As long as her desire for more ice cream is stronger than any conflicting desire, the Strength Principle requires that she will not intentionally do anything in order to resist this desire. The only way she could resist this desire is by having thoughts or other mental processes that occur non-intentionally and weaken this desire or strengthen some conflicting desire. For a chronic dieter, thoughts about physical health, for instance, might strengthen the desire to avoid gaining weight. But ego-depletion

would not prevent these thoughts or any other mental process that occurs automatically and effortlessly, because ego-depletion only affects mental processes involving executive control. Since resisting the desire for more ice cream is an example of synchronic self-control, the Humean approach predicts that ego-depletion should not affect how much ice cream people scoop into their bowls.

In fact, participants who tried to suppress their emotions on the video task ate significantly more ice cream than participants who did not try to suppress their emotions, confirming that ego-depletion impaired the ability to resist the desire for more ice cream. Since the Humean approach predicts that ego-depletion does not affect synchronic self-control, these results are not consistent with this approach. On the other hand, these results are quite consistent with the possibility that, while intentionally scooping more ice cream, people also intentionally exercise willpower in order to try to resist the desire for more ice cream.

Advocates of the Humean approach might try to defend the Strength Principle by arguing that the ice cream task is really an example of diachronic self-control. Suppose that when someone begins scooping ice cream, she desires two scoops of ice cream more than she desires to avoid consuming fat and sugar, but she desires to avoid consuming fat and sugar more than three scoops of ice cream. As she continues scooping, her desire for a third scoop starts getting stronger. Intentionally trying to resist the desire for a third scoop would not violate the Strength Principle as long as she begins trying to resist before this desire becomes stronger than any conflicting desire. In other words, resisting the desire for a third scoop is diachronic self-control.

If people suffering from ego-depletion cannot control the strength of this desire as effectively as people who are not suffering from ego-depletion, then ego-depletion increases the chance that the desire for a third scoop will eventually become stronger than the desire to avoid

consuming fat and sugar. If the ice cream task really involves diachronic self-control, then the Humean approach can provide an explanation for why ego-depletion causes chronic dieters to scoop more ice cream.

If the ice cream task involves diachronic self-control, then the Humean approach can explain any feeling of effort that participants experience on the ice cream task. Perhaps they experience a feeling of effort that comes from intentionally trying to control the strength of the desire for a third scoop.

The problem with this explanation of the feeling of effort is that not all effortful cases of synchronic self-control can be redescribed as diachronic self-control. There are some cases of resisting a desire, such as resisting the desire to stay in bed, that must involve genuine synchronic self-control, and these cases do involve the feeling of effort. If I am intentionally lying in bed, then the Strength Principle requires that my desire to lie in bed is stronger than any conflicting desire. Sometimes I get out of bed by resisting this desire, but the Humean approach requires that I can only resist this desire by non-intentionally having some thoughts or mental processes that affect the strengths of my desires. However, if I have these thoughts or processes occur automatically and effortlessly, then why does resisting my desire to stay in bed feel so effortful?

The Humean approach can explain the feeling of effort involved in diachronic self-control, because this approach can explain why people would intentionally try to control their desires to stay in the cold water and why they do not need to control their desires to pull out of the cold water. A similar explanation is not available for the feeling of effort involved in synchronic self-control, because the Humean approach requires that synchronic self-control involves mental processes that occur automatically and effortlessly.

I suppose advocates of the Humean approach could argue that certain thoughts and mental processes involve a feeling of effort even though they occur automatically and without any actual effort. However, these advocates of the Humean approach would need to explain why these thoughts and mental processes involve a feeling of effort when there is no actual effort, while other thoughts and mental processes do not. Some thoughts, like thoughts about how little sleep one got the night before, should strengthen the desire to stay in bed. Having these thoughts does not seem to involve this feeling of effort. Advocates of the Humean approach would need to explain why certain thoughts and mental processes involve a feeling of effort while others do not.

### **2.3 Conclusion**

To explain why we try to act one way rather than another, the Humean approach relies on the Strength Principle, which says we act according to the stronger of two or more conflicting desires. This principle allows the Humean approach to explain why people intentionally try to resist a desire in cases of diachronic self-control, but not for synchronic self-control. For synchronic self-control, the Humean approach requires that people resist their desires by non-intentionally having certain thoughts and mental process that affect the strengths of their desires.

However, experimental results suggest that ego-depletion only affects mental processes that people intentionally try to carry out, so the Humean approach predicts that ego-depletion would not affect synchronic self-control. Some studies, however, include tasks that seem to require synchronic but not diachronic self-control, and ego-depletion does affect performance on these tasks, which contradicts the prediction of the Humean approach. The Humean approach also cannot explain the feeling of effort involved in synchronic self-control.

Since these problems are consequences of the Strength Principle, I argue that we should abandon this principle. If people can act in ways that violate the Strength Principle, then what are the consequences for action theory? One consequence is that one can intentionally act contrary to a stronger desire. Exercising willpower might be an example of such an action. One might intentionally try to resist a desire to stay in bed even though this desire is stronger than the desire to get up.<sup>5</sup>

The strength of a desire is a measure of how much that desire influences various actions. The Strength Principle requires that all desires have strengths that are measurable on one common scale, so the strongest relevant desire always determines what a person does. Since there is only one measure of influence, the same desire cannot influence different actions in different amounts. Denying the Strength Principle eliminates this restriction, allowing one desire to have greater influence over one action than another desire, but less of an influence on some other action than that same other desire. For instance, the desire to get up may have a weaker influence on whether I get up than my desire to stay in bed, but the former desire may have a stronger influence on whether I exercise willpower. Denying the Strength Principle allows people to try to act in two conflicting ways at the same time, which is an interesting possibility for action theory.

---

<sup>5</sup> Mele (1997) discusses this possibility in detail while arguing against the Strength Principle.

### **3 SELF-CONTROL FAILURE: WHY IT'S NOT (COMPLETELY) MY FAULT**

People have strong desires to do what they know they should not do, from sleeping in when they know they should get up to relapsing with alcohol or addictive drugs. Sometimes we resist these desires by exercising willpower. Sometimes we try to exercise willpower but still fail to resist anyway. Self-control failure raises an important philosophical question: is someone fully responsible for doing something wrong if she tried to resist the desire to do so?

In Chapter 1, I argued that ego-depletion makes self-control more difficult. In this chapter, I will argue that people suffering from ego-depletion are not fully responsible for failures of self-control.

Suppose a recovering addict revisits his former heroin den intending to leave as quickly as possible without using heroin, but upon arrival he encounters sights, sounds and smells that vividly remind him of the way it feels to get high, eliciting a strong craving for heroin. If he tries very hard to avoid acting on this desire but ultimately fails to resist, is he fully responsible for relapsing?

Someone might argue that our recovering addict is responsible for relapsing because his addictive cravings are a consequence of a series of choices earlier in his life, or because these cravings are a consequence of his choice to put himself in a situation that would elicit these cravings. But perhaps he is not fully responsible for becoming an addict because he was not an adult at the time, and maybe he is not fully responsible for revisiting the heroin den because someone takes him there by force to repay a remaining debt. If so, then responsibility for relapsing, or at least full responsibility, would not trace back to becoming an addict or to revisiting the heroin den. Furthermore, perhaps he was not able to foresee the consequence of

relapse, because he lacked access to relevant information about heroin addiction before becoming addicted, or because he had good reason to believe he would not encounter certain temptations at the heroin den. We might think responsibility only traces back to something one did at an earlier time if one was to some extent able to foresee the consequences at that time. If full responsibility for relapsing does not completely trace back to becoming an addict or to revisiting the heroin den, then we are left with the same question as before: if he tries very hard to resist the desire for heroin, is he fully responsible for failing to resist this desire?

Relapsing at the heroin den is an example of self-control failure. I use the term “self-control failure” for situations in which a person tries to resist a desire and fails, such as when the recovering addict tries but fails to resist his desire for heroin. When he encounters the stimuli that elicit a strong craving for heroin, the recovering addict attempts to exercise willpower in order to resist. Heroin cravings are notoriously difficult to resist with sheer effort, which is why the recovering addict ultimately fails to resist in spite of trying very hard.

Is the fact that the recovering addict failed to resist even though he tried very hard relevant in any way to the question of whether he is fully responsible for relapsing? The reason why this question is so important to consider is that someone who lacks full responsibility for the way she acts deserves to be treated more leniently than someone who is fully responsible. Some circumstances, known as “excusing conditions,” make a person less responsible for the way she acts than she would be in other circumstances. If two college students are late for an exam, one because of sleeping in and the other because of a sprained ankle on the way to class that slows her down, then the second deserves more leniency than the first.

One reason to excuse someone for acting or failing to act in a certain way is that this person was not able to act differently. Philosophical, legal, and common, everyday discourse

take for granted that moral responsibility depends on what a person is able to do. If someone is not able to prevent something bad from happening, for instance, then she is not morally responsible for it. Suppose an innocent person will die from a poisonous snakebite, and the only way to save him is to run a long distance in a short amount of time to retrieve the antidote. You try to retrieve the antidote in time, but fail because you cannot run fast or far enough. Since you failed to save his life because you were not able to do otherwise, you are not responsible for his death. Our lack of responsibility in cases like this one follows from the following principle:

**Principle of Alternate Possibilities (PAP)<sup>6</sup>**

A person is not morally responsible for what she has done if she did it because she could not have done otherwise.

This principle links moral responsibility to what a person is able to do, and what it takes to be able to do something is disputed by libertarians and compatibilists in the free will debate. Since I want the arguments in this chapter to persuade both libertarians and compatibilists, I will avoid relying on either conception of free will.

While libertarians and compatibilists disagree about what it takes to be able to do something, they can agree about certain things one cannot do in certain cases. In the example I just gave, for instance, libertarians and compatibilists can agree that you are not able to save a person's life. The ability to save his life depends not just on whether determinism is true and whether free will is compatible with determinism, but also on your physical capacities to run fast and far enough. Even if libertarians and compatibilists do not agree about determinism and free will, they can still agree that your body's physical limitations prevent you from retrieving the antidote in time.

Libertarians and compatibilists can also agree about cases in which one cannot do something because of mental, rather than physical, limitations. Suppose retrieving the antidote

---

<sup>6</sup> This principle first received this name in (Frankfurt 1969).

requires a secret code. You know a complex algorithm for breaking the code, but performing this algorithm requires difficult mental arithmetic. You try to break the code, but you are not good enough at mental arithmetic to carry out the algorithm in time. Again, even if libertarians and compatibilists do not agree about determinism and free will, they can still agree that your cognitive limitations prevent you from retrieving the antidote in time. For both of these examples, the Principle of Alternate Possibilities entails you are not responsible for failing to save someone's life, regardless of whether determinism is true or whether free will is compatible with determinism.

Does this principle also apply to cases of self-control failure, in which someone tries but fails to resist a strong desire? Consider a similar case of an innocent person who will die from poisoning, but getting to the antidote requires crossing a shallow stream, and you have a severe phobia of running water due to a traumatic experience in your childhood. Regardless, you run to the water's edge and attempt to put your right foot into the stream, only to recoil in fear. After a few minutes of repeatedly trying to step into the water and recoiling in fear each time, he dies. Are you morally responsible failing to save him? If your desire not to step in the water actually made you unable to cross the stream, then the Principle of Alternate Possibilities entails that you are not morally responsible for his death.

This example assumes that you are not able to resist your desire to step into the stream. Are there any real cases of such irresistible desires? If we need certain capacities like willpower in order to resist certain desires, and if these capacities have limitations just like physical speed and endurance, and mental arithmetic, then perhaps the limitations of willpower make some desires irresistible. In Chapter I, I discussed a growing body of experimental results confirming that willpower is subject to such limitations. Exercising willpower depends on limited

physiological resources, and depleting these resources impairs willpower, making some desires more difficult to resist.

If willpower is limited in this way, it may follow that depleting these physiological resources enough can actually make a person unable to resist certain desires. Let's call this claim the "Inability Hypothesis:"

**Inability Hypothesis**

Some cases of ego-depletion are so severe that someone acts the way she does because she is not able to do otherwise.

For instance, if our recovering heroin addict resists the desire for heroin by exercising willpower, he will deplete some of the physiological resources on which willpower depends. The longer he exercises willpower, the more of these resources he will deplete. The Inability Hypothesis claims that it is possible for him to become so depleted that he is no longer able to continue resisting this desire. If this desire really is irresistible, then the Principle of Alternate Possibilities entails that he is not responsible for failing to resist. The Inability Hypothesis and Principle of Alternate Possibilities together entail that there are severe cases of ego-depletion in which people are not responsible for failures of self-control. If both the Inability Hypothesis and the Principle of Alternate Possibilities are true, then severe ego-depletion constitutes an excusing condition.

However, there are a couple of problems with this argument. First, the Inability Hypothesis is an empirical claim which could turn out to be false. In fact, denying the Inability Hypothesis is not, strictly speaking, inconsistent with the strength model or the experimental results discussed in Chapter I. Perhaps ego-depletion makes certain desires more difficult to resist, but there is no point at which a person becomes completely unable to resist that desire.

Ego-depletion could make a desire more difficult to resist even though a person could always resist this desire simply by trying harder.

In general, perhaps a person could always resist by trying hard enough regardless of how difficult some desire is to resist. Joel Feinberg illustrates this possibility by comparing it to a couple of colorful examples:

[S]trictly speaking, no impulse is “irresistible.” For every case of giving in to a desire, I would argue, it will be true that, if a person had tried harder, he would have resisted it successfully. The psychological situation is never—or hardly ever—like that of a man who hangs from a windowsill by his fingernails until the sheer physical force of gravity rips his nails off and sends him plummeting to the ground, or like that of the man who dives from a sinking ship in the middle of the ocean and swims until he is exhausted and then drowns. Human endurance puts a severe limitation on how long one can stay afloat in the ocean; but there is no comparable limit to our ability to resist temptation (Feinberg, 1970, pp. 282-283).

Feinberg seems to claim that there is an interesting difference between the psychological capacity of resisting desires and other capacities like physical strength and endurance: we can always successfully resist a desire by trying hard enough even though there are some physical challenges that we cannot overcome no matter how hard we try. One wonders why a resisting would be different from our physical capacities in this way.

Some examples, however, will suggest that Feinberg’s claim is quite plausible, and that the Inability Hypothesis may not be so plausible. One way to get someone to try harder is to motivate that person with incentives. The recovering addict would most likely resist the desire for heroin if commanded at gunpoint not to use heroin. He would also most likely resist for a large enough monetary reward. If a great enough incentive would get someone to try hard enough to successfully resist a desire, then someone could argue that this person is able to resist. Even if the recovering addict were suffering from a very severe case of ego-depletion, he would probably still avoid using heroine if threatened at gun point.

One reply to this argument would be to deny the claim that, if a strong enough incentive would get someone to act in a certain way, then that person is able to act in that way even without that incentive. Sometimes an incentive can actually change what someone is able to do. For instance, Al Mele offers the example of:

...the woman who, under ordinary circumstances, cannot even budge a 300 pound weight, but who, upon finding her child pinned under a 400 pound timber manages, due to a sudden burst of adrenalin, to raise the timber from his body. Surely, it would be misleading to say that she can lift 400 pounds, if we leave it at that. Rather, we should say that in ordinary circumstances she cannot do this (no matter how hard she tries), although in a certain kind of exceptional circumstance she can (Mele 1990).

The incentive of saving her child causes the woman to successfully lift the 400 pound timber, which does not imply that she would be able to lift the 400 pound timber even without this incentive. The incentive changes what she is able to do. Similarly, someone could argue that having a strong enough incentive not to use heroin changes what the recovering addict is able to do.

The incentive of a threat to this woman's child causes a physiological change that actually makes her strong enough to do something she normally is not physically strong enough to do. Perhaps the adrenalin rush of being threatened at gunpoint would provide the recovering addict with the physiological resources, and thus the willpower, to resist a desire he otherwise would not be able to resist. The fact that threatening him at gun point would get him to resist does not show that he is really able resist without such a threat.

One way to avoid this problem would be to focus on the example of monetary incentives rather than immediate, visible threats such as being threatened at gun point or having one's child trapped under a log. In addition to providing extra motivation to do something, these powerful negative incentives have the side effect of changing a person's physiological state in a way that

enhances what that person is actually able to do, whereas the monetary incentives might not have this side effect to the same extent.

Focusing on monetary incentives also reveals the reason for why incentives are relevant to what that person is able to do. The reason why the recovering addict would resist the desire for heroin for a large enough monetary incentive is that this incentive motivates him to try harder to resist. What the incentive objection shows is that the recovering addict could try harder to resist his desire, regardless of how severely depleted he is. If he would successfully resist this desire by trying hard enough, then he is able to resist.

However, this objection still overlooks another important possibility: Are there cases of depletion that are severe enough or cases of desires that are strong enough to overcome the effect of any possible incentive a person could have? A familiar example is the pain box from the classic sci-fi novel *Dune*. The protagonist Paul puts his hand inside of a box which causes a gradually increasing feeling of pain. If he removes his hand from the box too early, he will be killed by a poison tipped needle. While Paul does not remove his hand from the box until he passes the test, we can imagine the pain reaching an intensity for which no one would continue to resist the desire to remove his or her hand, even to avoid death, let alone for any amount of money.

To consider a real life example, almost everyone cracks under torture, which is unsurprising in light of ego-depletion. Torture victims are often deprived of sleep and food, and tortured repeatedly, which should deplete the physiological resources upon which willpower depends and prevent one from restoring these resources. If there is some possible torture of great enough severity and duration that we would not withhold the information sought even for a

later monetary reward no matter how large, then the incentive objection does not show that the Inability Hypothesis is false.

However, even if the Inability hypothesis is correct, cases of ego-depletion severe enough to entail a total lack of responsibility for self-control failure would be too rare for my argument to apply to ordinary cases of self-control failure in daily life. However, I want to argue that ego-depletion also affects how we should view our own common failures of self-control.

Since most cases of ego-depletion are not severe enough to satisfy PAP, perhaps there is a similar principle about moral responsibility that is less demanding than the Principle of Alternate Possibilities. If responsibility comes in degrees, then perhaps degrees of responsibility depend not just on whether one is able to do otherwise or not, but also on how much one's capacity to do otherwise is impaired:

**Capacities Principle**

If someone acts the way she does partly because of an impaired capacity, then she is less responsible for the way she acts than if that capacity were not impaired.

This principle is consistent with our intuitions about responsibility. Suppose a college student is late for an exam because she sprains her ankle on the way to class and limps the rest of the way there. With greater effort and pain, she could continue at her normal pace and arrive at the exam on time. Because her capacity for walking at a normal pace is impaired, continuing at her normal pace requires more effort than before her injury, so she slows down and hobbles the rest of the way to the classroom, arriving 15 minutes late.

Getting to class on time requires much more effort for her than it is for everyone else, which is why it would be unfair to hold her fully responsible for arriving late. Holding her fully responsible would be to hold her to a higher and more severe standard than everyone else, because arriving in time would not require nearly as much effort or pain for them.

On the other hand, if the student arrived 15 minutes late because she slept in, then, other things equal, it would not be unfair to hold her fully responsible. Everyone else had to wake up by a certain time to make it to the exam, and presumably waking up requires comparable effort for them. The Capacities Principle is quite consistent with the way we attribute responsibility to people in particular cases, and we rely on this principle frequently in everyday life.

However, the Capacities Principle has an important consequence about responsibility for self-control failure: If the Strength Model is correct about the limits of willpower, then people suffering from ego-depletion are not as responsible as they would be without suffering from ego-depletion. Recall from Chapter 1 that the Strength Model is committed to a particular explanation for the effects of ego-depletion: the Difficulty Explanation, according to which people suffering from ego-depletion have impaired mental capacities, which makes it more difficult to do anything that requires these capacities. Willpower is one of the capacities that becomes impaired, and willpower is necessary for resisting certain desires. Therefore, ego-depletion makes these desires require more effort to resist.

If someone fails to resist a desire at least partly because of an impaired capacity for willpower, then Capacities Principle entails that she is not as responsible for failing to resist this desire as she would be if she were not suffering from ego-depletion. In other words, ego-depletion makes people less than fully responsible for self-control failure.

For our recovering addict, resisting the desire for heroin over some period of time would continuously deplete the physiological resources that willpower depends on, which impairs willpower. If he eventually fails to continue resisting at least partly because of impaired willpower, then he is less responsible than he would be if he decided to shoot up even though he were not suffering from ego-depletion at all. If he decided to shoot up without trying to resist at

all, or before even arriving at the heroin den, then he might still be fully responsible. This disadvantage of suffering from ego-depletion takes some of this responsibility away, even if, strictly speaking, he could have continued to resist if he had tried even harder.

In this chapter, I raised the question of whether we are fully responsible for failures of self-control, and argued that we are not fully responsible when self-control failure occurs at least partly because of ego-depletion. One reason that ego-depletion could make us less responsible would be if we were completely unable to avoid self-control failure. However, it is not known if ego-depletion can be severe enough to make us completely unable to avoid self-control failure. Even if there is such a severe level of ego-depletion, most people do not reach this level of ego-depletion very often. Therefore, people suffering from ego-depletion do not satisfy PAP very often, if at all.

To argue that ego-depletion makes people less responsible for common, everyday failures of self-control, I introduced the Capacities Principle, and argued that we do rely on this principle for determining how much responsibility people have in various cases. If ego-depletion makes self-control more difficult, then this principle entails that people are less responsible for self-control failure than if they were not suffering from ego-depletion.

#### 4 PHILOSOPHICAL CONCLUSIONS AND FUTURE PROJECTS

In the preceding chapters, I argued that the Strength Model has two important philosophical implications: First, the Humean approach to motivation is incorrect about the claim that we only act on the strongest of two or more conflicting desires. Second, people suffering from ego-depletion are less than fully responsible for failures of self-control.

Since neither of these claims follows unless the Strength Model is correct, I argued for the Strength Model in Chapter 1. In Section 1.2, I argued that the Strength Model makes three important predictions: Fatigue Generality, Reduced Prefrontal Activity, and the Physical Substance Prediction. I offered several examples of experimental studies that confirm each of these predictions. Unless there is some better explanation for the results of these studies, these studies provide good evidence for the Strength Model.

The Strength Model is committed to a specific explanation for the experimental result on willpower and executive control: Difficulty Explanation. In Section 1.2, I considered three alternative explanations for these experimental results: the Belief Explanation, the Mood Explanation, and the Motivation Only Explanation. Each of these alternative explanations makes testable predictions, and I mentioned studies that disconfirm each of these predictions. Since the Difficulty Explanation accounts for these experimental results while these other explanations do not, it is a better explanation than these alternatives. The Strength Model does seem to provide the best explanation for the experimental results discussed in Chapter 1, so these results provide good evidence for the Strength Model.

In Chapter 2, I relied on this conclusion to argue against the Humean approach to motivation. I explained why the Humean approach has a strange implication about self-control: we never intentionally exercise synchronic self-control. Because of this commitment, I argued

that the Humean approach predicts that tasks involving synchronic self-control should be exceptions to Fatigue Generality. Performing an executive control task should not affect performance on a task that requires synchronic self-control. I mentioned studies that disconfirm this prediction. These studies, I argued, provide evidence against the Humean approach.

In Chapter 3, I relied on the Strength Model to argue that people suffering from ego-depletion are not fully responsible for failures of self-control. In other words, they are not as responsible as they would be if they were not suffering from ego-depletion. According to the Principle of Alternate Possibilities (PAP), we are not responsible for something unless we could have done otherwise. However, the Strength Model does not necessarily entail that we cannot avoid self-control failure when suffering from ego-depletion. The Strength Model only entails that certain capacities necessary for avoiding self-control failure are impaired. Therefore, I proposed principle that is less demanding than PAP: the Capacities Principle, according to which impaired capacities entails less responsibility for failing to do something. The more effort required for someone to do something, the less we should blame or penalize them for failing to do it. Since ego-depletion makes self-control require more effort, this principle entails that people suffering from ego-depletion are not as responsible for self-control failure as they would be if they were not suffering from ego-depletion.

While Chapters 2 and 3 address just two philosophical questions about willpower, the Strength Model probably has other philosophical implications as well. If the Strength Model is correct about the limitations of willpower, philosophers should try to determine what these implications are. As more experimental studies produce evidence for the Strength Model, philosophers should devote more attention to these implications.

## References

- Baumeister, R. F., Bratslavsky E., Muraven, M., and Tice, D. M. (1998) Ego-Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology* 74(5), 1252-1265.
- Baumeister, R.F., Heatherton, T.F., & Tice, D.M. (1994). *Losing control: How and why people fail at self-regulation*. San Diego, CA: Academic Press.
- Baumeister, R. F., Vohs, K. D., and Tice, D. M. (2007) The Strength Model of Self-Control. *Current Directions in Psychological Science* 16(6), 351-355.
- Elliot, R. (2003). Executive functions and their disorders. *British Medical Bulletin*, 65, 49–59.
- Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy*, 66(23), 829-839.
- Fuster, J. M. (2008). *The Prefrontal Cortex*, 4<sup>th</sup> ed. (Academic Press)
- Hedgcock, W. M., Vohs, K. D., & Rao, A. R. (2012). Reducing self-control depletion effects through enhanced sensitivity to implementation: Evidence from fMRI and behavioral studies. *Journal of Consumer Psychology*, 22(4), 486-495.
- Inzlicht, M., & Gutsell, J. N. (2007). Running on empty neural signals for self-control failure. *Psychological Science*, 18(11), 933-937.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541.
- Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion—Is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological science*, 21(11), 1686-1693.

- 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.
- Kennett, J. and Smith, M. (1996) Frog and Toad Lose Control. *Analysis* 56.2, 63-73.
- Hume, D. *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, 1978).
- Lobel, A. (1972) *Frog and Toad Together*, (New York: HarperCollins).
- McDowell, J. (1981) Non-cognitivism and Rule Following, in S. Holtzman, C. Leich (eds.), *Wittgenstein: To Follow a Rule* (London: Routledge & Kegan Paul 1981), 141-162.
- Muraven, M., & Slessareva, E. (2003). Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin*, 29(7), 894-906.
- Feinberg, J. (1970). What is so special about mental illness. *Doing and deserving*, 272-92.
- Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, 74, 774-789.
- Mele, A. R. (1990) Irresistible Desires. *Noûs* 24(3), 455-472.
- Mele, A. R. (1997) Underestimating Self-Control: Kennett and Smith on Frog and Toad. *Analysis* 57(2), 119-123.
- Mele, A. R. (1998) Synchronic Self-Control Revisited: Frog and Toad Shape Up. *Analysis* 58(4), 305-310.

- Naccache, L., Dehaene, S., Cohen, L., Habert, M., Guichart-Gomez, E., Galanaud, D., et al., 2005. Effortless control: executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia* 43, 1318–1328.
- Norman, D.A. and Shallice, T. (1986) Attention to action: willed and automatic control of behavior. In *Consciousness and Self-Regulation* (Davidson, R.J. et al., eds), pp. 1–18, Plenum Press
- Schmeichel, B. J., Vohs, K. D., and Baumeister, R. F. (2003) Intellectual Performance and Ego Depletion: Role of the Self in Logical Reasoning and Other Information Processing. *Journal of Personality and Social Psychology* 85(1), 33-46.
- Sinhababu, N. (2009) The Humean Theory of Motivation Reformulated and Defended. *Philosophical Review* 118(4), 465-500.
- Sinhababu, N. (2011) The Humean Theory of Practical Irrationality. *Journal of Ethics & Social Philosophy* 6(1).
- Vohs, K. D. and Heatherton, T. F. (2000) Self-Regulatory Failure: A Resource-Depletion Approach. *Psychological Science* 11(3), 249-254.
- Vohs, K. D., Baumeister R. F., Schmeichel B. J., (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(5), 883-898.
- Vohs, K. D., Baumeister, R. F., & Schmeichel, B. J. (2012). Motivation, personal beliefs, and limited resources all contribute to self-control. *Journal of Experimental Social Psychology*. 48 943–947.