NORTHWESTERN UNIVERSITY

Conflicted Rational Animals

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Philosophy

By

Abigail Bruxvoort

EVANSTON, ILLINOIS

December 2021

Abstract

To be tempted is to be conflicted, but the conflict is not one of oscillation between two good options. Rather, it is normally easy to act on temptation and difficult to act against it. But this is puzzling, because unless temptation is a force that acts *on* us, it's not clear why our motivation would sabotage itself in this way. In response to this, I offer a theory of inclinations as provisional decisions of the instinctive mind. Inclinations are what would guide our action if we were wholly creatures of instinct, but in the case of human creatures with the ability to question their inclinations and settle practical questions in light of their reasons, inclinations are also reason-laden, an inchoate grasp of the value of acting as we are inclined to act. This means that temptation is not an assault on rational nature in the form of an "urge" or the "flesh," but an inevitable conflict that arises in a creature with a dual human-animal nature. Our task, on my view, is not to beat down temptation with a strong will but rather to cultivate our inclinations across time.

Acknowledgements

As an undergrad, one of my professors told me that I ought to choose my graduate program carefully, since my graduate program would have more influence than anything else over the kind of philosopher I would become and the kinds of views I would come to regard as true. I think this point of advice was for the most part true, but in writing (and then reading) this dissertation I have been struck by the presence of my undergraduate teachers in it. My suspicion of Sartrean interpretations of Kant and Korsgaard and preference for Aristotle instead comes from Ben Lipscomb, and my interest in philosophy of mind and the ways in which we both are and are not rational was first fostered by Carlton Fisher. I am grateful to have had such fine first philosophy teachers and am happy to find that their influence still lingers.

For feedback on chapter drafts, I want to thank the members of NUMAP, Kyla Ebels-Duggan, and Tamar Schapiro. NUMAP was subject to many iterations of my chapters on responsibility and freedom in particular and I am grateful to have had such a safe and helpful forum for working out early versions of the ideas presented here.

Tamar graciously joined my dissertation committee relatively late in the project and was magnanimous in the face of my criticism of her work and helpful in her suggestions about where to go next. In addition to her feedback, this dissertation is deeply indebted to the questions raised in her work on desires and inclinations.

Thank you to Richard Kraut for serving on my committee. Somewhat to my surprise, the view I've ended up defending in this dissertation is relatively Aristotelian in key respects and so Richard, I hope I have not done Aristotle too much disrespect. Thank you for your support of

moral philosophy even outside of ancient Greece and for being so gracious about some of the twists and turns of my dissertation writing years.

I am lucky to have ended up at Northwestern for many reasons, Kyla being chief among them. Kyla's name shows up relatively infrequently in the body and footnotes of the dissertation, but anyone who knows her work will see that her influence is all over the dissertation, both in what is said and in the way my questions are framed. But beyond philosophical influence, I owe Kyla huge thanks for encouragement, support and flexibility both during my time in residence at Northwestern and my time dissertation writing from afar. I am so deeply grateful to have had a truly top-notch advisor.

I entered graduate school resigned to the idea that I would have terse and competitive relationships with my fellow graduate students. I was delighted to be almost immediately proved wrong in this expectation. Any big writing project requires a good deal of moral support and encouragement, and so for unceasing support and cheering on in addition to good advice, I want to thank my philosofriends: Annie Corbitt, Blaze Marpet, Breno Guimarães Santos, Gretchen Ellefson, Hao Liang, Kyle Scott, and my nemesis-friend Josh Kissel. In addition, thanks to Chelsea Egbert both for friendship and her kind assistance in proofreading my work.

In writing about animal agency in the first few chapters of this dissertation, I have tried to carefully attend to the animals in my life in effort to respect and capture both their alterity and familiarity. For inspiration regarding animal minds and animal action, as well as their constant friendly presence on lonely writing days, I want to thank my cats. They will no doubt regard this dissertation as *to-be-slept-on* or in the case of one, *to-be-shredded*, but living affectionately

alongside members of another species has influenced my thinking about my own human-animal agency and I do not take their presence in our home lightly.

To my family: thank you for not asking me about my dissertation. To my sister: thank you for merciless and affectionately mocking philosophers lest I become too vain. To my mother: thank you for being the sort of parent who is perfectly happy to have a child get a PhD in philosophy.

To my husband, Will Wilson: it's your turn next. I hope I will offer you half as much encouragement and unflagging confidence in your abilities as you have offered me when the time comes for you to write your own dissertation. You have buoyed my mood and talked me up when I needed it most in addition to putting up with my regular demands for silence. Thank you.

This dissertation would not have been possible without many people, but two remain to be singled out. The idea to write about temptation and in particular the difficulty of acting against tempting desires came from Steve White. Steve made me rewrite my seminar paper and this dissertation was borne out of that task. Beyond the original inspiration for the project, the chapters on resolution and responsibility in particular benefitted from his feedback. It is very bittersweet to complete the project without him. This project is dedicated in part to him with gratitude and affection.

The dissertation is also dedicated to my grandfather, Dennis Ribbens, who was the first person to suggest I ought to study philosophy. His intuition was right on that front, and my relationship with him was improved by the increased tolerance for debate occasioned by my studying philosophy. So to my slightly terrifying but loving and proud grandpa: this dissertation is dedicated to you too.

Contents

Introduction

Every human who lives long enough will experience inner conflict. In this dissertation I will consider a particular form of inner conflict: temptation. Even the most strong-willed among us know what it is like to be tempted to give up on a plan or goal. Sometimes tempting desires are high stakes: a "clean" politician is tempted to accept a bribe, or one partner in a marriage desires to cheat on the other. Others are mundane, like wanting to sleep in rather than get up and go to the gym, or do pretty much anything other than write one's dissertation.

Temptation often has religious connotations, at least in the West. In part, the religious connotations of temptation track the idea that we are tempted by a tempter, i.e. a devil. Furthermore, for much of its history, Christianity has had an unhappy and contentious relationship with embodiment, and thus there has been (and still is, in some circles) the tendency to think of temptation as essentially embodied: the flesh is the source of temptation and directs us towards sin.[1] Accordingly, on such views, becoming virtuous or saintly requires overcoming or totally controlling the body.

My own account of temptation is far less fraught. As I understand it, temptation is an inner conflict: more specifically, a conflict between a desire and a judgment. However, although I certainly don't think that temptation is "fleshly" in the sense gestured towards above, I do think

[1] For this history, we probably have Augustine's interpretation and emphasis on various passages from the letters of Paul to thank. In *The City of God*, for instance, Augustine regards lust as an "unwanted intruder" (Translated by Henry Bettenson, New York: Penguin Books, 2003, 577) and his autobiography is preoccupied with his own sexual transgressions, strongly emphasizing the lust and seduction that overcame him in his youth (see e.g. Books II-III in *Confessions*, translated by Henry Chadwick, Oxford: Oxford University Press, 1991, 24-51). For more on early Christian interpretations of Pauline teachings on the body and Augustine's view as the body, see Peter Brown's *The Body and Society* (New York: Columbia University Press, 1988, lx, 387-427).

our capacity to be tempted is fundamentally related to our animal nature, or more precisely, our dual human-animal nature.

Defining Temptation

In everyday language, temptation is used in a huge range of contexts. Sometimes "tempted" just means we want something or that it sounds appealing, as when we're tempted to buy a bold print instead of the classic grey shirt. We also use temptation to describe wanting to give up on commitments or goals: we are tempted to cheat on diets and give up our workout regiments. Sometimes experiencing temptation can be a good thing, like when we're tempted to take in and care for a scrawny homeless kitten. But sometimes temptation is not so good, if for instance we're tempted to steal. We mostly use temptation in the practical realm, but we can be tempted to certain beliefs too: we might be tempted to believe that an unfortunate event in the life of our nemesis was warranted desert for their irritating personality even as we realize that in truth it was just bad luck.

What best holds these wide range of experiences together is conflict. However, although conflict may be the genus of temptation, conflict has many species. For instance, we might experience conflict between two desires: imagine I want to eat pizza and I want to eat a burger, and I know I'm not hungry enough to eat both. But we wouldn't call these conflicting desires a case of temptation. Similarly, in the course of deliberating about what to do, we are often conflicted about what the best course of action will be. Imagine for instance feeling conflicted over what paint color will look best in one's kitchen. Here, the conflict is between two possible courses of action, or two judgments about what would look best, and again, we wouldn't say that trying to decide between pale green or grey for the kitchen is an experience of temptation.

For the purposes of this dissertation, I will define temptation as a conflict between a desire and a judgment, most often a judgment about what to do. However, it's important to note that I do not intend this definition to provide necessary and sufficient conditions. Rather, I am trying to define temptation according to the paradigmatic experience underlying our concept of temptation, and I think that experience is the experience of a desire in conflict with a judgment.

Outline of the Dissertation

In the first chapter, I consider an important feature of temptation: when we are tempted, it is normally easy to do that which we are tempted to do and difficult to act against our tempting desire. But this creates a problem: how can temptation both be ours (for surely it is not a force acting on us from without) and yet something in the face of which we seem to be passive, something with respect to which we are passive, and in the face of which it is difficult to act? In order to answer this question and work towards giving an account of "motivational pressure," the pressure that makes it easier to act on temptation rather than not, I turn to Tamar Schapiro's work on inclination and in particular examine her idea that inclination is the product of our instinctive mind, a remnant of our animal agency.

I take on a great deal of Schapiro's account, but ultimately, she and I part ways when it comes to the question of whether inclinations present us with reasons and the kind of failure that occurs when we act impulsively or instinctively on an inclination. Thus, in the second chapter I turn towards giving my own amended version of her account. I refer to my account of inclinations as a "provisional decision" view. Although inclinations are not decisions proper because to be inclined is not the same as intending, I place inclinations in the category of decisions broadly construed to flag that inclinations are oriented towards action and they would

direct our action if we were wholly instinctual creatures. “Provisional” serves to capture that inclinations are not yet decisions on my view. We often regard inclinations as provisional for the substantive reason that they are prone to error, but the deeper point is that inclinations are provisional because they are an imperative to act, not a settling of a practical question.

In the third chapter I pivot to a more practical question: how can we resist temptation? In particular, in the third chapter I ask whether there is a specifically rational way of resisting temptation, i.e., a method available only to rational agents that we should expect to be especially successfully in ideally rational persons. One obvious candidate is that we rationally resist temptation by making resolutions, which “hold off” temptation or make it so that tempting desires have no “sticking power.” However, on second look it’s not clear why resolutions should be of any special use in resisting temptation: if our first-order intentions and desires readily change in the face of tempting desires, why should we expect second-order intentions or desires (those constitutive of resolutions) to withstand temptation? Thus, I argue that resisting temptation requires a more encompassing approach, and that ultimately, the uniquely rational way to resist temptation is to become the sort of person who is not tempted. This way of putting it is rather too strong, but the underlying point is that resisting temptation requires us to have the right sort of inclinations in the first place and not only “fight” our inclinations after the fact.

In the fourth chapter I discuss the nature of our relationship with our inclinations and especially tempting inclinations. Are they robustly attributable to us? Can we be morally evaluated in light of them? After engaging with Angela Smith’s work on responsibility for attitudes, I conclude that we are not responsible for inclinations in Smith’s sense because we cannot justify inclinations in the robust sense required for answerability (and thus responsibility)

on Smith's account. However, I nonetheless argue that we should not conclude in light of this that inclinations are simply states that happen to us, like nausea or headaches. Inclinations are intentional and as such can be assessed for fit. Furthermore, we can cultivate or manage our inclinations to integrate them into our character and agency. Thus, I argue that we have a forward-looking responsibility to manage our inclinations. This means that any particular inclination might say effectively nothing about who we are, but we nonetheless have the responsibility to relate to that inclination in the right way so that it does not become part of who we are.

In the fifth and final chapter I shift to the social realm and the practice of tempting others. One way we might try to influence another person is by tempting them. In this chapter I ask whether it's all-things-considered worse to influence someone by tempting them as opposed to influencing them by other more "rational" means. I argue that tempting another person undermines their will in at least one and possibly two ways, and thus is to be avoided unless specifically justified. One, tempting another undermines their will and disrespects their right to choose for themselves insofar as it makes it more difficult for them to act as they intend. Two, tempting another often undermine a person's will by making tempting reasons appear very salient, thus making it difficult for them to clearly think through how they ought to act.

The underlying theme of the dissertation is that temptation is a lens into the kind of agency or rationality we have. On the one hand, my view understands our "animal" nature in a relatively rational way. I don't see our animal nature as a brute and fleshly assault on some citadel of reason that is who we truly are. Rather, I see our animal nature as an instinctive way of representing and responding to the world that is obviously effective as a mode of action for non-

human animals and is an important aspect of human agency as well. Furthermore, being a human-animal agent doesn't mean having two different modes of agency that function separately and you can alternate between them. Our rationality influences our instinct, shaping what we see as worth doing and our inchoate impressions of why an act is worth doing. Thus my overall attitude towards temptation is relatively neutral: more often than not, it's a bad idea to act as one is tempted. But neither do I see temptation as a sign of weakness, corruption, or immorality. It is in some cases a sign of immaturity, but in other cases it's just a sign that our agency continues to be animal as well as human, and thus is prone to a certain unique kind of conflict.

On the other hand, philosophers sometimes slip into thinking that because our rational nature can and should cultivate our instinctive nature, animality is something that adult human beings leave behind, and the adult's true self is defined by the rational ideals to which she is subject. Views of this variety tend to treat temptation as something that assails us from without. If temptation assails from without, we can blame tempting desires on our animal nature, or demons, or advertisers, and so on and so forth, and our true identity as rational creatures is left intact. But in fact, temptation is an *inner* conflict, and bringing the conflict of temptation inward makes the conflict deeper. There is no doubt that external conflicts can be thorny, but when a conflict is external, someone or something else is to blame. The source of the problem lies outside the self. But on my view, the problem of temptation – insofar as it is a problem – is an inner one, and thus the conflict of temptation is not a conflict between a rational agent and her desires but rather a conflict *within* rational agency, because rational agency in a human is from the start incomplete and conflicted. Even when cultivated, as it should be, human rationality remains animal.

Temptation and Weakness of Will

Is a dissertation about temptation really just a dissertation about weakness of will? I think not. Temptation and weakness of will are clearly related, since presumably many instances of weakness of will begin in temptation. Depending on how we define weakness of will, it may be essential to weakness of will that it involves succumbing to temptation as opposed to "coldly" acting against one's better judgment. [2] However, it's possible to be tempted and not succumb to temptation, and for this reason I think we should not collapse discussions of temptation into discussions of weakness of will. To do so would be to risk failing to give temptation proper shrift, and I think it is more productive to begin with temptation, the state prior to weakness of will, before turning to weakness of will in its own right. Furthermore, weakness of will has been hashed over relatively thoroughly in the philosophical literature. For the time being that discussion stands less in need of additional voices than do questions about temptation.[3] Thus weakness of will plays only a peripheral role in the body of the dissertation.

My view of weak-willed action follows that of Pamela Hieronymi, who argues that when we act weak-willed we are not strictly irrational (judging that it would be good to φ and simultaneously judging that it is not good to φ) but rather that when we are weak-willed we are settling two practical questions, one theoretical, about what we ought to do, and one practical, about what to do.[4] This is not to say that I think weak-willed action is unproblematic, but I think

---

[2] Davidson famously denies this. For example, "We succumb to temptation with calm; there are also plenty of cases where we act against our better judgment and which cannot be described as succumbing to temptation" ("How is Weakness of Will Possible?" in *Essays on Actions and Events*, Oxford: Clarendon Press, 2011, 29. Shortly after he offers his famous tooth-brushing example which purports to show that we can be weak-willed without succumbing to an inclination (30). I find it very odd to describe the tooth-brushing case as an instance of weak-willed action.

[3] To paraphrase a point made by Myles Burnyeat in "Aristotle on learning to be Good" – the real question is not how we can act akratically but how anyone is *not* akratic (in *Essays on Aristotle's Ethics*, edited by Amélie Rorty, Berkeley: University of California Press, 1980, 85-6).

[4] "It is quite possible (though no doubt irrational) to conclude that you have reason to φ without also deciding to φ,

the problem with weak-willed action is substantive: it is wrong insofar as the theoretical judgment about what the agent had best reason to do was correct and the agent settled the practical question poorly.

Temptation and Virtue

Many people assume that temptation is essentially immoral. Certainly temptation has some of this connotation in everyday language: often when I tell people that my philosophical research is on temptation, they assume that my research is about the taboo and are disappointed to learn that my work is not about all things naughty and sexy. So should I add to my definition the further idea that temptation is essentially immoral? Or, in virtue theoretic terms, is it incompatible with virtue to be tempted?

Temptation features prominently in the literature on this question. The debate over temptation and virtue occurs most often in Aristotelian contexts, and accordingly it is most often put in terms of continence. Consider, for instance, the following from John McDowell: "If someone needs to overcome an inclination to act otherwise, in getting himself to act as, say, temperance or courage demand, then he shows not virtue but (mere) continence."[5] Similarly, Rosalind Hursthouse writes,

> The continent character is the one who, typically, knowing what she should do, does it contrary to her desires, and the fully virtuous character is the one who, typically, knowing what she should do, does it, desiring to do it. Her desires are in 'complete harmony' with her reason. … So Aristotle draws a distinction between two sorts of people – the continent or self-controlled, and the fully virtuous.[6]

---

or to decide to φ while acknowledging that you have most reason not to φ. This happens when the reasons one acknowledges, when reasoning theoretically, are not reasons that one employs, when reasoning practically" ("The Will as Reason," *Philosophical Perspectives* 23, no. 1 (2009): 207).

[5] "Virtue and Reason," *The Monist* 62, no. 3 (1979): 334.

[6] *On Virtue Ethics*, Oxford: Oxford University Press, 1999, 92-3. Consider also the following passage from Gregory Trianosky: There is a familiar distinction between two sorts of morally good people. The first always does his duty, or more, without regret, and without even being tempted to do anything else. The second is highly self-controlled. He too always does what is right, whether this is required or perhaps even beyond duty; but he must constantly exert

Both McDowell and Hursthouse, then, ascribe to what I shall call (following Karen Stohr) the "harmony thesis."[7] The harmony thesis holds that the ideal or fully virtuous agent feels no conflict or stress at performing the right action.[8] Her feelings are in harmony with what she does.

Although initially quite plausible, the harmony thesis is not without its critics. In Stohr's article "Moral Cacophony," for instance, she argues that in some cases virtue actually requires us to feel some conflict, stress or distress. For instance, Stohr considers a case in which the owner of a small company concludes that her company is no longer financially viable in its current form and that she will have to lay off some of her employees.[9] We are to imagine further that the owner is virtuous: "She will, of course, have applied all principles of fairness in determining whom to fire, provided them with the best severance package that she can afford, and thought very carefully about the best way to break the news."[10] But in spite of this, Stohr says that we would expect the owner to feel very bad about this situation – she might be anxious or sad before and during the layoffs, and afterwards she is likely to worry about the wellbeing of her former employees. Stohr finally concludes that

> Such cases are actually quite common. There are many actions that, although required, seem to be of a sort that good people should find difficult. People should find it difficult to deliver bad news to their friends. Parents should find it hard to punish their children. Teachers should find it hard to give low grades to students who are genuinely trying to do well.[11]

---

himself in deliberation and in choice to subjugate unruly, contrary inclinations. Following Aristotle, the first of these two may be called temperate, and the second *continent (*"Rightly Ordered Appetites," *American Philosophical Quarterly* 25, no. 1 (1988): 1).

[7] "Moral Cacophony: When Continence is a Virtue," *The Journal of Ethics* 7, no. 3 (2003): 340.

[8] Although the harmony thesis isn't specific to temptation, I think temptation is one of a variety of conflicting and stressful responses one might have.

[9] "Moral Cacophany: When Continence is a Virtue," 342-3.

[10] *Ibid*, 343.

[11] *Ibid*, 344

All of this, then, is evidence against the harmony thesis. For my part, I find this line of thought very persuasive. It seems clear to me that as long as we live in a less than ideal world, we will have to do things that are necessary – and even morally right – that are nonetheless painful and difficult to do. Given this, I think it is sometimes appropriate to show continence or experience temptation.

Furthermore, I think there are cases in which agents have multiple and incompatible courses of action available to them. In these cases, I think it is natural for agents to continue to be tempted by an option they had to forgo. Think, for example, of someone making a choice between two equally worthwhile career paths. It seems to me that it's normal and even healthy to continue to be somewhat tempted by the option one ends up rejecting. Of course, lingering temptation could be a sign of having chosen the wrong path, or a sign of lacking commitment or focus. But it can just as easily be a reflection of the fact that humans are finite and can't do everything they might like to do. And in situations of the latter sort, I don't think it's necessarily a sign of lacking virtue to continue to be tempted by the path you've passed over.

Given all of this, for the purposes of this dissertation, I will assume a non-moralized account of temptation. Although it is true in some cases (or perhaps even the majority of cases) that it is bad to be tempted, whether because the tempting desire is essentially wrong (e.g. a tempting desire to harm for fun) or merely imprudent in the moment (e.g. a desire to watch YouTube videos instead of doing one's work), in at least some instances it is good or virtuous thing to be tempted. [12]

[12] There is a related discussion about the goodness of temptation in the context of inverse akrasia. The case of Huck Finn has been a favorite example here, since Huck is tempted to defend and protect his friend Jim, a runaway slave, while Huck judges that the morally required thing is for him to turn him in. The case is most notably discussed in several articles: see for instance "Praise, Blame and the Whole Self" (Nomy Arpaly & Timothy Schroeder,

1. The Pressure to Act

Say that I am not a runner, but I want to become a runner. So I sign up to run a half-marathon and write out a carefully structured training plan for myself. However, I can see a potential flaw in my perfect plan: I know that I'm going to want to give up training. What should I do? The question, of course, needs specifying. I might answer the question by reading pop-psychology books about habit-formation and learn tips like "find an accountability buddy" or "make it easy to run in the morning by setting out your breakfast and running clothes the night before." More philosophically speaking, I might wonder if the anticipated difficulty of this endeavor is reason for me to believe that I'll fail to complete my training and in turn fail to run the race.[13] My question in this chapter, however, is how I should understand and in turn relate to the difficulty of completing my training.

Furthermore, say that in this situation I'm not just going to want to give up because I'm bored: I expect that I will be *tempted* to give up on my training plan. Or, if the prediction about the future seems too uncertain (maybe you're an optimist and you think I should hold out hope for discovering that I absolutely adore running), suppose that right now I'm tempted to give up on the training plan I've undertaken. Temptation is paradigmatically hard to resist. Regardless of the object of temptation, when we are being tempted it is normally easy for us to do the thing we are tempted to do and difficult to do whatever we judge we ought to be doing instead. So if I am tempted to give up on my training plan, I will find it in some sense difficult to continue my

---

Philosophical Studies 93, no. 2 (1999): 161-188) and "Moral Worth" (*Journal of Philosophy* 99, no. 5 (2002): 223-245). Again, this strikes me as a case in which we would say that it's a *good* thing to be tempted – Huck's temptation is a good thing. Thus, I take this kind of case to be further evidence against the harmony thesis.

[13] This is the question animating Berislav Marušić's book *Evidence and Agency*, Oxford: Oxford University Press, 2015.

training but easy to give it up. In fact, the difficulty of resisting is perhaps one of the defining features of temptation: if I've resolved to do my training run, and the thought of something else (going to the grocery store) neutrally floats through my mind, I'm not tempted. At best I'm just considering an alternate action. If on the other hand I *want* to do something else, perhaps stay on the couch and open Netflix, there is a sense in which it would be easy for me to stay on the couch and to some degree more difficult to go on my planned run. In short, temptation exerts on us what I will call "motivational pressure" or "motivational force." To clarify the question posed just above, my question in this chapter is how we should understand and relate to the motivational pressure temptation exerts on us.

1.1 MOTIVATION AND RATIONALIST THEORIES OF DESIRE

Sometimes in everyday language we speak of temptation as if tempting desire is an urge so strong there is no resisting it. We appeal to such concepts when we say things like "I couldn't help it!" or "I just couldn't resist." The underlying concept of desire at play here is what I will refer to as a "brute force" view of desire.[14] Most simply, brute force views hold that desire is a force that pushes us around. Good arguments against brute force views of desire have been well-rehearsed elsewhere, and so I will not lay out the problems with this view in great detail here.[15]

[14] Another label for this kind of view is a "hydraulic view" (although the hydraulic view might extend beyond desire.) Wallace escribes the view that human actions are a result of "psychological forces [within us] with respect to which we as agents are essentially passive" as a "hydraulic" view of action ("Three Conceptions of Rational Agency," in *Normativity and the Will: Selected Papers on Moral Psychology and Practical Reason,* Oxford: Oxford University Press, 2006, 55).

[15] See e.g. Chapter 2 in Tamar Schapiro's *Feeling Like It,* Oxford: Oxford University Press, 2021 and Kyla Ebels-Duggan, "Bad Debt: The Kantian Inheritance of Humean Desire," forthcoming in *The Idea of Freedom: New Essays on the Kantian Theory of Freedom, e*dited by Dai Heide and Evan Tiffany.

However, before moving forward, I will briefly summarize what I see as the three primary problems with this view of desire.

1. One, if brute force views are true, then our ability to resist any given desire is contingent, just like our ability to resist a strong current is contingent on the strength of the current. Even the best swimmer is helpless in the face of a very strong current. Although someone might hold that it's correct to say that our ability to resist a given desire is contingent, to embrace this conclusion is to embrace a revisionist theory of desire, since we normally regard desires – and even strong desires – as able to be resisted, at least by the "strong willed."[16]
2. Two, we can act on desires. If I desire to go to the beach, this desire can become my action. But if desires are like brute forces, then it's not clear how we can act *on* them. We can act in light of them, or around them (as we might act in light of a strong wind), but we cannot act on them.[17]
3. Three, desires standardly serve to make our actions intelligible. When I cite a desire as my reason for acting, this desire at least gestures towards what I saw as good about so acting (even if it leaves more to be said). But if on the other hand I explain an apparent action by telling you it was just the result of a force within me – perhaps a reflex – then I am not making my action intelligible but rather rejecting the idea that what I was doing was acting.[18]

Considering these problems with brute force type views of desire, we might be tempted to take a sharp turn in a different direction when it comes to desires, to what I will call a "rationalist" theory of desire.[19] The common theme of such views is the idea that desire is some kind of perception of goodness. Although there are various proponents of rationalist views, I will take Scanlon as the exemplar in the following.[20]

---

[16] Addiction might appear to be a counterexample to this case. For my purposes, I am setting aside addiction because it seems to me that any philosophical treatment of addiction must be empirically informed. However, I will note that it is at least highly controversial how to overcome addiction. At risk of simplifying, there is a debate between those who think addiction is best treated with medication and those who think addiction must also engage the agency of the addict. For a defense of agency in the face of addiction and the view that addiction is not compulsive, see "Psychopathology and the Ability to Do Otherwise," *Philosophy and Phenomenological Research* 90, no. 1 (2015): 135-163.

[17] *Feeling Like It*, 48-9.

[18] See for instance Anscombe's *Intention*, Cambridge, MA: Harvard University Press, 1957.

[19] In *Feeling Like It*, Schapiro uses the label "practical thinking views" instead. In her earlier paper, "The Nature of Inclination," she uses "rationalist" as well (*Ethics* 119, no. 2 (2009), 229-256).

[20] For instance, Dennis Stampe maintains, "The authority of desire is the authority of perception" ("The Authority of Desire," *Philosophical Review* 96, no. 2 (1987): 362). Similarly, Tal Brewer writes, "… desires consist at least partly in representations of reasons or values. This does not mean that desires are to be conflated with beliefs about

1.1.1 Scanlon has a "directed attention view of desire." He writes, "A person has a desire in the directed-attention sense that P if the thought of P keeps occurring to him or her in a favorable light, that is to say, if the person's attention is directed insistently toward considerations that present themselves as counting in favor of P."[21] This doesn't mean, however, that desires must all arise from deliberation about what would count in favor of something. Rather, Scanlon allows that desires are "unreflective" and can "assail us unbidden" and "conflict with our considered judgment of what we have reason to do."[22] This is possible because desires arise from our broader "tendencies" to see certain kinds of considerations as reasons, and these tendencies are not wholly under our control. [23] What's most important for my purposes in this chapter, however, is Scanlon's position on desire and motivation. He writes,

> … we should not take 'desires' to be a special source of motivation, independent of our seeing things as reasons. … 'Desire in the directed-attention sense' comes much closer to capturing the commonsense notion of desire, but this notion doubly fails to capture a unique or independent source of motivation. First, it is not the case that whenever a person is moved to act he or she has a desire in this sense … Second, when a person *does* have a desire in the directed-attention sense and acts accordingly, what supplies the motive for the action is the agent's perception of some consideration as a reason, not some additional element of 'desire.'[24]

In other words, Scanlon's account relies on an analogy between perception and desire. Although other theories in the rationalist tradition make more of the analogy than he does, and even though

---

reasons or values. … desires are seemings of goodness or of reasons for action. … To have an occurent desire is to find oneself attending to (or imagining) certain possible activities or courses of action as good" (*The Retrieval of Ethics*, Oxford: Oxford University Press, 2009, 25) and "[Desire] involves an attentiveness to, or tendency to dwell on, some sort of goodness or value in those things that they can incline us to pursue or promote." *(The Retrieval of Ethics*, 29). Or: "desires are best understood as consisting not just partly but wholly in appearances of reasons or values" (*The Retrieval of Ethics*, 34). Consider also "The desire that P is P's seeming good" from Graham Oddie's *Value, Reality, and Desire (*Oxford: Oxford University Press, 2005, 42).

[21] *What We Owe to Each Other*, Cambridge, MA: Harvard University Press, 1998, 39.

[22] *Ibid*.

[23] *Ibid*, 40.

[24] *Ibid*, 40-41.

the perceptual analogy is not at the forefront of Scanlon's account, he nonetheless relies on the idea of "attention being directed" and desire as a form of "seeing" vs. "judging."[25]

1.1.2 Scanlon's theory is well-equipped to avoid the problems of a brute force view of desires. But is this view adequate to capture the motivational pressure characteristic of temptation? For Scanlon, desires are not a special source of motivation.[26] Rather, on his view, what motivates a given action is the agent taking something to be a reason to act. When it comes to temptation, then, this suggests that Scanlon would be committed to the view that we simply favorably "see" reasons via the tempting desire that conflict with our considered judgment(s) about the reasons we do in fact have.[27] Some thinkers have tried to understand motivational force (and furthermore the motivational force of temptation) in this way. Jay Wallace appears to be one such example. For instance, he writes that temptation is "a psychological condition that facilitates the choice of an action the agent believes ill-advised, by *directing the agent's thoughts* onto the alleged attractions to be gained through that action."[28]

Scanlon's notion of "directed attention" is astute, since our attention typically is persistently directed towards the object of our desire when we are tempted. When I am tempted to eat more salty chips, my mind keeps returning to the chips and the favorable thought of eating them. Still, I think this is insufficient to capture the phenomenon of temptation. An obsessive daydream might lead me to think persistently and favorably about an object without making it difficult for me to follow through on my resolutions or without moving me to do any particular

[25] See e.g. *What We Owe to Each Other*, 40.
[26] *Ibid*, 40.
[27] *Ibid*, 40, 43.
[28] "Addiction as Defect of the Will: Some Philosophical Reflections," *Law and Philosophy* 18, no. 6 (1999): 648. Emphasis added.

thing. Sports fans, when watching a close game, clearly direct their attention to the game, and it is true that their focused attention will make it difficult for them to do other things. It might be difficult, for instance, to notice a smoky smell from the stove in the other room during the last few seconds of a race, or it might be difficult to tear oneself away from a riveting moment in order to get up and care for a fussy child. Still, I don't think this sense of "difficulty" captures the motivational pressure that's paradigmatic of temptation. These are chiefly cases of distraction, whereas temptation is more than a mere distraction. Richard Holton rejects Scanlon's account for similar reasons, writing that "In a guilty state of mind, the things that count in favor of the virtuous but forsaken course of action may come insistently to my attention; but that does not mean that I want to take it. What is missing in Scanlon's characterization is the idea that desire *pulls* me to a course of action …."[29]

However, the deeper concern with Scanlon's account and any accounts of desire that rely implicitly or explicitly on perceptual metaphors is that they fail to clarify our relationship to our perceptions and the force they exercise on us. On the face of it we seem to have two options. One, we can say that desire *qua* seeing-of-reasons is the exercise of our own practical agency. But in this case, it's not clear why our own agency, having settled the question about what to do in the form of a decision or resolution, would then sabotage this decision in the form of tempting desire even in a situation where on reflection the agent determines that their reasons still on balance support their initial decision. Or, two, we can say that desire qua seeing-of-reasons is the exercise of something else, in which case it acts *on* our agency. The latter view leads to the problems of the brute force view outlined above. Ultimately, I will adopt a version of the former

[29] *Willing, Wanting, Waiting*, Oxford: Oxford University Press, 2009, 102. Unfortunately, I think Holton swings too far in the other direction, going on to characterize tempting desires as urges or even cravings.

view and defend a view on which tempting desire has motivational pressure but is not straightforwardly an instance of "agency sabotaging itself."

1.2 SCHAPIRO'S DIAGNOSIS

Although I have been focusing thus far on cases in which I decide to φ and then am tempted to do something else such that it becomes difficult to φ, my question about the motivational pressure of temptation is an instance of a broader question about the nature of inclination and the kind of motivational pressure inclinations exercise on us. When I am tempted to give up on my half-marathon training plan, my temptation is what Tamar Schapiro calls an inclination. (What exactly I am inclined to do in this situation just depends on what I feel like doing in lieu of running.) By inclination, Schapiro means "a feeling that motivates you to do something, while still leaving you free to act on it or not. To have an inclination to φ is to *feel like* φ-ing, prior to and independent of having decided to φ."[30] However, Schapiro's understanding of inclination is a narrower concept than desire in general, since it's possible to desire something without feeling like doing it. It's also possible to desire something in what Schapiro calls a "placeholder sense," i.e. to desire something without feeling like acting any particular way.[31] We often appeal to desire in the placeholder sense in order to attribute action to

[30] *Feeling Like It*, 10.

[31] See "What are Theories of Desire Theories of?" *Analytic Philosophy* 55, no. 2 (2014): 136. In addition to placeholder desires, it also seems that we can have desires that are not primarily directed towards any particular action. For instance, if I desire that a particular sports team win their game, my desire in itself is not a desire to act in any particular way, since the sports team winning the game is not a state of affairs that I could possible bring about: there's no meaningful action I could take that would fulfill this desire, although there may be a host of related actions that are borne out of or spill over from my desire (example drawn from Nomy Arpaly and Timothy Schroeder's *In Praise of Desire*, Oxford: Oxford University Press, 2014, 95).

the agent as opposed to some other source. So for instance, an agent's placeholder desire makes her going to the grocery store intelligible as her own action as opposed to behavior borne out of coercion, etc. But of course it's possible to go to the grocery store and for the action to be genuinely your own without feeling like it in the specific sense of inclination. Thus, in the following I will use inclination in Schapiro's sense when I wish to refer to the state of feeling like doing something without being determined to so act, although when my discussion pertains to desire in a more general sense I will revert to using "desire" instead.

In her work on inclination Schapiro supplies three constraints on any adequate theory of inclination:

1. Non-voluntariness: you cannot be inclined at will. You can determine yourself to φ, and thereby φ, but you cannot determine yourself to be inclined to φ, and thereby be inclined to φ. You can, however, do something we call 'cultivating' your inclinations over time.
2. Deliberative role: being inclined to φ involves being faced with a directive, something that purports to tell you, in a provisional way, what to do.
3. Asymmetric pressure: insofar as you are inclined to φ, then other things being equal (i.e., in the absence of additional motives), it is easier for you to go along with your inclination than not to. The relevant sense of 'easier' is the correlate of the effort we call 'will power.' Will power, whatever it is, is an effort to oppose this asymmetric pressure.[32]

I think these constraints are correct. In this chapter, my focus is on the third: what Schapiro refers to as "asymmetric pressure" is what I have been calling the motivational pressure of tempting desire.

1.2.1 Rationalist views like Scanlon's are highly responsive to the second constraint in particular. Scanlon also seems responsive to something like the non-voluntariness constraint: he uses passive language in speaking about desires and acknowledges their recalcitrance in the face

[32] *Feeling Like It*, 31-2.

of judgment, which captures something akin to Schapiro's non-voluntariness constraint. Just like we cannot perceive at will, so we cannot desire at will. Both perception and desire arise passively. But what about the asymmetric pressure constraint? Although I think Scanlon's view flounders on the third, as discussed above, here too there seem to be some resources available to Scanlon from the perceptual analogy: normally when we perceive, it is easier to go along with and believe our perception than not.

Arguably Schapiro's primary objection to Scanlon's view and others like it is that they are committed to motivational monism, "the view that *you* are the source of your desiring, in the same sense that *you* are the source of your acting."[33] The rationalist about desire will see this as an advantage: in saying that we are ourselves the source of our desiring, we avoid brute force views on which desire is simply something that happens to us that we must work around, like a stomachache or a thunderstorm. The problem Schapiro sees is that it's hard to reconcile motivational monism with the passivity and recalcitrance of desires. As she puts it,

> How can you be passive in relation to your own practical thinking, if you are the source of that thinking? The practical thinking theorist might say: because the mode of thinking you are engaged in when you are desiring is, or is like, perception, rather than judgment. Ok, but then is perceiving something that happens to you, or is it something you do? In perceiving, is your mind just being pushed around, as if by a brute, external force that insistently directs your attention, independent of your free assessment of the weight of the evidence, or of the practical considerations? If that is the case, then regardless of where your mind ends up, the resulting judgment is not your doing. … Your mind has simply been overpowered. Suppose, alternatively, that perceiving is your doing. In perceiving, you are taking something to be a reason, albeit in a provisional or less-than-fully-committed way, on the basis of your assessment of the evidence, or of the practical considerations. But if, in perceiving, you are the one directing yourself to this provisional or less-than-fully-committed judgment … then why is your perception, provisional as it may be, recalcitrant to deliberate correction, in a way that your non-perceptual judgment is not?[34]

[33] *Feeling Like It*, 83.
[34] *Ibid*, 84.

This then is the dilemma we face in giving a theory of inclination and tempting desire that makes use of the perceptual metaphor: one the one hand, if perception (and then by analogy, desire) is just the beginning of theoretical and practical reasoning, then why can we not correct our perceptions and desires in the same way we can correct our deliberated judgments and intentions? After all, the force of the perceptual analogy was supposed to be that it can capture how desire figures in practical reasoning but also the passive origins of desire. But both perception and desire are recalcitrant, which means that they sometimes "get in our way" or conflict with our more considered judgments. On the face of it this seems to be at odds with the idea that desire is an expression of our own rational agency: why would we get in our own way with such recalcitrant states? Or more seriously, as Schapiro raises, can we really say that desire and perception are expressions of our own rational agency if we cannot exercise over them the sort of control we exercise over our intentions and beliefs?

On the other hand, if we eschew the idea that desire and perceptions are themselves expressions or output of our practical agency and say that they simply happen to us, then we return to all the problems of brute views of desire.

Of course, one way to resolve this dilemma would be to deny that recalcitrant desires really are recalcitrant. This might seem absurd on the face of it, but there is some plausibility to the view: it would hold that all apparent tempting desires are cases in which we have reason to pursue two (or more) good but incompatible courses of action. So for example, say that my alarm clock goes off but I slept poorly and so I am tempted to turn off the alarm clock and go back to sleep.[35] In this case, there is a sense in which my tempting desire to go back to sleep is

[35] This example is drawn from Angela Smith's "Conflicting Attitudes, Moral Agency, and Conceptions of the Self." *Philosophical Topics* 32, no. 1/2, (2004): 338.

recalcitrant, since it does not conform to my considered judgment that what I really ought to do in this situation is get up so that I won't be late to work. But there's another sense in which the desire to go back to sleep is not recalcitrant at all. It's not an "alienated" desire. My desire to go back to sleep will seem eminently sensible to me, and in fact I will recognize that were the circumstances different (if it were a lazy Saturday instead of a workday, perhaps), I would be perfectly happy to act on the desire and go back to bed. In other words, to use Scanlon's language of judgment-sensitivity, in these circumstances the desire to go back to sleep may not be sensitive to my judgment about what I have reason to do at this point in time, but it certainly is sensitive to my judgment about what is in-general worth desiring insofar as I judge that sleep is good and especially so when one is tired from tossing and turning all night long.

Schapiro ascribes something like this view to Scanlon in some of her earlier work on inclination. There, she writes,

> I consider Scanlon's position to be a version of what I want to call 'extreme rationalism.' … [Extreme rationalism] denies that there are distinctively passive and active motivational capacities, each making a different contribution to action. Instead, extreme rationalism holds that the soul is unity, in the sense that agency involves the exercise of one rational capacity.[36]

On this reading of Scanlon, there is only one motivational element in the Scanlonion soul, and that element only recognizes reasons. Thus, although there can be some kind of conflict in the agent, the conflict can only ever be the conflict between reasons that point to different courses of action. This doesn't mean that all conflict for Scanlon is like a difficult decision between two good things; he does allow for the possibility that we might see as a reason something that we firmly judge to *not* be a reason, at least not a reason to act in the current circumstances. But the

[36] "The Nature of Inclination," 241.

problem is again that this view struggles to account for motivational pressure and our experience of temptation. Why should continuing to see something as a reason while simultaneously judging it not to be a reason affect the ease or difficulty of doing that which we judge we ought to do?

In contrast to Scanlon's motivational monism, motivational dualism seems to have a much easier time accounting for the relevant phenomena. Specifying the exact sense in which our motivation is dual is one of the chief tasks for Schapiro's account, but roughly, motivational dualism on her view is the view that we have two sources of motivation and activity: practical thinking on the one hand and our "inner animal" on the other hand.[37] However, depending on how we cash out the details of motivational dualism, we will continue to face the exact same dilemma articulated above. We still have to tell a story about how both of the sources of motivation in an agent really are the agent, because if we cannot explain how the conflicting sources really are both the agent, then we have just slipped into a version of the brute force view on which desire is simply something that happens to us.

In the following, I will first consider Schapiro's own positive view. This "inner animal" or "instinctive mind" view is, I think, a step forward in understanding the motivational pressure of inclinations including tempting desires. However, I nonetheless think that Schapiro's view goes too far in certain respects, so after summarizing her view I will raise several objections to her view and then move on in the next chapter to argue for my own amended account.

1.3 THE INNER ANIMAL

[37] *Feeling Like It*, 60.

Schapiro summarizes her positive view as follows:

> A first-pass description of my positive view goes like this: when you are inclined but not determined to φ, what you are relating to is a part of you that has already determined itself to φ. You are relating to a part of you that has the structure of an agent, a part of you that is already doing something, in a consciously guided way. But unlike the part of you that decides whether to act on your inclination or not, the part of you that is the source of your inclination has the structure of a creature of instinct. Insofar as it acts, it sees and responds to the world instinctively.[38]

There is a lot here to unpack. The motivational dualism should be clear: Schapiro thinks our agency is carved up into what she calls our "instinctive mind," which is the source of our inclinations and is the part of us that sees and responds to the world instinctively, and on the other hand our deciding mind. The deciding mind is the mind that *responds* to an inclination and "is the kind of thinking through which you decide whether or not to act on your inclination."[39]

The advantage of this view is that it is well-poised to address the concerns raised about whether desire *qua*-perception is the output of your own practical thinking or a force that acts on you. Schapiro's answer is "neither," since your inner animal is you, and it has its own kind of practical thinking. On the other hand, the instinctive thinking of your inner animal is not the same as you own decision on Schapiro's view. She writes,

> When you feel thirsty, your inner animal is already seeing and responding to the world, moving itself through its instinctive practical thinking. Its activity is underway. But you – by which I mean, the part of you that determines yourself to act on or against your inclination – are not thereby determined. You are, instead, in a condition I call 'being drawn out of yourself.'[40]

Schapiro's concept of the "inner animal" or "instinctive mind" draws heavily on Christine Korsgaard's account of animal agency.[41] On Korsgaard's view, the key feature of animal minds

---

[38] *Feeling Like It*, 86.
[39] *Ibid*, 87.
[40] *Ibid*, 86.
[41] Both, however, note that their theories are not intended to be an account of the kind of mind that all animals have.

or animal agency is a teleological view of the world:

> … the world as perceived by the animal is organized around her interests: it consists of the animal's food, her enemies, her potential mates, and, if she is social, of her fellows, her family, flock, tribe, or what have you. To say this is just to say that the animal's representations of things come already loaded with the practical significance of various objects for her. She confronts a world of things that are perceived directly, without calculation or conscious interpretation, as things that are *to-be-fled*, *to-be-cared-for*, and so on.[42]

Because animal minds see the world teleologically, movement that responds to their representations of the world is action attributable to the animal as opposed to mere behavior. Korsgaard acknowledges that there is a spectrum here and it may be difficult in border cases to determine whether a given movement is an automatic response (as when a dog salivates upon smelling food) or an action (as when a dog chases a ball thrown by a human).[43] However, she nonetheless maintains that it is appropriate to speak of animal behavior as action when the movement in question is done in response to the animal's representation of its world and environment. And crucially, an animal's concept or representation of its environment has its own logic. Animals don't perceive the world in brute or arbitrary ways. Rather, as noted in the quote above, their perceptions of the world are *organized by their interests*. And since the animal sees and responds to the world in light of his own instincts, we can say that he is acting when he acts in light of those instincts:

> The principles that govern an animal's movements as he guides himself through his environment – the principles that govern his reactions to his perceptions – are what we may call his instincts. An animal's movements are self-determined when they are governed by his instincts, for when they are governed by his instincts, they spring from his own nature. An animal's instincts then are his will, the laws of his own causality.[44]

Both readily admit that there may be animals whose minds are better characterized in a different way.

[42] *Fellow Creatures*, Oxford: Oxford University Press, 2018, 38. See also Korsgaard's *Self-Constitution*, Oxford: Oxford University Press, 2009, 110-111.

[43] *Self-Constitution*, 111.

[44] *Ibid*, 104

Having seen the basics of Schapiro's view as well as the origins of something like the view in Korsgaard, let's first consider how the instinctive mind or inner animal holds up against the three criteria Schapiro identifies for a theory of inclination.[45]

1.3.1 When it comes to non-voluntariness, it might be tempting to conclude that Schapiro's view is quite similar to Scanlon's. Just as Scanlon meets the non-voluntariness criteria by using "seeing" language and emphasizing the passivity with which desire arises, so too a perceptual framework might seem to do the work for non-voluntariness on Schapiro's view. The instinctive mind sees the world teleologically and makes a proposal on the basis of how it sees the world. However, Schapiro's view in fact goes a step further on this point. She argues that the inclination is not simply our inner animal's thought about a given activity but rather is activity itself, for if the inner animal were simply thinking about action, it would be unclear how inclination is a form of motivation.[46] Furthermore, the view that instinctive minds first think about what to do and then separately initiate action on their instincts fails to capture how instinctive minds actually act. When an instinctive mind has an instinct, it simply acts.[47] This is not to deny what I discussed above about the intelligibility of animal action in light of their representation of the world, but simply to emphasize that instinctive animal minds do not first contemplate the world and then deliberate and act in light of what they see. If we assimilate instinct to thought, then we imply that instinctive animals act by first having a thought and then incorporating or taking up that thought into action. Rather, animals represent the world in practically loaded terms and their

[45] They are, recall: 1.) Non-voluntariness: we can cultivate our inclinations over time, but we cannot determine ourselves to be inclined at will. 2.) Deliberation: inclinations make proposals about what we should do and as such we can take them up for deliberation. 3.) Asymmetric pressure: other things being equal, when we are inclined to φ it is easier to φ rather than not to φ (*Feeling Like It*, 31-2).
[46] *Feeling Like It*, 107.
[47] *Ibid*, 107.

instincts automatically govern how or if they will act in light of those representations.

So on Schapiro's view, the non-voluntariness of inclination stems from the fact that our inclinations are the activity our inner animal. But does this then create a problem for the deliberation constraint? Schapiro herself raises this worry. She writes,

> [Inclinations] tell us what *to do*, given our circumstances. How can your awareness of your inner animal's *doing* amount to a directive telling you what *to do*? This seems like a category mistake. You are thirsty, and in being thirsty, you are aware of your inner animal initiating the activity of drinking. It is already moving itself, under the guidance of its teleological thinking. How does that provide the occasion for you, *qua* deciding self, to deliberate about whether to drink?[48]

Schapiro's answer to this question helpfully highlights her general view about how we relate to our inclinations. As she likes to emphasize, we *are not* our animal nature, rather we *have* an animal nature.[49] Unlike the instinctive mind, whose world is teleological, "The world [the deciding mind] encounters does not have practical necessity built into it. In this sense, the deciding mind is free, and bears the burden of its freedom."[50] Thus we cannot, at least not in the straightforward sense, take up our inclinations and decide to act on them. Rather, when we deliberate about our inclinations, we are considering whether we can take the activity of the inner animal and "give that activity a humanized form."[51]

Using the example of being inclined to eat chocolate cake and the German verbs *essen* (eating as humans eat) and *fressen* (eating as animals eat, often translated "feeding"), Schapiro

---

[48] *Feeling Like It,* 117.
[49] *Ibid,* 103.
[50] *Ibid,* 130. Schapiro expounds upon this point earlier in the book, writing "The instinctive mind cannot conceive of an object disinterestedly, as a thing that exists independent of its relation to that mind's interests. I interpret this to mean that the instinctive mind cannot draw a distinction between a substance, viewed in a way that is not practically salient, and the practically salient attribute that happens to attach to it" (111-12). And then shortly thereafter: "By contrast, the non-instinctive part of your mind, which I will call your 'deciding mind,' distinguishes between object and practically salient property. It sees the cake disinterestedly, as a non-teleological object, while leaving open the question whether that cake warrants, or calls for, or makes appropriate, any particular response, e.g. eating it" (112).
[51] *Feeling Like It,* 132.

expounds upon this point, writing,

> When you are inclined to eat the cake, your inner animal is already *fress-ing*, under the guidance of "chocolatey-to-be-eatenness," while your deciding mind is undetermined. And yet, this provides your deciding mind with the occasion to consider whether it can take this animal activity, and make it into something you can decide to do, from your standpoint as one whose responses are not built into the world you confront. … *Qua* deciding mind, you cannot think ["chocolatey-to-be-eateness"] in the way your inner animal thinks it. But you can take this thought as the raw material out of which to construct your own action under which you find it worthy of your choice.[52]

I am among those who have the worry that this introduces too much duality and renders inclinations that something that happen to you, and I will return to this line of objection shortly. However, for the time being, I want to consider Schapiro's account of how her view meets the asymmetric pressure constraint.

1.3.2 The primary problem with Scanlonian views of desire is their failure to adequately account for asymmetric pressure. The motivating problem of this chapter is the need for rationalist theories of desire to accommodate the difficulty of acting against temptation and the ease of acting on it is. So does Schapiro's view fare any better on this front? On the face of it, it seems intuitive that her view should be well-positioned with respect to asymmetric pressure: the instinctive animal goes for something, and it is easy for the deciding mind to go along with what the instinctive animal is already doing and hard for it to resist.

However, as Schapiro points out, things are more complex and interesting than this. Most concepts of willpower define willpower as a power we need to "oppose the pressure of certain inclinations and not others."[53] Thus we normally refer to exercising will-power when it comes to inclinations on which we think we should not act, and so will-power is almost always invoked in

[52] *Feeling Like It* 132.
[53] *Ibid*, 138.

service of ends like following a diet or work-out plan. It feels odd to speak of exercising will-power when it comes to inclinations "to do some lovely action, like helping those in need."[54] The problem with this common view is that it obscures a deeper question about the kind of pressure we succumb to when we are weak-willed. As Schapiro points out, "We need an account of what sort of pressure you are giving in to when you willfully deceive yourself, such that doing so counts as weak, rather than just wrong."[55]

So what kind of pressure is the pressure of inclination, on Schapiro's view? One tempting and natural thought is to say that the pressure on the deciding mind comes from proposals or reasons put forth by the instinctive mind. So take the example of being inclined to fight back. Schapiro writes,

> But what about the fact that these [physiological] responses are being guided by the thought, 'insulting-to-be-foughtness'? Doesn't this motivating thought, insofar as you are not alienated from it, put some kind of deliberative pressure on your deciding self. Isn't this thought, in itself, a proposal, or a prima facie reason, or a 'seeming' reason to fight back? No. 'Insulting-to-be-foughtness' is the thought of your instinctive mind, and that kind of mind cannot raise the question of whether or not the situation, viewed disinterestedly, warrants fighting back. … 'Insulting-to-be-foughtness' is neither a proposal, nor an argument, nor a judgment, nor a reason. It is not even a 'seeming' or 'prima facie' reason. *It is simply not a move in deliberation*.[56]

So her view is not that the instinctive mind pressures the deciding mind by trying to get it to go along with what the instinctive mind sees as the best thing to do. This is not possible on Schapiro's view because the proposals or activity of the instinctive mind are unintelligible to the deciding mind, which views the world as normatively neutral.

Rather, Schapiro accounts for the motivational pressure of inclination by appealing to the

---

[54] *Feeling Like It,* 138.
[55] *Ibid,* 140.
[56] *Ibid,* 147. Emphasis added.

deciding mind's temptation to give up the responsibility of deciding for itself. She writes,

> *The only thing that can pressure a free will is the burden of freedom itself.* Strictly speaking, our inclinations do not pressure us to do anything. But something about their nature, independent of variations in content, makes it the case that when we are inclined, we are faced with a perfect opportunity to flee the burden of our freedom. We are weak insofar as we freely take that opportunity. We are strong insofar as we do not.[57]

> [In the moment of drama] There are only two roads, only two rival principles, because fundamentally, these are the only two alternatives. You can accept your freedom and take responsibility for deciding what to do, or you can flee your freedom, and come as close as is humanly possible to acting without deciding, and without taking responsibility for doing anything. There is a high road and there is a low road, and the low road is not a positive course of action; it is the privation of one.[58]

So return to the case with which the paper began: when I resolve to train for a half-marathon, and I anticipate that I will be tempted to give up on my training plan and that it will be difficult to resist. What kind of difficulty am I anticipating? When I am in the moment of temptation and I don't want to go on my scheduled run, what kind of motivational pressure do I face? Schapiro's answer is that I face the difficulty or pressure of the burden of freedom. The force of my tempting desire to, for example, watch Netflix on the couch instead of doing my long run is not strictly speaking pressure from the desire itself, but rather the pressure of my own deciding mind that doesn't want to determine itself and decide like a human.

What this view gets right is our responsibility to decide for ourselves. If I impulse-buy a yurt because I wanted to buy one, and then explain to my spouse that I couldn't help buying the yurt because I wanted to buy one and was helpless to do otherwise in the face of my instinctive mind's activity, I have abdicated the burden of freedom and failed to accept the responsibility I have as an agent to decide for myself. But although we should not regard ourselves as

---

[57] *Feeling Like It*, 147.
[58] *Ibid*, 148.

determined by our inclinations, helpless in the face of an inner animal's wishes, Schapiro's view is stronger than this. Her view here includes the following further claims:

1. Incorporation Thesis: that if we do not incorporate our inclinations into our maxims and thereby humanize them, that we are acting weak-willed.
2. Burden of Freedom Thesis: that inclinations as such do not pressure our will, but rather our will pressures "itself" insofar as it fails to take up the burden of freedom.

The "Incorporation Thesis" is the label Henry Allison used for a thesis found in Kant.[59] Kant writes, "freedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim.*"[60] Allison takes this to mean that "for Kant an inclination or desire does not *of itself* constitute a reason for acting. It can become one only with reference to a rule or principle of action."[61] This is of course in keeping with Schapiro's view that the contents of our inclinations, the working of our instinctive mind, do not appear as reasons to the deciding mind.

So according to the Incorporation Thesis our inclinations do not themselves constitute reasons, and so to act on an inclination, we will have to transform it by incorporating it into our maxim. How does one do such a thing? Schapiro writes,

> Your incentive, as I will understand it, is your inner animal's guiding thought, e.g., 'chocolatey-to-be-eatenness.' *Qua* deciding mind, you cannot think this thought, in the way your inner animal thinks it. But you can take this thought as the raw material out of which to construct your own action-guiding thought, your maxim. *Your maxim will not be a stretch of teleological thinking. It will be a description of your action under which you find it worthy of your choice*. Your maxim will specify the what, where, when, how, and why of your activity, in such a way as to represent it as something worth it for you to do, given who you are, namely a free human being who happens to be situated in such-

[59] *Kant's theory of freedom*, Cambridge: Cambridge University Press, 1990, 40.
[60] *Religion within the Boundaries of Mere Reason,* Translated by Allen Wood and George Di Giovanni. Cambridge: Cambridge University Press, 1998, 6:24 (page 49).
[61] *Kant's Theory of Freedom*, 40.

and-such circumstances.[62]

This view draws on Schapiro's earlier claim that the instinctive mind sees the world teleologically. Earlier, she says that the instinctive mind cannot distinguish between substances and the practically salient attributes that attach to them.[63] So the instinctive mind does not perceive chocolate cake and then tag on "to-be-eaten" but rather represents the world in a practically loaded way. Unlike the instinctive mind, however, the deciding mind can distinguish objects and practically salient properties, and for this reason the deciding mind cannot simply take up the thoughts and activity of the instinctive mind. Rather it must take them as "raw material" for deliberation, deciding to incorporate the inclination into a maxim – or not.

Returning then to Schapiro's account of asymmetric or motivational force, Schapiro holds that we are weak-willed when we do not incorporate our maxims but rather "act" instinctively. I use scare quotes because as we shall see, she thinks we cannot be properly said to "act" instinctively. But if we cannot act instinctively, and furthermore inclinations cannot even pressure our deciding mind, how then are we going to account for the asymmetric pressure of inclination? Again, Schapiro's answer to this is the Burden of Freedom Thesis which holds that the apparent pressure of inclination is in fact the pressure to fail to be free, to fail to decide as a free being and instead to allow the instinctive mind to take over.

1.4 AGAINST THE BURDEN OF FREEDOM

I find this account of asymmetric pressure unsatisfactory on several accounts. First, the

[62] *Feeling Like It*, 132. Emphasis added.
[63] *Feeling Like It*, 112.

view sits uneasily with the experience of anticipating difficulty in the course of carrying out an inclination. What am I anticipating and planning for when I expect to be tempted? Am I anticipating that it will be a burden to be free? This seems unnecessarily complicated and at odds with the actual experience of temptation: most simply, when I anticipate that I will be tempted, I anticipate that I will not want to do what I've decided or resolved to do and that I will want to do something else instead. Of course, Schapiro might just respond that this doesn't yet explain *why* my wanting will be difficult, and that is correct; this response needs to further say why wanting to do something else places motivational pressure on me. I will try to give such a response in the following chapter. Still, how can it be a burden to be free when you have already decided what to do? In cases like my anticipating that I will be tempted in the course of my training, the work of the deciding mind has been done. Let's further assume that my decision was just that, the decision of my deciding mind, and then any inclinations were properly incorporated. So how is it that I anticipate my freedom will be a burden? In this case, I've already exercised my freedom in making my decision. Here too, one might respond that the pressure in such cases is the pressure to give up maintaining one's freedom. But such a response at least requires an explanation as to why it is a burden to maintain one's freedom in the face of a contrary inclination but relatively easy to do so when there is no contrary inclination. Absent such an explanation, I prefer to think that the inclination *itself* is a source of motivational pressure, since in this case freedom is not a burden until there is an inclination that conflicts with a free decision.

1.4.1 The second problem with this view concerns what happens when you act on a tempting desire. If inclinations as such cannot pressure the deciding mind, what are we to say about

instances of weak-willed action in which we "act" on inclination without incorporating it into a maxim? On many perspectives – my own included – the resulting movement is not in fact mere movement but is perfectly intelligible action. But how can it be intelligible as action, not mere behavior, and attributable to the agent, if inclination as such has no content that could pressure the deciding mind? Again, remember that Schapiro doesn't view inclination as reasons-responsive. She writes,

> The chocolate cake is in front of you. You are aware of your nonvoluntary physiological and conscious capacities already going for it. Your inner animal has initiated its *fress-ing*, guided by its instinctive thought, 'chocolatey-to-be-eatenness!' The other part of your mind, your deciding mind, does not share in precisely this thought. Instead it sees the chocolate cake as an object that exists, at least in principle, independently of its to-be-eatenness. It sees the cake disinterestedly, as an object in the world that is not essentially for its own consumption. … [Insofar as you occupy a deciding mind,] you are in a position to try to construct a maxim that will allow you to eat the cake in a way that accords with your principles. … [But] The presence of your instinctive mind gives you, *qua* deciding mind, another place to go. You can allow yourself to inhabit your instinctive mind, to come as close as possible to being the thinker of its thoughts. You cannot fully inhabit this mind, but you can act as if you could. When you do this, you conceive of the cake *as if* from the standpoint of one *whose nature it is to be an eater of chocolatey-to-be-eatenness.* You can conceive of the cake as if you were wholly a creature of instinct. And although you cannot decide to *fress*, and thereby *fress*, I want to suggest that you can do something that comes close to *fress-ing* without deciding to do so. This is possible if you can *attach* to 'chocolately-to-be-eatenness' as your incentive, without taking responsibility for incorporating it into your maxim.[64]

So if you eat the chocolate cake by attaching to 'chocolately-to-be-eatenness' as our incentive, without taking responsibility for incorporating it into your maxim, Schapiro says you are acting as if you were an instinctive mind, although you cannot fully inhabit this mind. But if you eat chocolate cake without incorporating your incentive to do so into your maxim, what are we to say about the resulting movement? Is it intelligible as action and furthermore, your own action?

Schapiro is ambiguous on this point, perhaps purposefully. She writes, "When you take

[64] *Feeling Like It*, 151.

the low road, you are acting without really doing anything, well or badly. By the same token, you are acting without really knowing or taking responsibility for what you are doing."[65] Then later, more pointedly raises the question: "How, then, is [weak-willed action] a form of willing? I don't know that I have a clear answer to this, except to say that in fleeing, you are not overpowered by an external force, and you are not responding to the force of a reason."[66] This response is understandable in that both of these are reasonable criteria as a theory of weak-willed action: we don't want weak-willed action to consist in being overpowered by an external force, and we want to account for how it is weak, and one plausible way of doing so is to deny that it responds to the force of a reason. But the question still remains: what are we to say about the "movement" resulting from weak-willed action? Can we call it action and attribute it to the agent?

By all accounts, it's action. Of course, it's possible to eat chocolate cake as a sleepwalker, or in the grips of an obsessive compulsion. But the normal case of eating chocolate cake without properly incorporating the incentive of the instinctive mind into one's maxim is not either one of these cases. The normal case is perfectly intelligible as action, and the agent would be perfectly capable of explaining the action. I know Schapiro is reticent to say that the agent in such a case would act for reasons, but I find it entirely natural to say that the agent would be able to give us their reasons for eating the chocolate cake. Why did they eat it? Because it looked good. Because it was chocolate. Because they think chocolate tastes good. Unlike eating chocolate cake at 2 am under the influence of Ambien, the normal case of eating cake instinctively is perfectly intelligible as action because we can take the thoughts of the instinctive

[65] *Feeling Like It*, 152.
[66] *Ibid*, 154.

mind as reason-giving. We may not view those reasons as decisive or good reasons, but we recognize them as the right type of consideration to be a reason. This is of course something Schapiro has denied, but in the following I will build on and defend this view of inclination.

If on the other hand we agree with Schapiro and say that weak-willed action is some form of willing although it is not a response to reason, how are we to respond to weak-willed, instinctual "action" done by another person? Say the chocolate cake you eat instinctively is not just any chocolate cake but is my chocolate cake, lovingly baked for my friend's birthday party and waiting in the kitchen for the party this evening, and you eat it. If you were truly in the grips of an external force, I could not hold you accountable, at least not in the fullest sense. Both Schapiro and I agree that in this situation you are not in the grips of an external force when you acted instinctively. But I worry that on Schapiro's account I still cannot hold you accountable. How can I hold you accountable for eating my cake if you were not responding to a reason in so doing– or we might say, *acting* at all? I am not implying that I should hold you accountable for instinctually eating cake in the same way I might hold you accountable if you deliberately and maliciously ate the cake in order to spite both me and my friend. But I think I would be within my rights to level a complaint against you for instinctively eating my chocolate cake in a way that would not be appropriate if you ate the cake during a sleep walking episode.[67]

Is it an objection to this view that you might reasonably respond "I wasn't thinking!" when I confront you about eating my cake and ask what you were doing or thinking. Is this perhaps some evidence, even if partial, that when someone eats cake instinctually it's not really action or willing, properly speaking, but is rather an instinctual and thus automatic behavior? I

[67] Furthermore, my complaint against you would be *that you ate the cake,* not that you failed to live up to the burden of freedom and acting instinctively. Thanks to Kyla Ebels-Duggan for raising this point in correspondence.

think not, for several reasons. One, as Schapiro herself points out, acting instinctively (as opposed to incorporating one's incentive and deciding) is perfectly compatible with a great deal of thinking about one's action.[68] So the absence of thought is not necessary for weak-willed action. Furthermore, there's another perfectly natural reading of the retort "I wasn't thinking!" available to us. On this reading, the person who says "I wasn't thinking!" is admitting that no, they did not have good reason for their action, and in fact if they had stopped to consider for even a moment they would have realized, for instance, that I baked the cake last evening while talking about the party and yelled "Don't eat the cake!!" again on my way out the door this morning. But saying that your action was unjustified is not the same as saying that your action is unintelligible or mere movement.

1.4.2 The final concern I want to raise with Schapiro's view is that we seem to have slipped off the careful tightrope between inclinations as the sort of thing that can be taken up for deliberation and inclinations as having some kind of motivational force. It is tricky to reconcile the passive nature of inclination with their role in rational deliberation, but I worry that Schapiro's view ends up embracing their passivity and force at the cost of agential unity. That way of putting it is a bit disingenuous, since a key tenet of Schapiro's view is that the agent is *not* unified in the sense of having one motivational source. Rather, what I mean is that it's not clear on her view why inclinations are ours as opposed to things that happen to us.

> Consider again how the deciding mind relates to the instinctive mind on her view:
>
> 'Insulting-to-be-foughtness' is neither a proposal, nor an argument, nor a judgment, nor a reason. It is not even a 'seeming' or 'prima facie' reason. *It is simply not a move in deliberation.*[69]

[68] *Feeling Like It*, 155.
[69] *Ibid,* 147. Emphasis added.

> *Qua* deciding mind, you cannot think ["chocolatey-to-be-eateness"] in the way your inner animal thinks it. But you can take this thought as the raw material out of which to construct your own action under which you find it worthy of your choice.[70]
>
> You can allow yourself to inhabit your instinctive mind, to come as close as possible to being the thinker of its thoughts. You cannot fully inhabit this mind, but you can act as if you could.[71]

But this is the problem: if we cannot fully inhabit the instinctive mind, and if the thoughts and activity of the instinctive mind are not even moves in deliberation, not reasons, then in what sense is the instinctive mind ours at all? We can "take it as the raw material" out of which to construct our maxim, and we can to some degree – but not fully – inhabit this mind. But these are not sufficient for concluding that the instinctive mind is our own mind. I can also take your stated preferences as the raw material for constructing my maxim, and I can sort-of inhabit your mind inasmuch as I can imagine what things look like from your perspective. But that is not thereby reason to conclude that your mind is part of me, my mind, or my agency.

Schapiro is attentive to this problem. In her words, we must give an account of why having an inclination is a "motivational condition" rather than an event that the deciding self observes.[72] In response to this challenge, she writes,

> Insofar as you are aware of yourself as being *inclined* to fight back, you do not see yourself as undergoing a disorganized and unintelligible aggregate of experiences. You are not regarding your experiences as, say, symptoms of an illness, which may have the unity of an immunological response, but which do not have the unity of a purposive response to the world as you see it. Nor are you meditating, dissociating from your inclination *qua* motivational condition. Rather, having an inclination is a motivational condition insofar as you regard your experiences as constituting a purposive response in you, oriented toward achieving an aim that is at least minimally intelligible to you.[73]

---

[70] *Feeling Like It*, 132.
[71] *Ibid*, 151.
[72] *Ibid*, 126-7.
[73] *Ibid*, 127.

If we interpret this quote to mean that the instinctive mind's activity constitutes a motivational condition because its activity and perspective are purposive and teleologically laden, and thus intelligible to us, then I think this response is right. But if we interpret her response to instead simply say something along the lines of "your inclinations are yours because you normally regard them as yours in some sense," then we must ask the further questions of whether the agent is right to regard them as his and if so, why he is right in so doing.

In other words, it seems that we have returned to the dilemma that motivated Schapiro's view. The primary reason for rejecting practical thinking views of desire is that they make desire too deliberative, the output of practical thinking, and thus have a hard time explaining why we cannot desire at will and why certain desires are recalcitrant. Schapiro's view aimed to resolve the dilemma by introducing dualism into our motivation: we have a deciding mind, and an instinctive mind, and both minds are ours, but they are also independent. Thus the hope was that we could capture how inclination is deliberative *and* non-voluntary and a source of motivational pressure. But now we have returned to this dilemma with the worry that Schapiro's view goes too far in the opposite direction. How can inclinations be deliberative if inclination is not even a move in deliberation?[74]

One natural thought is to take Schapiro's view and amend it back in the direction of Scanlon and practical thinking type theories. On this view, the thoughts and activity of the instinctive mind *are* moves in deliberation, and the instinctive mind does present reasons to the deciding mind, even if our grasp on those reasons is partial and non-propositional. This is the view I will develop in the following chapter.

---

[74] *Feeling Like It*, 147.

2. Intending Instinctively

On my view, inclinations are what I will call "provisional decisions" of the instinctive mind. In other words, our instinctive mind is an aspect of our agency that continues to be animal. If we were wholly creatures of instinct, i.e. if we lacked a deciding mind, the instinctive mind would guide our action. This is why inclinations have force and why they are like decisions: they are products of a part of our agency that functions to guide action, albeit guide it in a way differently than what we standardly think of when we imagine deliberating and forming an intention. However, although we have instinctive animal agency, we are not creatures wholly of instinct. We have deciding minds, which is to say that we can direct our attention towards our attitudes themselves, including inclinations, and ask whether they are fitting and whether we ought to act on them. Bringing the deciding mind into the equation doesn't just add to our agential capacities, making it possible for us to act instinctually *or* act decisively, but furthermore changes our instinctive mind. Because we are not wholly creatures of instinct, our inclinations cease to be decisions (or more precisely, decision equivalents) but rather function as *provisional* decisions, because having an inclination is not the same as forming an intention. My account retains Schapiro's emphasis on duality and the idea that the force of inclination crucially depends on our animal agency, but unlike Schapiro I think in the standard case our inclinations are reason-laden, and thus the deciding mind can inhabit inclinations and act on them. That is, inclinations are not just imperatives to act but are standardly imperatives-to-act-because-x, even if they are not experienced in propositional form.

2.1 IMPERATIVES AND REASONS

In one sense of the word "instinct," instincts are wholly automatic tendencies to act in particular ways; literally reflexes or reflex-like movements. However, as I'm using the phrase, to say that we have an instinctive mind is not just to say that we have a collection of reflexes or tendencies but first and foremost to say that we experience the world teleologically, i.e. that our instinctive or animal mind sees the world "already practically interpreted." *Qua* animal, a human "confronts a world of things that are perceived directly, without calculation or conscious interpretation, as things that are *to-be-avoided, to-be-chased, to-be-investigated, to-be-eaten, to-be-fled, to-be-cared-for*, and so on."[75] Thus the instinctive mind is not just a capacity for brute or mechanical "action" but is rather the capacity to represent the world in practically loaded terms and act in light of this representation.

One important difference between the instinctive mind account and, for example, Scanlon's account of desire, is that the instinctive mind's teleological understanding is a world of immediate imperatives. But the imperatives given by the instinctive mind are not perceptions of reason. Thus, whereas Scanlon emphasizes desires as seeing reasons, Korsgaard's idea of the teleological understanding of the world emphasizes that instinctive minds see the world in terms of imperatives to act.

In the human case, however, the adult human's teleological understanding of the world is not just the representation of a world of imperatives to act, but rather an understanding of the world in which we see certain things as *to-be-done-for-reason-x*.[76] Of course, inclinations are not

[75] *Fellow Creatures*, 38.

[76] What about the animal case? In the case of non-human animals, I think we cannot say with any certainty. It seems

literally experienced in this form, but I think this description more closely mimics the nature of inclination than a pure imperative.[77] Try to imagine a pure imperative with no rational content. What would it even be to experience such a state? Just a nagging sense that I ought to do something? A voice in my head repeatedly saying "do φ?" Would it be the same as another person telling me that I ought to do something, regardless of my assessment of what they think I should do? My body literally beginning to move towards executing the imperative, like a reflex?[78] On the contrary, in the standard case we experience an imperative as *to-be-done* for some reason, even if that reason is not fully articulate. I don't just experience a pure imperative to eat the cake, rather, I see the cake as to-be-eaten *because* I see it as chocolatey and delicious. I'm not just struck with a nagging sense that I should go to the beach, but I see the beach as to-be-visited *because* it is sunny and the water is blue.

If inclinations had no rational content, it is hard to see how they could exercise motivational force. Inclinations might literally put our body in motion, making it hard to resist

---

reasonable to conjecture in the case of young animals that they experience the world in terms of pure imperatives: the newly hatched bird knows to open its beak towards its parent returning with food before it could have possibly learned that you open your beak in order to be fed with something hopefully delicious and nourishing. In the case of adult animals, on the other hand, one assumes that the animal experiences the imperative to act in response to a representation, which comes much closer to the human case of acting in light of a reason. For example, while a kitten may try to eat litter, not a suitable food source for felines, a healthy adult cat normally only eats when they smell and see as appropriate food. Thus we might speculate that the kitten simply experiences the imperative "Eat!" and tries to eat anything they can easily put in their mouth and chew, while the eating adult cat seems to be doing something much closer to delicious-to-be-eaten. However, for obvious reasons this is all speculative. We cannot say for certain what it is like for any given animal to experience the world teleologically and this experience no doubt varies across species.

[77] By "pure imperative" I mean an imperative devoid of rational content. We might also think of such imperatives as urges or reflexes, but I am trying to keep my description as free of theory as possible. We might also call such imperatives "brute" or "blind," but brute imperatives are confusingly different from brute forces like currents or strong winds, and "blind" is ableist in this context.

[78] I don't actually think this is Schapiro's view of inclination, but it's striking to me that her descriptions of inclination do sometimes veer in this direction. For instance, she writes, "From your standpoint as a free, deciding self, the fact that your heart rate is increasing, and that your muscles are tightening, and that your respiration is growing shallow, and that your attention is being insistently focused on your foe, makes it easier to execute the decision to fight back, should you make it" (*Feeling Like It*, 146).

the movements of the body in the same way it's hard to swim against a strong current, but as discussed in the previous chapter, this is not the sense of force relevant to inclinations.[79] Inclinations could also appear as imperatives from a foreign entity, namely the instinctive mind. But again, unless the imperatives had rational content such that to consider an inclination is to inhabit a certain understanding or perspective on goodness, imperatives from the instinctive mind would have no more force than another person yelling instructions at us that we can discern no particular reason to follow. Such instructions might be distracting, thus making it hard to act otherwise, but the imperatives would not move our motivation in the way essential to motivational force. Imperatives from other people only exercise motivational force on us when the person has some special authority over us, or if we are especially deferential and eager to please others, or if the imperative independently strikes us a good way to act. In other words, imperatives have motivational force only when we perceive that we have reason to obey them, whether that reason is internal to the content of the imperative or not.

Consider the example of the inclination to eat chocolate cake. The inclination is experienced as an imperative, but part of the motivational force of the imperative is having one's attention drawn to what would be good about eating the chocolate cake. It is difficult to tease apart the representation of the chocolate cake as delicious and worth eating (what I am calling the "reason") from the imperative *to-be-eaten*, since the instinctive mind perceives them simultaneously,[80] but insofar as they can be teased apart when held up for examination by the deciding mind, the imperative clearly depends on the representation of the reason to act. Thus if I

[79] See e.g. *Feeling Like It*, 48-9.

[80] On this point I agree with Schapiro that it would be a mistake to say that the instinctive mind views an object neutrally and then attaches practical significance to it through some stretch of reasoning.

see a delicious chocolate cake or contemplate the idea of a delicious chocolate cake, I may furthermore see it as *to-be-eaten*, but if I realize that the delicious-chocolate-appearing cake is in fact a doggy liver cake for a dog's birthday party, the imperative *to-be-eaten* will rapidly disappear in the absence of the representation of the cake as delicious and worth eating.

Furthermore, what we see as reason to act a given way is arguably part of the content of the inclination to so act: if I have the inclination to fight because I am a boxer and I am amped up and ready to win a match, this inclination is importantly different from the inclination to fight because I'm furious and insulted.[81] This then is the sense in which I am retaining Scanlon's emphasis on desire as a tendency to see as a reason. When we are inclined to eat chocolate cake, our inclination is dependent upon our seeing reason to eat the chocolate cake and furthermore our inclination is likely to focus our attention on this reason or reasons. This is why acting on an inclination or desire standardly serves to make the ensuing action intelligible: it tells us something about what the agent saw as desirable in so acting.[82]

[81] I suspect that a similar point holds for the inclinations of at least some animals. If my cat has the inclination to eat something because it is delicious, she will refuse to eat it if it turns out to be a food she doesn't like after all. But if her inclination is more like eat-because-starving-anything-will-do, she may be willing to eat anything even vaguely resembling a meal, for instance leaves, sticks, and rocks. (Rescues that take in starving stray cats often find that their stool is full of such objects, none of which are nutritious for a cat and presumably none of which are delicious to a cat.) Thus although the cat may have the inclination to-be-eaten in both cases, it can be practically salient to distinguish between inclination motivated solely by deliciousness (my cat wanting her favorite treat right after a big meal), inclination driven by both hunger and deliciousness (the standard case for domestic cats living as pets), and inclination driven solely by hunger (the case for starving cats).

[82] Tal Brewer argues this same point, writing, "… even the simplest human desire carry the phenomenological traces of some inchoate sense of the good that hangs on fulfilling them. Consider, for instance, the varying outlooks on value that can inflect a desire to eat a particular piece of cake. There is a subtle phenomenological difference between: (1) desiring to eat a piece of cake, even though one does not like sweets, because one is very hungry; (2) desiring to eat that same piece of cake because one has had it in the past and one knows it to be delicious; and (3) desiring to eat the same piece of cake because one's elderly uncle has baked it and fawning over his baking is the family's ritual manner of acknowledging him. Each of these desires carries traces of a different picture of the good that might hang on eating the cake" (*The Retrieval of Ethics,* 28).

2.1.1 However, several gaps remain with respect to this account, especially concerning the idea of seeing reasons. Consider again the chocolate cake. We represent the cake as delicious and thus the instinctive mind sees it as *to-be-eaten* and provisionally decides to eat the cake. But it is also possible, and indeed is often the case, that we simultaneously represent the cake as delicious (thus seeing reason to eat the cake) *and* judge that we have reason to avoid the cake, perhaps being in the grips of a regrettable diet on which all sugar is strictly forbidden. So how can we reconcile this difference? Isn't this proof that inclinations are not reason-laden at all, since in cases like these the inclination to eat the cake persists in the face of the judgment that the cake is to be avoided at all costs?

In addressing this kind of recalcitrance, Scanlon often appeals to a distinction between seeing and judging.[83] The distinction needs further development, but I think the idea of the instinctive mind's teleological understanding of the world can supply just that. Korsgaard observes that

> Perception first evolved in animals who are not highly intelligent, and would have been useless if all it did was flood their minds with neutral information that needs to be processed by intelligence or reason before it can be of any use. So the world comes to an animal already practically interpreted… [84]

Thus I think we should add to Scanlon's distinction between seeing and judging the idea that our instinctive mind, the part of our agency that we share with non-human animals, represents the world in practically loaded terms according to its own needs and interests.[85] This is why our

---

[83] See e.g. *What We Owe to Each Other*, 40.

[84] *Fellow Creatures*, 38.

[85] This highlights another potential ambiguity in the account as I've described it thus far: although it may be tempting to think that cake essentially has the properties *delicious* and *to-be-eaten* and our teleological understanding of the world is just our instinctive mind directly perceiving these properties in the cake, in fact the teleological understanding of the world is more like the instinctive mind imposing its own interests and needs onto its perception of the cake, or viewing the cake through the lens of its own interests and needs, such that the cake is automatically seen as *delicious-to-be-eaten*, as opposed to seeing the cake as practically neutral and having to

instinctive mind sees the world primarily in terms of what will bring pleasure and comfort and safety and protect from pain and difficulty, although it can be shaped and cultivated over time. And in the case of chocolate cake, this is why the instinctive mind may continue to see the cake as *delicious-to-be-eaten* even when the deciding mind judges from a reflective distance that we do not in fact have reason to eat the cake. Thus in the following when I refer to "seeing" reasons I refer to the reasons inchoate in an inclination, whereas "judging" reasons will refer to considerations regarded as reason-giving by the deciding mind.

However, it is important to note that I regard both reasons seen and reasons judged as reasons. I take my account of reasons from the work of Pamela Hieronymi, and so I regard reasons as "items in pieces of actual or possible reasoning."[86] However, when reasoning we can use what the instinctive mind sees as much as what the deciding mind judges, and so the difference between the reasons of the instinctive and deciding mind is not that one mind sees what are merely apparent reasons, but rather that we might in reasoning come to regard the reasons seen by the instinctive mind as insufficient grounds to settle an intention to φ.[87] To give a simple example of this, say that the chocolate cake I am inclined to eat is sitting on a nearby table at a restaurant: it is a slice of birthday cake about to be eaten by a child on what appears to be their birthday. I might wish to eat the cake on account of it looking very delicious, and its apparent deliciousness is indeed a reason to eat it, but that reason is wholly inadequate when

---

deliberate coolly about how to conduct oneself with respect to the cake. Schapiro makes this point in *Feeling Like It* (112).

[86] "Reasoning First," In *The Routledge Handbook of Practical Reason*. Edited by Ruth Chang and Kurt Sylvan. New York: Routledge, 2020, 355. She goes on say, "Reasoning is thought organized in a certain way: directed at a question or conclusion. Thus, I would suggest, reasons are considerations that either bear or are taken to bear on a question" (355).

[87] In other words, my view is that the deciding mind discovers rather than "creates" reasons. It is one thing to say that the deciding mind has the freedom to question apparent reasons and deliberate about them in settling practical questions. (I affirm this.) It is another thing to say it has the freedom to create reasons. (I deny this.)

compared to all the other reasons I have for leaving the child and their cake in peace. In fact, if I am a well-functioning adult, in this situation my instinctive mind will also see the situation as one in which the child's food is *not-mine-to-be-ignored*, and insofar as I am inclined to eat the cake, I will almost certainly be inclined to get a slice of the same cake for myself and not inclined to steal the child's cake for myself.

2.1.2 Thus far I have been emphasizing, contra Schapiro, that inclinations are reason-laden. Our instinctive mind sees the world teleologically and thus its provisional decisions are borne out of its practically loaded representation of the world, but this does not mean that the instinctive mind sees the world in terms of pure imperatives for which we can discern no reason. However, one might object that not all inclinations are reason-laden in this way. It seems intuitively possible that our instinctive mind can direct us to act – e.g., can see the world in terms of an imperative to act – even when the imperative is devoid of reason. One natural thing to say about such cases is that in these moments, the problem is that the reasons of the instinctive mind cannot or are not being communicated to or understood by the deciding mind. Consider, for example, the inclination to startle or run in the face of a harmless loud noise in a person with PTSD. Presumably the person in question experiences no sense of why it would be good to startle or run, but we might nonetheless think that PTSD is a pathology in which there is a great divide between the instinctive and deciding minds: the instinctive mind represents a harmless loud noise as a source of great danger and thus decides to run in the face of danger, even as the deciding

mind consciously registers no threat, thus seeing the imperative to startle or run as totally devoid as reason.[88]

My response to this objection is to allow that it does seem possible to experience "pure imperatives," by which I mean the imperative that normally constitutes an inclination but absent the perception of a reason also involved in an inclination. In addition to PTSD type cases, I suspect that some whims are pure imperatives of this sort. However, I want to further emphasize that pure imperatives are not the standard case of inclination in a well-functioning adult. If inclinations in general were like this, we would regard inclinations as puzzling reflexes or tics, not provisional decisions that are candidates for our action. In other words, if inclinations were pure imperatives, we would have to specifically come up with reason to follow the imperative before we would act on it in the strong sense of action. By way of analogy, say that I am on a walk when someone driving by yells to me, "Stand in the crossroads and look." Unless I know the person, we are both participants in some cultural context that makes sense of the incident, or the person in some way appears to pose a threat and I obey them for the sake of avoiding harm, I will see no reason to follow the imperative. If I am going to act on this imperative, I will have to give it a reason myself, perhaps concluding that the person is a potential threat and it's best to cooperate with their instructions until they're at a safe distance. Or, I might glance over and see some object in the crossroads, and thus decide to follow their imperative because their strange

[88] Although I by no means intend for this example to be a definitive example of PTSD, my description of the case is drawn from the descriptions of PTSD presented in Bessel van der Kolk's *The Body Keeps the Score*, New York: Penguin Random House, 2014.

instructions coupled with the presence of something in the road gives me reason to think that there is something to be found.

The same thing is true of pure imperatives given from within rather from without. Presumably Warren Quinn's Radio Man case is a case of this sort. Radio Man feels an urge to turn on every radio he sees but does not see anything good about radio's being turned on.[89] If turning on radios is to be intelligible as his own action rather than a reflex or tic, Radio Man will have to specially incorporate the imperative by coming up with reason to act on the imperative. The reason need not be deep; he might simply make a game out of turning on radios as soon as possible or turning on as many radios as possible. But he will nonetheless have to find some kind of reason in order to act on the imperative.

But again, inclination as I am defining it is not like this. The boundaries are blurry, and I will not take a firm position on whether, for instance, young children have reason-laden inclinations, or inclinations of a related but different sort, or no inclinations proper at all but only pure imperatives. In other words, I am defining inclination according to the paradigm case of a well-functioning adult, and in such a person to have an inclination is, among other things, to have some grasp the reason why the object of one's inclination is to be done.

The emphasis on adulthood is not an incidental feature of my definition. In the well-functioning adult, there is a good deal of congruity between what the instinctive mind sees as a reason and what the deciding mind judges to be a reason. But this is a strong claim that immediately needs clarification. In saying this, I am not taking back what I say above about the divergence between what is seen as a reason vs. judged as a reason. It is totally implausible to

[89] "Putting Rationality in its place." In *Morality and Action*. Cambridge: Cambridge University Press, 1994, 246-7.

think that our teleological understanding of the world just stems directly from what we judge to be our reason, such that if I decide to tackle a difficult project like writing a dissertation, my instinctive mind will from then on out automatically see writing as *to-be-done* such that I am strongly inclined to write every time I have the chance. To take the view that the instinctive and deciding mind share a practical perspective is arguably to collapse the distinction between the instinctive and deciding minds. On the other hand, however, it is a sign of well-functioning agency for an agent, over time, to learn to (mostly) want to do what they judge they should be doing. Consider the chocolate cake in a restaurant example above: as has been pointed out, although a toddler or young child may specifically want to take cake from someone else in order to eat it, well-functioning adults should not want such a thing.[90] A well-functioning adult might want to get a slice of cake for themselves, but they aren't inclined to take cake from someone else in order to eat that exact slice themselves. Or, consider a world-class athlete: surely no athlete is inclined to train at every scheduled training session, but it is no doubt impossible to become a world-class athlete without being generally inclined to train. The young athlete with raw talent and no training will need to *want* to train, or *learn to want* to train, if they are going to hone their skills and develop in their sport. This process of habituation is in large part the integration of what they see as reason vs. what they judge as reason.

[90] See Barbara Herman's "Making Room for Character," in *Aristotle, Kant and the Stoics: Rethinking Happiness and Virtue*, edited by Stephen Engstrom and Jennifer Whiting, Cambridge: Cambridge University Press, 1996, 46.

2.2 PLEASURE AS PRACTICAL COGNITION

In order to understand the process of habituation that shapes our teleological understanding of the world and leads our instinctive mind to see certain actions as *to-be-done* automatically, we must take a detour into the idea of pleasure. In some sense, it is odd that we are this far into a dissertation about temptation and pleasure has hardly been mentioned. Lay people immediately associate pleasure with temptation in general and inclination in particular: why is it that we want to stay on the couch and find it hard to get up and go running? Surely part of the answer is that it is pleasant to lie on the couch and decidedly unpleasant to go for a run (at least from the perspective of someone who wants to avoid running by staying on the couch). Still, we must quickly qualify talk of pleasure when it comes to inclination. Certainly I do not wish to advance a view on which pleasure is the object of inclination as such. It is difficult to get one's language right, but Korsgaard helpfully summarizes the matter as follows with respect to the inclination to dance:

> We should not say, in this kind of case that pleasure is really the object of the inclination, rather than dancing. Dancing is the object of the inclination; to say that it is pleasant is just to say that there is a natural incentive for dancing [i.e. pleasure is what makes dancing attractive]. … To say that dancing is pleasant is not to say exactly what the incentive is: 'pleasure' is just a kind of dummy word that indicates that there is one. … You can explain the incentive, sometimes, but at bottom it is not something articulable. We just gesture at the fact that there is an incentive, a positive attraction, a suitableness of the object to our nature, when we say that the object is pleasant. Having said that, let's nevertheless call the incentive, for short, 'pleasure.' Then the point of saying that pleasure is the person's incentive is not to say that he likes the pleasure rather than the dancing, *but rather that the pleasure, the sheer joy of the movement, is what he likes about the dancing*.[91]

[91] *Self-Constitution,* 120-121. Emphasis added.

This is a helpful start: something about pleasure captures our attraction to the objects of our inclination. We can immediately see how this connects with the description of inclinations as reason-laden above: inclinations are not just pure imperatives, the sense that something is *to-be-done*, but to have an inclination is furthermore to be attracted to something about the object of one's inclination. Above I described this attraction as the fact that inclinations are reason-laden, and although talk of reasons sounds much colder than talk of pleasure and attraction, I think they are the same thing. In the case of an inclination, what we see as good about doing something (for instance dancing or eating the cake) is inextricably linked to our attraction to or the pleasure we take in the object of our inclination.[92]

2.2.1 I think we can further flesh out this view by appealing to a broadly Aristotelian account of pleasure.[93] On this view, pleasure and pain are practical perceptions.[94] That is, when we find something pleasant, we are struck by its value or worth, which is to say that pleasure is way of perceiving good. More specifically, all action depends on practical cognition or the identification of one's object of action as in some sense good.[95] Practical cognition might work through

---

[92] Although I will focus on pleasure and attraction to an object of inclination here, I think the same view holds with respect to pain and aversion.

[93] More precisely, what I say in the following uses broadly Aristotelian ideas about pleasure for the sake of building a theory of inclination and is not meant to be a fully faithful account of Aristotle's actual views. In particular, I will not be consider Aristotle's views on the role of pleasure in a life of *eudaimonia* and the idea that pleasure is a completion or perfection of an activity. For more on these topics, see Verity Harte's "The *Nichomachean Ethic*s on Pleasure," in *The Cambridge Companion to Aristotle's Ethics*, edited by Ronald Polansky, Cambridge: Cambridge University Press, 2014 and Dorothea Frede's "Pleasure and Pain in Aristotle's Ethics," in *The Blackwell Guide to Aristotle's Nichomachean Ethics*, edited by Richard Kraut. Oxford: Blackwell Publishing, 2006.

[94] In saying this, I am rejecting the idea that pleasure is a sensation. See for instance Korsgaard's discussions in *Fellow Creatures*, 160-2. See also Brewer, *The Retrieval of Ethics*, 117.

[95] My use of the phrase "practical cognition" for the kind of practical cognition that picks out a worthy object of action is drawn from Jessica Moss's book *Aristotle on the Apparent Good: Perception,* Phantasia, *Thought, and Desire.* Oxford: Oxford University Press, 2012 (see esp. p. 10). As will become apparent, the entire ensuing discussion of pleasure as perception of the good is influenced by Moss's exegetical work. However, it is important to clarify that I do not follow Moss in fully embracing what she calls "practical empiricism." That is, I think intellect or thought can set ends and discern goodness on its own. For more on Moss's rather Humean, empiricist reading of Aristotle, see "Was Aristotle A Humean?" in *The Cambridge Companion to Aristotle's Ethics*, edited

thought, or in the terms I've adopted from Schapiro, the deciding mind might judge that a certain act is good, but pleasure is a different and more basic form of practical cognition that is available to non-human animals as well as humans. To say that pleasure is a form of practical cognition is to say that when we take pleasure in something, we perceive it as good. Again, pleasure as practical cognition is not the same thing as **judging** that a thing is good.[96] Judgment is a form of practical cognition, but the most basic form of practical cognition is perceptual and works through pleasure. Thus for example I might judge that I ought to do something (go for a swim in a cold lake) and furthermore judge that one of the reasons to do the thing in question is that I will find it pleasant once I am underway, but not be inclined to do the thing in question. In such a case, I am not perceiving swimming pleasurably and thus my instinctive mind is not perceiving these activities as worthwhile.[97]

Because pleasure is a basic form of practical cognition, pleasure is motivating. Or, more precisely, when the object of our pleasure is something we can do, to take pleasure in something

---

by Ronald Polansky, Cambridge: Cambridge University Press, 2014.

[96] Moss writes, "Cognitively, [finding pleasant] is an extremely simple and basic way of finding something good. It need not involve *thinking* that the thing is good; it is a purely sensory experience which may be independent of or even at odds with thoughts about goodness" (*Aristotle on the Apparent Good,* 30). Consider also Brewer: "We can affirm that pleasure involves running appearances of goodness or value without implying that it requires one to judge or even to have a tendency to judge that what one is doing or experiencing is valuable. Pleasure requires only a vivid appearance of value. Hence it is perfectly possible to take pleasure in an activity that one judges on reflection to be valueless" (*The Retrieval of Ethics,* 130).

[97] Aristotle further emphasizes that we don't just directly perceive as pleasant but we also have *phantasia*, i.e. the capacity to represent to the mind things we have perceived before, and *phantasia* can furthermore have an affective component i.e. be pleasurable or painful. For my purposes we don't need to worry too much about *phantasia* as opposed to direct perceptual contact – it is just sufficient to know that we can pleasurably represent something we are not currently in direct perceptual contact with, remembering or anticipating. This is what makes it possible for instance for me to be inclined to eat chocolate cake even when there is not chocolate cake in my presence, or what makes it possible for a predator to start hunting for food even when no prey is immediately perceptible. See Moss's *Aristotle on the Apparent Good,* Chapter 3 (48-66).

is to see it as *to-be-done*.[98] It is probably easiest to see this in the case of a non-human animal who cannot judge an object good or grasp its goodness in thought because it lacks both language and capacity for thought. In a non-human animal, to perceive something with pleasure is for the animal to see that object as good and *to-be-pursued* or *to-be-done*. Pleasure for this animal is both motivating (tells them to go for the object/represents it as to-be-pursued) and a basic form of cognition; it is how they apprehend the goodness of an object for them. Thus when everything is functioning correctly, an animal gets what is good for her by going for what pleases her. In a healthy and well-functioning cat, the pleasure he takes in eating meat both motivates him to eat and picks out the meat as something that is good for him.

It is tempting to say that when we perceive something as pleasant, we construe it as *to-be-done* for the reason that it is pleasant. This is close to accurate as a description of an action we take pleasure in purely for the reason that performing it will be pleasant, but there are many cases where it would be a distortion of the worth we see in some action we are inclined to perform to say that the reason for performing it is purely the pleasure of performing the action. (As Korsgaard helpfully puts it in the quote above, it is not that we like the pleasure rather than the dancing, but what we like is the pleasure of the dancing.) Instead, on the view I am advancing, pleasure itself is a form of cognition. *Pleasure is the way we conceive the objects of inclination.*[99]

---

[98] "Pleasure is indeed by its nature essentially motivating, but not necessarily appetite- or action-inducing. In some cases, pleasurably perceiving or contemplating something motivates us simply to carry on perceiving or contemplating it" (*Aristotle on the Apparent Good*, 27).

[99] Consider the example of being inclined to fight: it's not right to say that I see the person who insults me *as to-be-fought* and my reason for acting thus would be "pleasure!" To say that inclinations are pleasurable perceptions is compatible with saying that the reason to fight is the insultingness, but the instinctive mind sees fighting as good and *to-be-done* in this situation of insult because it perceives them pleasurably. (Or, more accurately, in this case we should probably say that the instinctive mind perceives the insult with pain and sees it as too be overcome or removed via fighting.)

Or, more specifically, the animal perceptual faculty is designed so as to regard beneficial objects with pleasure, and this is true in human as well as non-human animals.[100]

2.2.2 Bringing the discussion back to inclinations and the instinctive mind, we are now prepared to say that to have an inclination for something, to see it as *to-be-done* is to pleasurably perceive it. This doesn't mean that the instinctive mind is like some simple machine that's always programmed to seek feelings of pleasure, or that the instinctive mind only takes biological objects that are clearly related to pleasure, e.g. food and sex. On the contrary, pleasure is the efficient cause of inclination, not the object of inclination.[101] Pleasure tracks what is good for the organism. The instinctive mind goes for what it perceives pleasurably because that is the instinctive animal way of discerning the animal good. In a healthy animal living in the right environment, it's a pretty good way for discerning and acting towards the good. But in the case of a human whose action can and should be governed by other modes of practical cognition and considerations beyond what is pleasurable, the aim is not to ignore and override one's instinctive mind. Rather, the aim is to integrate and cultivate one's instinctive mind, to learn to pleasurably perceive objects and actions beyond the merely biological, including, for example, the morally good.

Adding pleasurable perception to the idea of inclination has several advantages. One, it explains why we cannot cultivate our inclinations simply by contemplating or judging that some action is good but must instead act. We cannot force ourselves to take pleasure in some object by simply contemplating it, and in the case of an action, we learn to perceive it pleasurably by

[100] Moss, *Aristotle on the Apparent Good*, 37-8.
[101] *Ibid*, 66.

acting and experiencing it as such.[102] This account can also explain why cultivating our inclination often begins with bribery, i.e. we often get ourselves to want to do something new by linking it to something else we already perceive pleasurably.

We often assume that habituation works simply by making something second nature, and I think this is an important element in habituation, but it cannot explain the important connection between habituation and inclination. It may be habit for me to take a certain route to work but that does not mean I am inclined to take that route in the sense of specifically feeling like taking the route. Instead, when habituation shapes an existing inclination or brings about a new one it is because we have begun to take pleasure in so acting.[103] This also explains the difference between unreflective action (action not done on the basis of explicit deliberation) and action done on an inclination: often when we act unreflectively we act on an inclination ("It just looked so good! No, I wasn't thinking!") but it's also possible to do demanding actions unreflectively, and in fact this explains why it's often best to avoid deliberation in such cases. In such cases, if we think too much about the action at hand, we may begin to perceive the right thing to do as painful and our instinctive mind will want to do the more pleasurable thing instead since the instinctive mind is guided by the goodness picked out by pleasure.

---

[102] Myles Burnyeat writes, "It turns out that Aristotle is not simply giving us a bland reminder that virtue takes practice. Rather, practice has cognitive powers, in that it is the way we learn what is noble or just" ("Aristotle on Learning to be Good," 73) and "I should now like to suggest that the prominence given to pleasure in these passages is the key to our problem about how practice can lead to knowledge" (76).

[103] Thus on my reading, when Aristotle famously says "we become just people by doing just actions, temperate people by doing temperate actions, and courageous people by doing courageous one," I read him as referring not to our acquiring virtue through brute habituation or mere association but rather to the fact that acting is one way in which we shape our emotions and feelings, and we shape feelings through action because acting gives us access to the goods embedded in certain activities (*Nichomachean Ethics*, Translated by C.D.C. Reeve. Indianapolis: Hackett Publishing Company, 2014, II.1).

2.2.3 However, although pleasure is the way non-human animals and human babies pick out what it good for them and what is *to-be-done*, they are simply aware of things as *to-be- done* and do not further grasp *why* the action is *to-be-done*. Does this undermine my insistence that inclinations are reason-laden? I think not. It's not just that pleasure is a pure imperative, an affective non-cognitive state of *to-be-doneness*. Pleasure is still a mode of practical cognition. Thus it has "an irreducibly cognitive aspect" and straddles the conative and cognitive.[104] To pleasurably perceive something is to be aware of it as *to-be-pursued*, but pleasure is also "a state by which one registers or discerns a property of the object: the relational property of contributing to one's flourishing."[105] Thus pleasure is a mode of cognition proper to the instinctive mind.[106]

In this sense, pleasure and pain are the way animals perceive reasons.[107] Of course, they do not perceive them as such. The point is rather that instinctive minds perceive acts as *to-be-done* because pleasure is a perception of goodness. The non-human animal or baby presumably doesn't grasp this connection, but in an adult human with thought we do. Think for example of non-practical perception, i.e. theoretical perception. The concepts humans acquire as they grow shape the range of objects that are perceptible for them. Thus for instance the human child or adult who has acquired the concept of a "rose" simply perceives the rose as opposed to a collection of colors and shapes that we associate with rose-ness. The analogous point holds true

---

[104].Moss, *Aristotle on the Apparent Good,* 43.

[105] *Aristotle on the Apparent Good*, 43. Consider also, "feeling pleasure in something *is* finding it good" (*Ibid,* 29).

[106] That pleasure is a mode of cognition explains why it is appropriate to try to free ourselves of certain pleasures "insofar as they involve distorted outlooks on value" (Brewer, *The Retrieval of Ethics*, 145).

[107] In *The Sources of Normativity,* Korsgaard writes, "Suppose for instance the animal needs nourishment. It perceives that by getting hungry. It finds this unpleasant and is moved to get something to eat. Don't be confused here: it is not that the pain is an unpleasant sensation which gives the animal a reason to eat. The animal has a reason to eat, which is that it will die if it does not. It does not know that it has that reason, but it does perceive it. The sensation in question is the sensation of hunger, not of pain. But an animal is designed to perceive and revolt against threats to the preservation of its identity, such as hunger. When it does that, it is in pain" (Cambridge: Cambridge University Press, 1996, 150).

in the practical realm: our maxims shape the range of objects we can practically perceive. As we grow and learn what it is to act-for-a-reason we don't just contemplate pure acts (at least not in the standard case), but rather acts-for-some-reason or acts-in-some-context.

Again, consider the non-human animal case instead, although I will be unfairly mechanistic to the animals in order to make the point. One way pleasure could work is that when an animal perceives some act type with pleasure the animal goes for that act type any time it's in a situation in which it's possible to act accordingly. Thus we could imagine an animal that gorges itself any time food is present because it regards food as pleasurable and thus as *to-be-eaten*, period.

But this is not how pleasure works in an adult human being. Our pleasure isn't judgment sensitive in Scanlon's sense; we take pleasure in and perceive as good things we might judge to be ill-advised for any number of reasons. But what we take pleasure in and in turn our inclination is nonetheless much more reason-laden than an animal who just keeps going for food any time it's available to eat. As discussed above, well-functioning adult humans aren't inclined to eat the cake on the plate of the child at the adjacent table in the restaurant, which is another way of saying that they don't take pleasure in it or perceive it pleasurably. Well-functioning adults don't take pleasure in the idea of eating cake even when they are already uncomfortably full after a large meal. Or setting aside what we think are the ideal set of inclinations to have with respect to cake, the broader point I am trying to make is that actions, not just act-types, are objects of practical perception for adult human beings. Adults (and older children, for that matter) can

practically perceive and thus pleasurably perceive *actions*, acts-done-for-a-reason and not just in act-types or ends of action.[108]

2.3 INCORPORATION

Thus far I have been trying to hold together two ideas that seem to be in tension. On the one hand, I want to insist that inclinations are reason-laden. By this, I don't mean that inclinations are authoritative, but that to have an inclination is to inhabit a perspective on the goodness of so acting, even if the agent is less than fully articulate about the nature of that goodness. On the other hand, I have insisted that the reason inclinations have motivational force and are often recalcitrant to our considered judgments is because they are an expression of our lingering animal agency, the agency we share with non-human animals. The apparent problem with endorsing both of these theses simultaneously is that animals do not have reasons in the same way humans do.

An objector might suggest that the solution to this tension is to embrace the dualism and say instead that instinctive mind simply sees acts as *to-be-done* on account of perceiving them with pleasure, but in creatures with dual motivation (creatures with both instinctive and deciding minds) the deciding mind automatically imbues reason to imperatives or interprets them as reason-laden, thus making inclinations appear both reason-laden and motivationally forceful.

---

[108] Relatedly, humans normally take pleasure in activities not just as pure sensations or act types but in light of their underlying meaning or value. Consider an example Brewer uses to make this point: "Considered from a purely physical point of view – e.g. as patterns of skin contact with particular textures and pressures – the same touch that is pleasurable in a voluntary sexual encounter might be entirely unpleasant in the context of a sexual assault. The pleasure of something as basic as a sexual caress, then, requires that one see the activity that causes the pleasure as good" (*The Retrieval of Ethics*, 134).

In some sense, I think this proposal is right. Clearly the fact that we experience inclinations as reason-laden is a result of the fact that we have a deciding mind. Furthermore, when it comes to the automatic and unconscious workings of the mind, we cannot sort out what is "instinctive" and what is "deciding" because when we turn our attention to the automatic and unconscious it ceases to be such. For this latter reason in particular, I will not try to say more about this idea. It may be that what I am calling the deciding mind "gives" reasons to inclinations unreflectively. But I find that the labels "instinctive mind" and "deciding mind" become less useful when we are talking about the unreflective anyways.

So it may be that we incorporate our inclinations unreflectively or automatically, subconsciously interpreting our inclinations in terms of the reason or worth of doing what we are inclined to do. This sense of incorporation is importantly different from the kind of incorporation referred to in what is called the "Incorporation Thesis," but the underlying question in both cases is whether there is some kind of gap to bridge between a desire as a motivationally forceful state and a desire as a perception of value on which we can act. After all, agency is not always automatic and unconscious. Often we deliberate about what to do, including deliberating about whether to act on an inclination, and we might well wonder whether when we engage in such deliberation we are considering proposals for action on which we could immediately act, or whether our inclinations are not as such appropriate grounds for action.

Grant that inclinations are in some sense reason-laden, by which I mean that we can inhabit our inclinations and that to do so is for us to have some kind of inchoate understanding of the goodness of acting on that inclination. It is not obvious that this is furthermore sufficient to allow us to act directly on inclinations without incorporating them. After all, on Schapiro's view

(and Korsgaard's too, although less centrally), inclinations are essentially provisional. Consider the following:

> Being inclined to φ is the starting point, rather than the end point, of practical deliberation. To see what I mean, consider the following. Once you have determined yourself to φ (through decision, or intention, or whatever volitional concept you prefer), you need a further reason to reopen the practical question you have just settled. Upon determining yourself to φ, it makes no sense to raise the question, 'should I φ?" unless your circumstances have changed, or you have discovered some reason to second-guess the integrity of your original deliberation. If you do not have such a reason, it makes no sense to do so. In fact, in reopening such a question without a further reason, you are arbitrarily second-guessing yourself, perhaps out of obsessiveness, or free-floating anxiety, or a lack of self-respect. The situation is different if you are merely inclined to do something. If, upon being inclined to φ, you simply ask yourself, "should I do what I am inclined to do?" you are not arbitrarily reopening a question you have already closed. This is so, even if you have no further reason to raise that question. Being inclined to φ opens a practical question; determining yourself to φ closes one. I call this the 'provisionality' of inclination's deliberative role. To be inclined is to be faced with a directive that is provisional in nature, subject to further deliberation.[109]

Can we act directly on an inclination? Schapiro says no, because inclinations are provisional, and in order to act (or act in the fullest sense) on an inclination we must incorporate that inclination. But why? Why does it make sense to raise the question "should I φ?" about an inclination and not about a decision? In the following, I will consider answers to these questions by examining arguments in favor of the view that we cannot act on inclinations directly, that is, arguments in favor of the Incorporation Thesis.

2.3.1 Recall that the label "Incorporation Thesis" originates in the work of Henry Allison. About Kant, he writes:

> … the intentional actions of a rational agent are never 'merely' the causal consequences of the agent's antecedent psychological state (or any other antecedent conditions for that matter) but require, as necessary condition, an act of spontaneity. The claim that this spontaneity is an ineliminable component in rational agency is what, for reasons that will become obvious, I call Kant's 'Incorporation Thesis.'"[110]

---

[109] *Feeling Like It,* 79.
[110] Allison, *Kant's Theory of Freedom*, 5.

And then, later on:

> … for Kant, an inclination or desire does not of itself constitute a reason for acting.[111]

This is a helpful starting point for thinking about why we must incorporate inclinations. What Allison gestures towards is the Kantian idea that to "act" on desire is for our action to reduce to a merely causal relationship between the agent and her movements because desires do not in themselves constitute reasons for acting.

But what does Allision mean by "of itself"? If this means that the mere fact that we have an inclination is not normally a reason to do anything, then the view is correct. The mere fact of wanting something does not settle that we should act on our desire. But we must be able to say more about why this is: if desires are provisional and require incorporation before we can act on them, is this because are inclinations are pure imperatives such that it would make no sense to act on them unless we specifically incorporate or take them up, giving ourselves reason to act according to the imperative in question? Or is it for some other reason?

I think the former cannot be the whole story. It seems possible to experience a pure imperative, and it is indeed true that we would need to incorporate such imperatives by discerning a reason to follow the imperative or straightforwardly coming up with reason to follow the imperative. However, as I emphasized earlier in the chapter, the paradigmatic case of inclination is not like this, and thus this motivation for the Incorporation Thesis is not helpful in understanding whether we can act directly on inclinations understood in the reason-laden sense.

2.3.2 Let's try again, considering what I shall call "The Heteronomy Motivation." Maybe the reasons inclinations are provisional, the reason we must incorporate inclinations before we can

---

[111] *Kant's Theory of Freedom*, 40.

act on them, is because when we act on inclination, we are determined by something within us that is not us. For example, when a non-human animal acts, there may be a sense in which he chooses the act-type he performs, but he does not choose his purposes. The purposes of his actions are given to him by instinct. To use an example Korsgaard discusses, when an antelope ducks into the tall grass upon seeing a lion, there is a sense in which she chooses the act (ducking into the tall grass) but her purpose or reason for ducking (to avoid the lion) is given to her by nature. In the case of an antelope, her acting on instinct is what makes the action hers, because she is her nature, and her nature gives her the instinct to duck. But on a familiar conception of rational agency, who we are is not our biology or given to us by nature, but what we choose. So although the antelope may exercise a kind of antelope-autonomy or autonomy-equivalent by acting on her instinct, when the human acts on instinct she is determined by her biology which is not an expression of her rational agency, let alone a chosen principle of action, and thus is not truly an expression of who she is.

However, I think this motivation for the Incorporation Thesis cannot be right either. It is true that inclinations are importantly biological on my account, but I wholly reject the idea that inclinations are always biological and instinctive in origin. It may be that we cannot get fully to the bottom of, for example, the inclination to eat chocolate cake without talking about the human instinct to eat food in general but especially sweet and fatty foods, but it is also a mistake to think that the inclination to eat chocolate cake reduces to a biological instinct to eat sweet and fatty things. It is also true that human inclinations can and should be cultivated over time, and in some cases this cultivation can result in very substantial changes, whether that means acquiring a new inclination to for example read obscure novels or shaping the instinctual inclination to for

example hoard good things for yourself into the inclination to enjoy good things together with others.

One might object in response that focusing too narrowly on biology is not the best version of the heteronomy motivation. Perhaps all we need is the observation that inclinations happen *to* us: we are passive with respect to them and cannot control them. States that happen to us cannot be expressions of who we really are, and so to act directly on an inclination is to be determined by something outside of us, or so the reasoning goes. The problem with this version of the heteronomy motivation is that more needs to be said in defense of the idea that states we cannot control are not expressions of who we are, and absent such a defense, this motivation begs the question insofar as it insists that we cannot act on inclinations directly because they happen to us when whether inclinations happen to us or are expressions of our agency is one of the very points under consideration.

2.3.3 On my view, inclinations stem from our animal nature. They are the output of our instinctive mind. Thus one resource for understanding why we must incorporate inclinations is to contrast animal action with human action: perhaps the reason we must incorporate our inclinations is because, absent incorporation, we are acting like an animal. Of course, this sense of "acting like an animal" is not base, not animal-like on account of its gluttony or promiscuity or violence, but animal-like in that is fails to be an act done for a reason. Still, the view needs motivating. Why might acting directly on an inclination fail to be a fully human action? What, precisely, is the difference between human and animal action? Korsgaard writes,

> Is a lioness who protects her cubs from a marauding male lion then acting for a reason, or rationally? Perhaps we do not know exactly how to think about the lioness's mental

> representations, but she is an agent, not a mechanism, and it seems clear that there is some sense in which she does what she does in order to protect her cubs.[112]
>
> We may insist that there is something different in the human case, something that does involve the faculty of Reason. The human is aware of the reason as a reason; she identifies the good-making properties of the action under the description 'good' or 'reason' or 'right,' or some such normative description. She does not act merely in accordance with a normative consideration but *on* one.[113]

Unlike animals, humans act *on* reasons, not just in accordance with them. So if we are to act directly on an inclination, without incorporating it, we must in so doing be acting on a reason.

What does it mean to act on a reason? We should reject outright the idea that acting on a reason requires a high degree of reflection or conscious endorsement. Clearly we can act automatically for a reason.[114] So acting on a reason does not require us to be consciously aware of our reason or ourselves as acting on a reason prior to or during the action. But again, we can ask: what does it mean to act on a reason? Elsewhere Korsgaard emphasizes that it means to choose a maxim that can be universalized, but I think we can set aside the emphasis on universality and even maxims for the time being.[115] The more basic point Korsgaard is trying to

---

[112] "Acting for a Reason" in *The Constitution of Agency,* Oxford: Oxford University Press, 2008, 213.
[113] *Ibid,* 213-214. Emphasis added.
[114]The contrast Pamela Hieronymi draws between "organized thought" and "explicit deliberation" is relevant here. She writes, "… reasoning is organized thought, not explicit deliberation. Explicit deliberation is a conscious activity that unfolds across time. Organized thought need not be. I can take reasons to bear on, or to settle, a question without explicitly deliberating about that question," ("Reasoning First," 355). Elsewhere, Hieronymi gives an example that helpfully illustrates the point: she writes, "A recent on-line poll asked whether Michael Vick deserves to play football again. I can answer immediately, without deliberation. I believe I can also provide reasons, and I believe the reasons I provide may in fact be the reasons for my judgment (and not merely a 'post-hoc' rationalization). But my judgment followed no deliberation upon those reasons. (I think this is a deep point: reasons can explain judgments even when they do not precede them in occurrent thought)" ("The Will as Reason," 218, footnote 20). Importantly, this is a point Korsgaard affirms. She writes, "acting on a rational principle [as opposed to being merely caused] need not involve any step-by-step process of reasoning, for when a principle is deeply internalized we may simply *recognize* the case as one falling under the principle, where that is a single experience" (*Self-Constitution,* 107).
[115] Consider the following passage from Korsgaard's essay "Self-Constitution in the Ethics of Plato and Kant": "So here is the free will, completely self-governing, with nothing outside of it giving it any laws. And along comes an inclination, and presents the free will with a proposal. Now inclinations, according to Kant, are grounded in what he calls 'incentives,' which are the features of the objects of those inclinations that make them seem attractive and eligible. Supposed that the incentive is that the object is pleasant. The inclination says: end-E would be a very

make is just that if we are to act on an inclination, our action cannot just be the causal output of an inclination, the effect the inclination had on us. The action must instead stem from our own self-determination, and this means the action must derive from not from our pure attraction to the object of inclination but from the principles that tell us what to do in the face of an attraction to some end or act-type and determine what counts as a reason for us.[116]

However, to act on a reason is not just to act according to your principles, since animals have principles too. The instincts given to animals by nature determine how they respond to

---

pleasant thing to bring about. So how about end-E? Doesn't that seem like an end to-be-produced? Now what the will chooses is, strictly speaking, actions, so before the proposal is complete, we need to make it a proposal for action. Instrumental reasoning determines that you could produce end-E by doing act-A. So the proposal is: that you should do act-A in order to produce this very pleasant end-E. Now if your will were heteronomous, and pleasure were a law to you, this is all you would need to know, and you would straightaway do act-A in order to produce that pleasant end-E. But since you are autonomous, pleasure is not a law to you: nothing is a law to you except what you make a law for yourself. You should therefore ask yourself a different question. The proposal is that you should do act-A in order to achieve pleasant end-E. Since nothing is a law to you except what you make a law for yourself, you ask yourself whether you could take that to be your law. Your question is whether you can will the maxim of doing act-A in order to produce end-E as a law" ("Self-Constitution in the Ethics of Plato and Kant," in *The Constitution of Agency*, 109).

This raises a related but slightly different motivation for the Incorporation Thesis than the one I am considering here. On this motivation, the reason we cannot act directly on an inclination is because it is not properly speaking an object of choice: it is not yet in the form of something we can choose as our action. However, I think we should be skeptical of the idea that an inclination does not present us with an object of choice. If inclinations were pure imperatives, this might be true, or if inclinations were simply vague attraction to some end with no particular action attached to the attraction, this might also be true. But inclinations as I've defined them are not like this. Inclinations involve both the sense that an action is to-be-done and they are reason-laden, which is to say that inclinations carry an inchoate representation of the reason for doing something. In this context, we might then say that even though inclinations are not maxims in propositional forms, there is a maxim inchoate in an inclination. My view here is influence by Talbot Brewer's theory of desire, and by Kyla Ebels-Duggan's exploration of having reasons for value commitments about which we are inarticulate ("Beyond Words: Inarticulable Reasons and Reasonable Commitments," *Philosophy and Phenomenological Research* 98, no. 3 (2019): 623-641). Brewer writes, "[Desires] present us with a deliberative problem in the straightforward sense that we must decide, when confronted with them, whether it really makes sense to move our bodies in the way that they incline us to move them. Bu they also present us with an inchoate sense of how or why it would count as good to act as they incline us to act, and this provides us with the richer and more interesting practical problem of interpreting and assessing our sense of the point of these actions" (*The Retrieval of Ethics*, 29). Thus I think inclinations can be objects of choice, although I am of course setting aside for a later time the further question of whether we must have a universalizable principle or maxim in order to count as acting.

[116] An incentive, Korsgaard writes, "is a motivationally loaded representation of an object. It presents the object as desirable or aversive in some specific way – as a thing to be eaten, to mated with, avoided, fled, investigated, defeated or what have you. The principle determines, or we may say describes, what the animal [human or non-human] does, or tries to do, in the face of the incentive" (*Self-Constitution*, 109).

incentives and serve as their principles. Furthermore, since an animal's movement is determined by her own principles and thus her own self, an animal can be said to act rather than simply move or react. In this sense, animals can be said to "have reason" for acting.

What makes human action unique is not the ability to have reason but the ability to evaluate reason, and this makes the difference between acting in accordance with a reason and acting on a reason.[117] Korsgaard writes,

> [Human beings] are aware, not only *that* we desire or fear certain things, but also that we are inclined to act in certain ways on the *basis* of these desires or fears. We are conscious of the potential grounds of our actions, the principles on which our actions are based, *as potential grounds*. … [this] sets us a problem that the other animals do not have. For once we are aware that we are inclined to act in a certain way on the ground of a certain incentive, we find ourselves faced with a decision, namely, whether we should do that. We can say to ourselves: "I am inclined to do act-A for the sake of end-E. But should I?[118]

Although Korsgaard is not generally known for having a reason-laden view of inclination, what I like about this passage is that it leaves open a space for inclination to be reason-laden. Here, Korsgaard does not reject inclination because it is a pure imperative with no rational content or

---

[117] Although I will go on to draw a different conclusion than she does, on this point Schapiro and I are in agreement. She writes, "I am also going to reject the idea that the salient difference is the capacity to double-check. For there may indeed be a sense in which a creature of instinct can double-check its teleological thinking. A cat, I imagine, can double-check to see whether that mousey-to-be-eatenness really is mousey-to-be-eatenness. But if the cat's mind is as I have characterized it, what it cannot do is raise the question whether that mouse is really a mouse, conceived apart from its to-be-eatenness. *Nor can it raise the question whether, given that that thing is a mouse, eating it is appropriate. It cannot deliberate in these ways, because it cannot view objects in the world non-teleologically, and then take responsibility for deciding how to respond in relation to them*. The deciding mind, by contrast, is forced to take on this responsibility. The world it encounters does not have practical necessity built into it. In this sense, the deciding mind is free, and bears the burden of its freedom" (*Feeling Like It*, 130). Emphasis added.

[118] *Self-Constitution*, 115. On the next page, she goes on to add, "It is within the space of reflective distance that the question whether our incentives give us reasons arises. In order to answer that question, we need principles, which determine what we are to count as reasons. Our rational principles then replace our instincts – they will tell us what is an appropriate response to what, what makes what worth doing, what the situation calls for. And so it is in the space of reflective distance, in the internal world created by self-consciousness, that reason is born" (116). Although they are strange bedfellows, on this point I think Korsgaard and Alasdair MacIntyre agree. In his book *Dependent Rational Animals*, he writes that in growing up from infants to adults, we go from "merely having reasons to being able to evaluate our reasons as good or bad reasons and by so doing to change our reasons for acting and in consequence our actions" (Chicago: Carus Publishing Company, 1999, 71-72).

some other kind of brute force at work within us. To the contrary, what Korsgaard says here invites the idea that the reason we need to incorporate inclinations, or more specifically in this case, choose our principles, is exactly because inclinations have rational content.[119]

Consider a specific example, the now-familiar case of being inclined to eat chocolate cake. When I have this inclination, when I am inclined to eat the cake for the sake of its chocolatey deliciousness, I can ask: but should I? And when I do eat the cake, even if my action is automatic and not consciously deliberated, my action answers or embodies my answer to that question, even if I later come to judge that I answered the question poorly. This means that we do not need to worry that acting on an inclination is a form of acting like an animal, having reason but not acting on reason.

The key difference between human and animal action then is not conscious deliberation, but the capacity to evaluate the grounds of one's action, to ask "But should I really φ?" about any given action, and over time to cultivate one's principles of action, the principles that dictate what counts as a reason.[120] In other words, acting on an inclination is a form of acting on a reason when the actor is human, a creature with the capacity to evaluate her grounds for action. This is true even in the case where the actor does not consciously consider whether her grounds for action are good. To use Scanlonian language that I will revisit in greater depth in a later chapter, we might say that the reason acting on an inclination counts as acting on a reason instead of acting merely in accordance with a reason is because I am answerable for so acting, even if the

[119] This is again the point I have repeatedly made in this chapter with respect to inclinations: the reason we can deliberate about acting *on* inclinations rather than in light of them is because they are reason-laden.
[120] Korsgaard places an especially strong emphasis on choosing one's principles, for example writing, "But an animal does not choose the principles of his own causality – he does not choose the content of his instincts. We human beings on the other hand do choose the principles of our own causality – we choose our own maxims, the content of our principles" (*Self-Constitution,* 108).

action was done on inclination. Even ill-advised actions performed on the basis of inclination represent my answer to a normative question, since I am an animal with the capacity to ask and answer such questions, and in this respect I am importantly different from the non-human animal acting on instinct.

This has the possibly surprising result that incorporation as an activity is not a matter of humanizing one's inclination in the moment of choice, not a discrete task we must undertake with respect to a particular inclination, but a task of habituation and interpretation across time. Incorporation understood as an activity refers to the process of cultivating inclinations across time in light of our capacity to evaluate reasons and ask normative questions. Or, speaking from the first-person deliberative perspective, the point is that we do not need to examine and consciously incorporate every inclination as such every time we are aware of an inclination. Rather, the important thing is that inclinations are the kind of thing we *can* subject to examination and questioning if we choose, and in this sense, they are already candidates for incorporation into our rational agency. The more pressing task from the practical perspective is cultivation, but that is a theme for a different chapter.

2.4 PROVISIONALITY REVISITED

Still, Schapiro might object at this point that I'm failing to notice that inclinations are provisional. One primary reason for the Incorporation Thesis was the idea that inclinations are provisional in a way that, for example, intentions are not, and so we need to say something about how we transform inclinations from provisional states into decisions. Thus one might object that

although inclinations can be evaluated and we can ask whether the inclination provides us with good grounds for acting as we are inclined, this fails to address the point that it's appropriate to raise questions about one's inclinations in a way that it's not appropriate to question one's intentions. But why? Why is it appropriate to question inclinations in a way that it's not appropriate to question intentions?

Non-human animals are deeply immediate. Perhaps some non-human primates have the ability to think about the future and what would be in their best interest over time, but it seems clear that most animals lack the capacity to think about their future interests. Evolutionarily speaking, the instincts of animals may be in their best interests because having roughly those instincts is what enabled the species to survive, but from the perspective of the animal himself, his purposes are "local and concrete."[121] Especially in a species with deciding minds and a social web built around the capacities of the deciding mind, the immediacy of instinct is what makes it prone to error. To take the case of an inclination for chocolate cake, our instinctive mind might see the chocolate cake as chocolatey and delicious and thus *to-be-eaten* and might fail to represent the situation as one in which the cake is *to-be-avoided* because it is actually the property of the roommate who plans to bring it to a party tonight.[122]

---

[121] "I am not at all inclined to deny that the other intelligent animals do things on purpose, but I would expect these purposes to be local and concrete—to eat something, mate with someone, avoid punishment, have some fun, stop the fight—but not to do what is best for themselves on the whole" (Korsgaard, "Morality and the Distinctiveness of Human Action," In *Primates and Philosophers: How Morality Evolved.* Edited by Stephen Macedo and Josiah Ober. Princeton: Princeton University Press, 2006, 102).

[122] I hesitate to describe the situation this way, since I think instinct in human beings actually can take on social objects and thus for instance could for instance highlight the salient features of this situation as chocolatey! Delicious! And also Warning: possible social violation! In other words, I don't want to give the impression that instinct is narrowly biological and only takes objects like *delicious-food-to-be-eaten*. In a social species, our inclinations can also be social, and we might for instance be eager to please and make other people like us or eager to avoid offense and retributive attitudes.

So one reason it is acceptable to challenge one's inclination, asking "Should I really eat this chocolate cake?" even absent a special reason to think that one's inclination is wrong, is because inclinations in general are prone to error. Inclination and especially uncultivated inclination is very often a poor guide to action.

Furthermore, I think substantive norms of agency also play a significant role in our diverging attitudes on questioning inclinations and decisions. In part, this is to repeat the concern raised above about inclination's propensity to error, since to say that we should treat inclinations are provisional on account of their tendency to misguide is not to say that they are unsuitable as such for action but rather than it is generally a wise principle of action to treat one's inclinations as provisional in light of this. However, avoiding error is not the only substantive norm at play here. There are people who think that it's generally misguided and alienating to challenge one's inclinations. Furthermore, although we normally regard a person who constantly questions their decisions as flighty, anxious or indecisive, we can also redescribe this behavior as spontaneous and flexible. In order words, I think our norms for challenging inclination vs. decisions are for the most part not norms stemming from the nature of moral agency, but are rather substantive moral norms about how we should relate to inclinations and decisions.

However, the point goes even deeper than this: it's not just that inclination generates decisions with a high degree of error. To say that inclinations are provisional decisions is not just to say that they are decisions, but there's a rather high chance that the decision might be wrong, and so it should be double-checked. Rather, labeling inclinations as provisional decisions is supposed to capture the deeper point that inclinations do not determine the deciding mind. I have long since dismissed the view that inclinations are forces acting upon us that might determine

our action if they reach a certain sufficient strength. But even on the view that inclinations involve the appearance of reasons, the capacity of the deciding mind just is the capacity to call into question whether or not a given consideration settles the question of whether to φ. To this extent, I do agree with Korsgaard and Schapiro that a certain kind of freedom is definitive of the deciding mind, or in this case we might also say rational will. If we regard inclinations as settling practical questions such that we cannot reopen or raise practical questions, then we are simply denying that we have a deciding mind at all.

So take a particular example: say that I have the inclination to buy myself coffee, and I am inclined to drink coffee because I find it delicious (as opposed to being inclined to drink coffee because I am restless and need an excuse to leave my work and take a break, etc.). I ask myself, "But should I really buy coffee?" In so asking, I assume that I have the option to reject this inclination. (In contrast, it would not make sense to ask myself, on the precipice of a sneeze, "But should I really sneeze?") If I decide to buy coffee, it may be for the reason embedded in my inclination, that it is delicious. I might also decide to buy coffee, but do it for a different reason: perhaps in deliberating I realize that I haven't had any coffee yet today and I'm going to need some caffeine in order to make it through the long afternoon without dozing off. Or I might forgo my inclination for coffee, because even though coffee is delicious and caffeine would help me stay awake in the afternoon, my cardiologist has told me to stop drinking so much caffeine and another cup of coffee would put me over my daily limit.

So in this scenario, when I ask "But should I really buy coffee?," I ask this question from the perspective of the deciding mind and from this perspective regard it as unsettled what I should do. However, in challenging or double-checking the imperative put forth by my

instinctive mind, I do not thereby regard my inclination as having no reason-giving force whatsoever. And crucially, I do not think any kind of special transformation or incorporation is required in order for me to act on the inclination.[123] It might be that I transform my inclination or choose an entirely different principle of action as a result of deliberating, but this is not a requirement of acting. On my view, to say that inclinations are provisional is simply a reflection of the fact that we are creatures with deciding minds, and because of this, when we act, we settle practical questions. An inclination however results from our instinctive agency, and although we can act on an inclination and thus settle our will in accordance with that inclination, to have an inclination is not yet to have an intention. Why not? Because we have the capacity to ask, "But should I really act thus?" about any given inclination.

2.5 EVALUATING THE ACCOUNT

In summary , my view is that inclinations are provisional decisions. This way of describing things is misleading insofar as non-human animals do not make decisions, but I nonetheless prefer to call inclinations provisional decisions because in a being with a deciding mind, inclinations function as would-be decisions. To experience an inclination is to see something as *to-be-done* (and thereby to be motivated to do it) and furthermore have at least a partial, if inarticulate, understanding of the goodness of so acting. In other words, inclinations aim at determining our action. However, clearly inclinations are not identical with decisions

[123] It may be fair to say that acting on an inclination requires *interpretation*, i.e. the interpretation of rational content embedded in the inclination. But this seems clearly distinct from the sort of "humanization" Schapiro for instance thinks we need in order to transform our inclinations.

proper, in part because we can have an inclination without yet having decided what to do, but also because we have control over our decisions in a way we do not over our inclinations.[124] Thus I refer to inclinations as "provisional" although my use of this word is weaker than Schapiro's. On my view, "provisional" primarily serves to flag that inclinations are not decisions, although when we act on an inclination they become such. At this point I want to briefly return to the problem motivating chapter 1 to see how this theory fares.

2.5.1 An adequate account of inclination must walk a very fine line. On the one hand, it must allow for recalcitrance. The problem with for instance Scanlon's account is that he assimilates desire too closely to practical thinking, and thus it becomes hard for him to say why certain desires, i.e. inclinations, have such motivational force, and furthermore it becomes difficult to explain why some desires are thoroughly recalcitrant in the face of judgment. On the other hand, neither do we want to make inclination a totally foreign force within the human agent. Inclination has to be something we can act on and something that can help make our action intelligible.

Schapiro's three criteria for an adequate theory of inclination nicely capture what's needed in order to navigate this tension. By way of reminder, she holds that an adequate theory of inclination must meet the practical thinking, non-voluntariness, and asymmetric pressure constraints.[125] So how does inclination as provisional decision fare with respect to these constraints?

Practical Thinking: Throughout this chapter I have emphasized that inclinations are reason-laden. I mean for this phrase to capture that inclinations are not necessarily conscious

[124] I will return to this topic in much greater depth in chapter 4.
[125] *Feeling Like It,* 31-2.

awareness of reasons as such or awareness of reasons in propositional form, but that an inclination represents the goodness of acting a certain way. In light of this, I think it's clear that my view meets the practical thinking constraint. In fact, I think practical thinking is even stronger on my view than Schapiro's, since I am happy to say that the conclusions of the instinctive mind are (*prima facie*) reasons for the deciding mind and that the moves of the instinctive mind are potential moves in deliberation even if they can and sometimes ought to be called into question or double-checked by the deciding mind.

Non-Voluntariness: Recall that the non-voluntariness criteria for inclination is as follows: "you cannot be inclined at will. You can determine yourself to φ, and thereby φ, but you cannot determine yourself to be inclined to φ, and thereby be inclined to φ. You can, however, do something we call 'cultivating' your inclinations over time."[126] Are provisional decisions under our control? No, because what we see as *to-be-done* and thus what we are inclined to do is a result of what we perceive pleasurably, and this is not something we can do at will. I will furthermore argue in a later chapter that neither are inclinations judgment sensitive attitudes, which introduces an even deeper sense in which they are not under our control.

Asymmetric Pressure: what about asymmetric or motivational pressure? This is, of course, the primary flaw of Scanlonian theories of desire: they have a difficult time explaining why it is hard to act against an inclination and easy to act on one. Intuitively, it is tempting to appeal to pleasure here, and certainly it's intuitive to think that the reason it's hard to resist inclinations is because inclinations are for pleasure and thus very strong. But this view should be rejected for multiple reasons. One, it is not right to say that inclinations take pleasure at their

[126] *Feeling Like It*, 31.

object in every case or that pleasure is our reason for acting anytime we act on an inclination. Two, this view assumes a brute force conception of desire on which certain desires are "very strong forces" in the face of which we are helpless. Thus, in order to understand asymmetric or motivational pressure, we will need to take a different approach.

Humans and non-human animals are capable of acting even when they lack a deciding mind. To use the human case, toddlers or young children with no (or very little) capacity to ask themselves whether they really ought to do something are nonetheless capable of intelligent and purposive action. If a toddler sees some cake that she would like to eat and grabs a fistful and puts it in her mouth, it's not that her limbs are flailing at random and she happened to move her hand first cake-ward and then face-ward. Neither is the toddler engaged in mere stimulus-response when she eats cake: she is not like a motion-sensitive light bulb that's turned on by movement, triggered to eat cake every time the stimulus of cake presents itself. This being said, of course a toddler eating cake is different from an adult eating cake, and we do and should respond differently to these two actions. But my point here is that our capacity for a basic form of action is not something we lack and only acquire once we grow or grow into a deciding mind.[127] Thus children are capable of motivation (where motivation is understood to involve something beyond mere movement and flailing limbs) even without exercising a deciding mind, and the capacity to motivate oneself without the intervention of the deciding mind is not something we lose somewhere along the way to adulthood.[128]

---

[127] This is a point made by MacIntyre in *Dependent Rational Animals*, especially chapters 2, 3, and 4. At the end of chapter 4, he writes, "What I am suggesting then is that adult human activity and belief are best understood as developing out of, and as still in part dependent upon, modes of belief and activity that we share with some other species of intelligent animal" (41).

[128] Perhaps we should speak of the instinctive mind not as our "inner animal" but our "inner toddler" or "inner child."

In other words, the instinctive mind and the deciding mind are both capacities for action. Of course, the instinctive mind cannot raise practical questions; it cannot ask "But should I really φ?" or the more open ended "What to do?" But focusing too much on the inability of instinctive mind to pose practical questions of this sort runs the risk of obscuring that the instinctive mind is perfectly capable of guiding an agent purposefully through the world. Furthermore, in the case of human beings who are both rational and animal, our deciding mind can raise practical questions and answer them, but our deciding mind can also turn to the instinctive mind for answers, although it is not bound by the instinctive mind. This can happen in a forward-looking sense, when we ask what we ought to do and act as we're inclined, or in a backward-looking sense, when we look at an action done on instinct and ask why we acted thus and our inclination both explains and justifies our action.[129]

The reason I emphasize that both the instinctive and deciding mind are capacities for settling questions, however, is not to over-intellectualize or rationalize the instinctive mind, but to emphasize that the motivational force of inclination is just the force of our own motivation. When I am trying to train for a half marathon and I'm inclined to lounge on the couch instead of following my training plan, my tempting desire to give up on my training plan isn't a foreign pressure acting on me. It's not a strong wind, it's not even literal movement in my body outside the purview of agency, like digestion. The force or motivational pressure of my tempting desire is the force of my own agency. Now, importantly, I do not think the force just is the force of my practical thinking or deciding mind. If that were true, then temptation could only ever be a

[129] More accurately, "purports to justify," since of course we might act badly. So if for instance I blurt out one of your deepest secrets because I want to seem interesting in a conversation with very witty and cool people, my action is justified in the sense that I can be held responsible for it as an action. It is not mere behavior like a reflex or digestion, but it is also not justified in the stronger sense of being a good action.

conflict between two good but incompatible courses of action, like wanting to vacation in two equally wonderful places but only having time to visit one. Rather, we need to maintain some space for the recalcitrance of tempting desire. Schapiro might say we need a bit more dualism in our account. But this is the advantage of thinking of the instinctive and deciding minds as two very different capacities for settling practical questions, or two different routes for action. The division within our own motivation, our dual-capacity for determining ourselves to act, means that one part of us can be moved towards one end (for instance, lounging on the couch) when the other part has already decided to pursue a different end (training for a half marathon). In other words, Schapiro is right to describe inclination as a kind of two-in-oneness.

So the instinctive mind is capable of moving human and non-human animals to action. In the case of a wholly instinctive animal, this means that the conclusions of the instinctive mind play the role of decisions. In the case of an animal who has both the capacity to see the world teleologically and respond from instinct and the capacity to pause and call a given response into question, the decisions of the instinctive mind are provisional decisions. This means that from the perspective of the deciding mind, these decisions have force, they appear as imperatives. After all, were we a creature wholly of instinct, the decision would not be provisional but would simply be a seamless transition of instinct responding to a representation of the world in purposeful movement.

However, in the case of a human being, inclinations are not decisions proper but are rather provisional decisions, because from the perspective of the deciding mind, our inclinations do not determine us. Our deciding mind is the capacity to ask whether we ought to φ and settle that question, not the capacity to look on and observe as inclination moves us to act, and so to

say that inclinations determine the deciding mind is just to deny that we have a deciding mind. I will take for granted that we do have the capacity to ask and then settle whether or not to φ, i.e. that we do have a deciding mind. In an animal with a deciding mind, something important changes with respect to the role of instinct. Instinct is one way we can be moved to act, but it is no longer the sole determinant of our action.

What then accounts for asymmetric pressure? The pressure arises from tension between the inclination and deciding mind.[130] When we have an inclination to φ but then ask the question "Should I really φ?", or when we have an inclination to φ even though we've already decided not to φ, our inclination purports to guide our action. Since our instinctive mind is capable of moving us to act all on its own, it's as if our instinctive mind has already made a decision and is frustrated to be hanging around waiting for the deciding mind to catch up. After all, the thoughts of the instinctive mind are how we would "decide" (or more precisely, act) if we lacked a deciding mind. And the instinctive mind does not cease making decisions when it belongs to a creature who can also distance herself from the instinctive mind's perspective and ask whether she really ought to do what the instinctive mind proposes.

[130] When my cat Albie sees slinking-cat-to-be-chased and so has an incentive to chase another cat, Char, I take for granted that he does not have the capacity to question whether he really should chase Char. But this also means that there's not a sense in which it's easy for Albie to act on his incentives and difficult for him to avoid them. He just acts. Thus although we sometimes speak as if inclination or desire is intrinsically forceful, but in fact inclinations are only "forceful" in the sense of making it easy or difficult to act when they are considered from the perspective of a deciding mind asking how to act. Thus there's a sense in which only humans have inclinations, whereas wholly instinctual animals simply have instincts or incentives.

3. Rationally Resisting Temptation

In the previous two chapters I've offered an account of why it is hard to resist temptation. With that account in hand, however, we might turn towards what we need to do in order to resist temptation. We know that it is easy to act on tempting desires and difficult to resist them, but what are we supposed to do in the face of this fact? Although I've repeatedly claimed that the deciding mind is not determined by temptation, thus far I haven't provided very much hope for carrying this out in practice. To the contrary, having insisted that tempting desires are provisional decisions of the instinctual mind and thus expressions of our own agency, you might think that this encourages a kind of passive acceptance of one's tempting desires. "Ah well," one might say. "My instinctive mind wants to binge watch a show on Netflix because it's a relaxing and soothing thing to do and far less anxiety provoking than writing. My instinctive mind is just trying to take care of me *qua* animal. Might as well just go along with its instinctive wisdom."

In this chapter I will try to forestall this worry with a discussion of resisting temptation. One obvious starting point when it comes to resisting temptation is resolution, since temptation and resolution appear to go hand-in-hand. Folk wisdom tells us that if we have any hope of achieving something difficult, we will have to make a resolution to do the thing in question. We don't stumble by accident into running ultramarathons, being good romantic partners, or writing books: we first make resolutions about these things.[131] After all, in the absence of a resolution to do a difficult thing, we can simply change our mind and decide to do something else at any point. Resolutions, in other words, draw a line between intentions that we can revise on a whim and

[131] Of course, many cases of romantic partnership involve something beyond a resolution: a promise to another person. However, it is also common and natural to make resolutions regarding important relationships.

those that we cannot rationally revise without special justification.

But are resolutions appropriate philosophical fodder? Might they be instead a matter for empirical study? After all, one natural question to raise about resolutions is whether they're of any *use*. Is making a resolution actually an effective way of accomplishing one's goals? This is clearly an empirical question, and an important one at that. However, prior and adjacent to this empirical question about resolution are philosophical questions. They are:

1. What is a resolution?
2. Why should we grant resolutions rational authority over our actions, given that the moment of temptation involves the evaluation that it would be best to act on the temptation?
3. Why should we expect resolutions to be of any use in resisting temptation?

The first question matters but is in many respects the least interesting of the three questions. It's clear that resolution is some kind of extra-committed intention: I might form the intention to have fried rice for lunch because I like fried rice and I have leftovers readily available in the fridge, or I might form the intention to have fried rice for lunch because I'm committing to a gluten free diet and fried rice is the only gluten free meal I have available in the home. In the case where I form the intention to eat fried rice just because I like it and it's available, I clearly remain open to changing my mind. Perhaps lunchtime rolls around and I decide to have quesadillas instead. But in the second case, where I resolve to have fried rice, it's clear that my intention carries with it the further thought that I should not change my mind or go back on this intention unless there are extenuating circumstances.

When we ask what a resolution is, we are trying to understand this element of "extra commitment." Does the extra commitment take the form of a second-order intention, a first-order intention plus the intention to not reconsider? Or an intention-desire pair, where we have a first

order intention and a desire to not reconsider? The reason I say this question is perhaps the least interesting is because it seems to me that philosophers are most likely to answer question 1 in light of their answer to question 3. I will say more about question 3 shortly, but if our focus is on giving an account of why a resolution would help an ideally rational agent to accomplish her goals (without committing oneself to empirical claims about how actual, non-ideal agents manage to accomplish their goals), then it is natural to allow our answer to question 3 to drive our answer to question 1, as is I think the case with the two accounts of resolution I will consider below.

The second question is the one that preoccupies most of the existing philosophical literature on resolutions.[132] For example, the following case posed by Bratman is representative of the cases motivating this literature:

> Consider Ann. She enjoys a good read after dinner but also loves fine beer at dinner. However, she knows that if she has more than one beer at dinner she cannot concentrate on her book after dinner. Prior to dinner Ann prefers an evening of one beer plus a good book to an evening with more than one beer but no book. Her problem, though, is that each evening at dinner, having drunk her first Pilsner Urquell, she finds herself tempted by the thought of a second: For a short period of time she prefers a second beer to her after-dinner read. This new preference is not experienced by her as compulsive.[133]

The question is whether it would be rational for Ann to act on a prior preference given that her preferences have now changed. However, for the most part I will set aside this problem in this chapter: I do not assume that our preferences in a given moment determine what is most rational

---

[132] The current philosophical literature on resolution is concerned primarily with this question. See for instance: Chrisoula Andreou, "Temptation, Resolutions, and Regret," *Inquiry* 57, no. 3 (2014): 275-292; Michael Bratman, "Toxin, Temptation, and the Stability of Intention," in *Faces of Intention*, (Cambridge: Cambridge University Press, 1999), 58-90; Michael Bratman, "Temptation Revisited," in *Structures of Agency*, (Oxford: Oxford University Press, 2007), 257-282; Michael Bratman, "Temptation and the Agent's Standpoint." *Inquiry* 57, no. 3 (2014): 293-310; Sarah K. Paul, "Diachronic Incontinence is a Problem in Moral Philosophy." *Inquiry* 57, no. 3 (2014): 335-355.

[133] Toxin, Temptation, and the Stability of Intention," 74.

for us to do in that moment. I will not provide a full argument for this claim here, but in brief, it seems to me that this way of thinking uncritically adopts many of the assumptions of, for instance, rational choice theory in economics, and moral philosophers should not feel beholden to adopt assumptions or at least not adopt them uncritically.

The final question, then, is the primary object of my attention in this paper: why should we expect resolutions to be of any use? As I've already mentioned, this question is closely related to empirical questions about the efficacy of resolutions. However, there is space to consider this question philosophically. We might for instance hold fixed the idea that resolutions are at least of some use in resisting temptation and then attempt to give an account of why resolutions help rational agents resist temptation. Or, we might think in terms of an ideally rational agent: not so ideally rational that they do not experience temptation at all, but ideally rational enough to respond to temptation in a fully rational way. Thus, while a less rational agent might need to use commitment devices[134] or other such strategies that tackle temptation "sideways," as it were, you might think that resolution is a way for a rational agent to face their temptation head on.

[134] The idea of a commitment device is most often used in economics, psychology or public health. Formally speaking, "a commitment device [is] an arrangement entered into by an agent who restricts his or her future choice set by making certain choices more expensive, perhaps infinitely expensive, while also satisfying two conditions: (a) The agent would, on the margin, pay something in the present to make those choices more expensive, even if he or she received no other benefit for the payment, and (b) the arrangement does not have a strategic purpose with respect to others" (Gharad Bryan, Dean Karlan, and Scott Nelson, "Commitment Devices," *Annual Review of Economics* 2, no. 1 (2010): 673). A classic example of a commitment device is Odysseus tying himself to the mast of the ship in order to hear the sirens.

3.1 RESOLUTION

In his book *Willing, Wanting, Waiting*, Richard Holton introduces the idea of a resolution, writing, "Resolutions serve to overcome the desires or beliefs that the agent fears they will form by the time they come to act, desires or beliefs that will inhibit them from acting as they now plan."[135] So the purpose of a resolution is to hold off beliefs and desires that might prevent me from acting as I now intend. But what is a resolution? According to Holton, a resolution is a pair of intentions, one first-order and one second-order. It is "both an intention to engage in a certain action, and a further intention to not let that intention be deflected."[136] So someone who resolves to quit smoking forms a first-order intention to quit, and a second-order intention to stick with the first-order intention. This second-order intention doesn't generate new reasons (we would have a boot-strapping problem if it did), but instead entrenches the first-order decision in response to reasons for not-reconsidering that decision.[137]

There is something intuitive about this view of resolution, since it captures the phenomenology of resolving: deciding to do something, and furthermore deciding to not let yourself be distracted or deterred from your goal. However, in the following I will consider an objection raised against this account of resolution by Alida Liberman in her article "Reconsidering Resolutions."

3.1.1 Liberman argues that resolutions understood as a two-tier intentions are not effective in resisting temptation. She summarizes her objection as follows:

It seems that the second-order intention should succumb to the same temptation

[135] Holton, *Willing, Wanting, Waiting*, 77.
[136] *Ibid*, 11.
[137] *Ibid*, 146.

> to which the first-order intention is susceptible … Why do the very same considerations that tempt you toward watching yet another episode of your favorite TV show – say, your burning desire to find out what happens next, and your aversion to working – not *also* tempt you to reconsider your resolution to turn off the TV and get to work on your paper?[138]

Liberman develops this objection through an argument she calls "Temptation Transmission." The argument comes in two stages. The first stage lays out two background principles,[139] and the second gives the actual argument which consists of three premises.[140] The details of her argument are compelling, but for our purposes we need only consider her conclusion. She writes,

> Since the temptation to Φ leads to the appearance of an *equally strong* reason to abandon the first-order intention to Φ *and* to abandon the resolute second-order intention, the second-order intention cannot do any meaningful work in blocking temptation and preventing judgment shift.[141]

In short, Liberman argues that resolutions do not work and that the second-order intention is not an effective source of rational resistance against temptation because that intention is itself vulnerable to temptation's effects.

Liberman goes on to offer her own solution to the efficacy objection. The third premise in her argument relies crucially on the rational requirement to avoid akrasia, and so Liberman thinks that we must identify a mental state that is not subject to an "anti-akrasia norm" to block

---

[138] Alida Liberman, "Reconsidering Resolutions," *Journal of Ethics & Social Philosophy* 10, no. 2 (2016): 4.

[139] The first background principle is: "Temptation Claim: Temptation works by altering the appearances in favor of there being a reason to do the tempting thing, from the agent's perspective" (Liberman, "Reconsidering Resolutions," 6). The other background principle is simply a version of means-end transmission: "Rational-Means Reasons Transmission (RMRT): Where E is an intentional action, if it appears to a rational agent A that (1) there is a reason of strength X for an agent A to attain end E, and (2) M is the only rationally permissible way to attain E, then there will appear to A to be a reason of strength X for A to do M" ("Reconsidering Resolutions," 6).

[140] "(1) When Φing is an action about which an agent has formed a resolution, an apparent reason to Φ (stemming from temptation) necessarily leads to an apparent reason to intend to Φ.
(2) An apparent reason to intend to Φ necessarily leads to an apparent reason to reconsider the intention to avoid Φing (call this intention "Intention 1").
(3) An apparent reason to reconsider Intention 1 necessarily leads to an apparent reason to reconsider the intention not to reconsider Intention 1 (call this intention "Intention 2")" ("Reconsidering Resolutions," 9).

[141] *Ibid,* 15.

temptation. She turns to desire to play this role.

> **Second Order Desire Account (SODA)**: Resolving to Φ involves intending to Φ, and *desiring* not to reconsider the intention to Φ.[142]

On this view, then, a resolution is a (first-order) intention coupled with a (second-order) desire. For example, if I resolve to clean my office, I intend to clean *and* I desire to not reconsider my intention. Liberman thinks that there are two notable advantages to her view over Holton's. One, it doesn't fall prey to the Temptation Transmission argument because the argument responds specifically to resolution understood as a two-tier intention[143] and two, it better captures data about resolutions.[144]

The chief problem with SODA is that we can apply the general gist of the Temptation Transmission argument to Liberman's own positive account. Why should we expect our second-order desires to persist through temptation, when our first order desires so frequently fail in the face of temptation? Doesn't temptation paradigmatically come in the form of desire, swamping other relevant desires? So how then is desire supposed to play the role of resistor to (tempting) desire?

In order to see how this works, consider an example Liberman uses in defense of SODA: the resolution to not eat donuts. In this case, we have three pertinent desires:

---

[142] "Reconsidering Resolutions," 18.

[143] Desires are not subject to the same rational norms as intentions, and so it is rationally permissible to have apparent reason to reconsider a (first-order) intention while simultaneously desiring not to reconsider that intention.

[144] Liberman notes that resolution comes in degrees and that some are stronger than others, and "Appealing to desire as a necessary component of a resolution gives us an easy and efficient explanation of how resolutions can vary in strength …We can explain the strength of a resolution as a direct result of the strength of the agent's desire to avoid reconsideration" ("Reconsidering Resolutions," 20). She further claims that the strength of the resolution desire determines the degree to which the agent is successful in resisting temptation. She writes, "In general, the degree to which an agent is resolute in Φing seems to depend not on how strong the temptation to Φ is, but on how much the agent cares about whether she Φs … suppose I resolve not to drink any beer at a party tonight, and I care very much about whether I keep this resolution. In such a case, it seems that even extremely tempting beer … will not be very likely to sway me to break my resolution" (32).

> Desire 1: The desire to *not* eat donuts which corresponds to an intention to not eat donuts (Intention 1). First order desire.[145]
>
> Desire 2: The desire to persist in the intention to not eat donuts. Second order desire.
>
> Desire 3: The desire to eat donuts. Tempting desire.

So imagine that I form an intention to not have a donut on the grounds of Desire 1, and furthermore I desire to persist in this resolution. But then I enter the break room and a colleague has brought in fresh donuts, and I am tempted to eat a donut after all. In the face of this temptation, imagine that Desire 1 drops away. I no longer desire to not eat a donut; in fact, I desire the opposite. So why should we expect Desire 2 to persist after Desire 1 has disappeared? Liberman's answer to this is as follows:

> The second-order desire can persist when the first-order desire does not because the second-order desire is held for additional reasons. I might desire to avoid eating a donut because I do not want to ruin my supper, or because I do not want to get powdered sugar on my shirt, or because I want to heed my doctor's advice to consume less sugar, etc. I desire to remain firm in my intention to avoid eating donuts for another reason: because I care about carrying out my donut-avoidance plan and being an effective agent regarding the baked goods I consume … my desire to carry out my plan is a desire about what kind of agent I want to be; this sort of desire is resistant to temptations that press on the content of the plan itself.[146]

So the agent is likely to lose Desire 1, the first-order desire not to eat donuts, in the face of temptation. Why not think that the tempting desire also puts pressure on the second-order desire, the desire to persist in the intention not to eat donuts? Liberman claims that the second-order

---

[145] It may seem odd that to speak of a first-order desire to not eat donuts, since normally we associate the temptation to eat donuts with desire, but we don't really associate the resolution to give up donuts with desire. However, Liberman herself speaks this way on pages 22-3, and I think this language is natural in that we have a (placeholder) desire to Φ whenever we intend to Φ. However, I do think that Liberman ought to draw a distinction between placeholder and substantive desires, since there are interesting and relevant differences between the two categories. (Recall I draw the distinction between placeholder and substantive desires from Schapiro's "What are Theories of Desire Theories of?")

[146] "Reconsidering Resolutions," 22-3.

desire will not cave because it is "held for additional reasons." In other words, the agent's second-order desire to not reconsider is supported by all of her first-order reasons, *and* her reason to be an effective agent, someone who follows through on her plans and intentions.[147] Our interest in being an "effective agent" thus is an additional reason, separate from our first-order reasons, and it is this additional reason that is supposed to bolster second-order desires in the face of temptation.

3.1.2 However, the heart of Liberman's critique of Holton's views is that temptation applies just as much to our second-order intentions as it does to first-order. Although she defends her own view from this objection by appealing to "being an effective agent" as an independent, additional reason that bolsters the second-order desire, I think this appeal is ultimately unsuccessful.

First, although being an effective agent, the sort of person who follows through on his commitments, is a worthwhile aim, being an effective agent doesn't require us to follow through on every single resolution. Take the example of resolutions regarding difficult athletic pursuits that require a demanding training regimen. One common reason people undertake such pursuits is to demonstrate to themselves what they're capable of, and in this sense, being an effective agent is among their motives. However, too much rigidity in following one's training plan is a detriment. Obsessive adherence to one's training plan is likely to lead to injury or burnout. Rather, what is needed – and is arguably harder to achieve – is flexible consistency. The point

[147] Although Liberman doesn't highlight this aspect of being an effective agent, it seems to me to share similarities with accounts that emphasis the temporally extended nature of practical rationality. For instance, Thomas Nagel writes that one sees "oneself as a temporally extended being for whom the future is no less real than the present" (*The Possibility of Altruism* (Oxford: Oxford University Press, 1971), 69.). As Michael Bratman points out in "Toxin, Temptation, and the Stability of Intention," simply recognizing that one is temporally extended demands some concern for one's future (85-86).

holds in general as well: in order to be a generally effective agent, yes, you must be willing to stick with your resolutions. But perfect adherence to one's resolutions is not effective agency. It is obsessiveness that is likely to backfire.

Furthermore, if our response to temptation is to count up the reasons we have in favor of not reconsidering vs. the reasons we have in favor of reconsidering, it's not clear that the plentiful reasons on the side of not reconsidering will be decisive in favor of staying resolute. Given that temptation paradigmatically involves the presentation of reasons, it may not matter that we have independent, additional reason for the second-order desire (or the second-order intention, on Holton's account). We cannot weigh our reasons in a neutral deliberative space, because temptation makes certain reasons more salient and thus affects our ability to weigh reasons objectively. In the moment of temptation, our interest in being an effective agent may not count for much.

So whether we understand them as two-tier intentions or an intention-desire combo, resolutions are supposed to entrench or freeze our reasons for our original intention by forestalling reconsideration of that intention. The problem with this strategy is that temptation itself involves the presentation of reasons, and those reasons affect our first and second-order intentions and first and second-order desires. When tempted, we'll feel that we lack good reason to act as we initially intended to act, and this is a key part of why temptation corrupts our rational agency so easily: it involves the appearance of reasons. Part of being rational is responding to a landscape of changing reasons and updating one's intentions accordingly. This means that from the first-person perspective, resolutions aren't going to be consistently effective, since when tempted, we're faced with reasons to do as we're tempted *and* reason to give up the second order

desire or intention to not reconsider.

3.2 HOW TEMPTATION AFFECTS US

In short, one reason both Holton and Liberman's accounts go awry is because they fail to attend carefully to temptation in its own right. Thus, in order to give an account of how to rationally resist temptation, we must have a clear and internally consistent account of the "enemy" in hand. Although both Holton and Liberman do briefly define temptation, gesturing in their accounts towards Scanlon's view of desire as a state involving the appearance of reasons, they also take on the additional and conflicting idea of desire being an "urge" or "pull."[148] A charitable interpretation of this move is that they recognize the limitations of Scanlon's account when it comes to motivational pressure, but I think appealing to "urge" language is not the right corrective to this gap in Scanlon's account. I will start by presenting Holton's view of temptation, and then move to defending my own account of temptation and how tempting desires affect our deliberation.

3.2.1 Holton's view begins with temptation as it relates to judgment shift, since the two are inseparable on his account. He explains his view in several different passages, writing,

> I argue that temptation frequently works not simply by overcoming one's better judgment, but by corrupting one's judgment. It involves what I call judgment shift…This in turn gives rise to the problem of understanding how one can resist [temptation]: the impetus to resist cannot come from the judgment that resistance is best.[149]

> The change in valuation [judgment shift] is not the *origin* of the process that leads to the

[148] Liberman for her part adds the idea that desire comes in degrees of "strength" in discussion of SODA ("Reconsidering Resolutions," 16, 20).
[149] *Willing, Wanting, Waiting*, 97.

> subjects yielding to temptation: it is rather itself caused by [their] awareness that they are likely to yield … If the change in valuation is not the source of the process that leads to yielding, what is? What causes the subjects to yield is desire, in one sense of that rather broad term.[150]

Then, discussing a particular case of temptation studied in experiments done by psychologists Karinol and Miller, Holton applies his view as follows:

> So, to sum up, what I think is happening in [the experiments] is this: the tempted children *find their attention focused* on the immediately available sweet; as a result they *find themselves with a strong urge* to ring the bell to get it; and, as they become aware that *they are likely to succumb to this urge, they change the evaluation of their options* so as to avoid cognitive dissonance [emphasis added].[151]

Scanlon's influence is clearly present in this description: "the children find their attention focused on." But Holton departs substantially from Scanlon as well in his characterization of tempting desire as a pull or urge, and in the claim that we change our judgment in response to the recognition that we're going to succumb to the urge.[152]

In short, I think there are two problems with this broad view of desire. First, Holton's understanding of desire is at odds with the rest of what he says about temptation and resolution. Consider again his description of how temptation works in the study:

1.) The children find their attention focused.
2.) As a result of this focused attention, they experience a strong urge.
3.) They realize that they are going to succumb to this urge.
4.) They shift their judgment in favor of temptation in order to avoid the cognitive dissonance of not doing that which they judge to be best.

How is resolution supposed to be effective, on this account? As I understand Holton, resolutions work by holding off reconsideration, since the second-order intention involved in reconsideration

---

[150] *Willing, Wanting, Waiting,* 101-2.
[151] *Ibid,* 102.
[152] At one point Holton directly states, "What is missing in Scanlon's characterization is the idea that desire *pulls* me to a course of action: that I have an *urge*, or, in more extreme cases, a *craving*, something that moves me to do it" (*Willing, Wanting, Waiting,* 102). I will discuss Scanlon's view of desire in greater detail below.

is the intention to not reconsider the first-order intention. The problem with this is that there's no reconsideration in the summary above. It would naturally fall in step 2 or step 3: perhaps we're prompted to reconsider after having our attention focused, or perhaps the urge of temptation just *is* the urge to reconsider, and so after experiencing the urge, we're likely to reconsider. If there's no moment of reconsidering whether or not to act as resolved, there is no moment at which to choose whether to act as resolved or as tempted, and we simply slide into judgment shift and succumbing to temptation, or resist temptation simply because the tempting urge is too weak to be a real threat. This leads me to my second criticism.

Setting aside the consistency of Holton's view, the more significant reason for rejecting an "urge" conception of tempting desires is that this conception renders us passive in the face of our desires. This approach conceptualizes desire or inclination as a force that acts upon us, which removes our agency in the face of desires. In other words, a view of inclination as an urge makes our ability to resist that inclination entirely contingent upon the strength of the inclination.[153]

Just like a current is a force outside of me with which I struggle, so is an urge-desire something outside of me, something against which I struggle. But this is not in fact what inclinations are like. As discussed in the previous chapters, inclinations are not forces outside of us that act upon us. They arise from within our agency, and when we struggle with an inclination, we are struggling with ourselves, not something foreign to us. Grappling with our own inclinations isn't the same as wrestling with another person or a strong wind. Complicated

[153] I am shifting here to speak of "inclinations" instead of "desires," because although Holton and Lieberman use the label "desire," I think Schapiro is right to observe that "desire" is often unhelpfully vague. I think from context it is clear that Holton and Lieberman are talking about desires that affect our motivation and make it difficult for us to act otherwise, i.e. inclinations, which is why I will revert back to the label "inclination" in the following section. However, again, unless specifically denoted by the label "placeholder desire," readers can assume that when I speak of "desires" I refer to the broad category of desires that includes inclinations and/or that I am simply speaking of inclinations.

though the relationship may be, our inclinations are part of us.

Consider for instance being moved to act as a result of an inclination. If having an inclination is like being pushed and pulled by an external force, then whatever results from that inclination is not properly understood as our action but is instead mere effect or behavior. About this, Schapiro writes, "If my desire pushes me around like an ocean tide, then it is hard to see how its effects can, in principle, count as my actions, unless action is just a way of being pushed around."[154] As Schapiro points out elsewhere, if inclination is a brute force like a tide or wind, it's not clear how we could ever act *on* an inclination.[155] I can act in light of the tide or the wind, but I cannot act on them in the same way I can act on a strong desire to scream or eat cake or start dancing.

Furthermore, construing inclinations as brute forces doesn't just create problems for acting *on* inclinations, it also creates a problem when it comes to resisting tempting desires. Take a current in water: my ability to swim against a current is ultimately not up to me. It's up to me whether I've learned how to swim, or whether I try to resist. But there are some currents so strong that even very skilled swimmers cannot resist despite their best efforts. Similarly, conceiving of inclinations as forces that act upon us makes inclinations like currents: some of them will be perfectly manageable, forces we can resist. But others will simply be too strong, and we will be helpless in the face of them. This means that viewing inclinations as urges renders us unfree in the face of inclination, and whether we are able to resist the inclination isn't up to us but is instead contingent upon the force of the inclination.[156]

---

[154] "What are Theories of Desire Theories of?", 4.
[155] *Feeling Like It*, 49.
[156] I intend my remarks in this section to be neutral with respect to views on free will. Although I suppose some determinists might argue that the correct way to conceive of our agency is to view ourselves as predicting what we'll

It is not an accident, in other words, that Holton's view of judgment shift holds that temptation causes judgment shift because we *predict* that we will succumb to temptation. In other words, on his account of judgment shift, I view myself from the outside and realize that the tempting inclination is too strong to overcome, and so I predict that I will succumb and change my judgment about what's best to do to be in keeping with my prediction. But this is an odd and problematic account of how tempting desire affects us. Relating to our inclinations in this way involves abdicating responsibility for ourselves as agents, viewing ourselves from a third-personal perspective instead of occupying our agency from its own perspective, the first-personal.[157] Take the following scenario:

> Vinny resolves to spend his Saturday catching up on a complex project for work. He then learns that several of his friends are planning to drive into the country and visit Vinny's favorite vineyard on Saturday, and that he's invited to join them. Vinny loves to get out of town on the weekends, and he furthermore enjoys the food and wine at this particular vineyard. When he reflects on the tempting desire to skip his work and go instead to the vineyard, he realizes that he's probably going to succumb to the temptation. Given his prediction that he will succumb, Vinny just decides in advance to give up on his resolution and tells his friends he'll join them on Saturday.

The problem with this scenario is that Vinny replaces the first-personal perspective of the deciding mind with the third-personal. In the terms introduced in the previous chapters, we might say that he allows the instinctive mind to settle the deciding mind. But even this is not quite strong enough: it's not just that his instinctive mind has made a *prima facie* decision to go to the vineyard and his deciding mind goes along with it. Rather, Vinny decides what to do *on the basis* of his prediction about what he will do, which is to say that he doesn't properly decide. He

---

do, rather than deciding what to do, I take it that this is a minority position and would generally be regarded as a *reductio* of the view in question.

[157] My way of framing this issue is drawn in part from Berislav Marušić's *Evidence and Agency*, 122-136.

doesn't settle the practical question of "What should I do?" but rather substitutes it with a theoretical question, "What will I do?" He takes the perspective of an observer, not the perspective of an agent responsible for his own action.

Still, one might object that the problem here is not with the conception of inclination as something that acts on us, necessitating third-person, predicting stances towards ourselves, but rather that the problem lies in Vinny's failure to distinguish between predicting and deciding. In other words, the problem is that he confuses the two activities. This objection holds that sometimes inclinations really *are* too strong to be resisted. In such cases, we ought to recognize that we are unlikely to be able to resist the inclination, although we should not treat this prediction as good reason for then deciding that the best thing to do is to act on the tempting desire, as Vinny does. On this line of thought, perhaps Vinny should keep trying to maintain his resolution and just wait and see what happens on Saturday: maybe he'll be successful in working, maybe he won't. Or, perhaps Vinny should take his prediction that he'll succumb as a sign that his initial resolution was poorly formed and unrealistic and revise on the grounds that his initial resolution was ill-thought.

I find both practical recommendations dissatisfying, since both continue to treat Vinny as a bystander to his actions. However, the objection helpfully highlights the fact that when we predict that there's a good chance we'll succumb to a tempting desire, we typically either alter our plan for executing the intention or resolution in question and/or go on to make a decision that will alter the context in which we decide, in order to avoid succumbing to the tempting desire after all. Take Vinny: imagine instead that Vinny decides to give up his Friday night leisure time and gets his work done then, in order to free up his Saturday for a trip to the vineyard. This

would be an instance of altering his plan for executing the resolution. Or, perhaps Vinny instead makes plans with his spouse to go out for dinner at a local restaurant on Saturday, knowing that he couldn't make it back from the vineyard in time for dinner but that he'll nonetheless have plenty of time for working. If he takes this option, he alters the context or ecology in which he will deliberate and act on Saturday afternoon. If he makes plans with his spouse, he may still feel tempted to drive to the vineyard on Saturday. But having made plans, he'll be able to resist his temptation because this new situation or context will make the vineyard less tempting.

In short, I think we should reject the idea that inclination is a force that acts upon us. For one, construing inclination as a force makes inclination out to be something external to our agency, and inclinations are part of our agency. Second, I don't think that our ability to resist a given inclination is entirely contingent on the strength of the inclination being sufficiently weak.

3.2.2 Recall that in the previous chapter I argued that inclinations are provisional decisions of our instinctive mind. Crucially, I want to maintain that inclinations are moves in deliberation inasmuch as they involve the appearance of reasons, but that they put pressure on the deciding mind because they purport to settle our action. I think this represents the amendment Holton was after with respect to Scanlon's account insofar as it goes further towards explaining the motivational pressure inclinations exercise on our will, but it has the advantage of avoiding the problems associated with describing tempting desires as urges.

It's clear that it's difficult to resist temptation. Arguably an attitude doesn't count as tempting if it's easy to resist. But why is it difficult to resist temptation? Not because temptation is a force that acts on us from without, but because when we are tempted our own agency is in tension, part of it (the instinctive mind) directing us to act as it wills and the other part (the

deciding mind) asking "But should I really φ?" or even more open-endedly, "What should I do?" So tempting desire is not arational. Part of the difficulty of resisting tempting desire is the difficulty of resisting the reasons latent within it. Furthermore, strong tempting desires can make certain reasons very salient, so that it seems as if we have very strong or decisive reason to do something, even though that may not in fact be true.

When we are specifically tempted to give up on a resolution, we may lose our grasp on the reasons we had for forming that intention, but that does not mean that we lose those reasons altogether. Rather, it's as if we "forget" or "lose sight."[158] For instance, if I resolve to exercise more but then am tempted to stay on the couch when it comes time to go to the gym, it's not that in being tempted my reasons for exercising are no longer relevant to my situation. No, the reasons that led me to resolve to exercise more in the first place are still relevant considerations; I've just lost sight of them because the temptation made other reasons salient.[159] When tempted, it's not just that my instinctive mind is figuratively yelling imperatives at my deciding mind and the imperatives have no sticking power. Rather, inclinations appear as imperatives for which we have good reason. In the gym case, my inclination to stay on the couch will include my being drawn to the comfy couch, the annoying long drive to the gym, and the physically strenuous and unpleasant workout waiting for me upon arrival at the gym.

3.2.3 But this leads us back to the place at which the discussion began: if temptation works by

[158] Thus, succumbing to temptation often seems subjectively rational in the moment but later occasions regret.

[159] Perhaps surprisingly, this account of how temptation affects us is compatible with how Holton understands the effort involved in resisting temptation. He writes, "One maintains one's resolution by dint of effort in the face of the contrary desire" (*Willing, Wanting, Waiting,* 118), and then later adds "[…] the effort involved has to be a kind of mental effort. It is the mental effort of maintaining one's resolutions; that is, of refusing to revise them. And my suggestion here is that one achieves this primarily by refusing to reconsider one's resolutions" (121). In other words, according to Holton's own lights, it is difficult to resist temptation not because resisting temptation is like swimming against a current, but because we must set aside the reasons temptation makes salient and instead affirm our resolutions. In other words, it seems that Holton's account of tempting desire is confused.

the instinctive mind directing us to act a certain way and thus making certain reasons very salient, then the saliency of those reasons affects our ability to maintain our resolutions in the face of temptation. Forming a resolution is not an automatic out from being affected by temptation. Tempting desires are throwing reasons for consideration into the ring, and they may appear to be excellent reasons even in circumstances where we've formed a resolution to the contrary. After all, as I said above, the deciding mind doesn't tally our reasons from a neutral or dispassionate perspective. We can try to ignore our inclinations when deciding, but tempting reasons will affect our overall tally of what we have best reason to do unless we specifically intervene to wholly discount tempting reasons in our deliberation.[160]

Imagine for instance that Bryn has resolved to stay home and eat simple homemade meals over the weekend in order to save money. But then on Friday afternoon she gets a text from a friend inviting her to join a group of people at her favorite restaurant for dinner. It would be natural for Bryn to feel tempted to join them: she would have the company of friends, her favorite food, and no work preparing for or cleaning up after dinner. Although it might be rational for Bryn to refuse to reconsider her dinner plans given her initial reasons for resolving to stay in and any independent, additional reason she has for refusing to reconsider, it's not clear why these reasons will be compelling in the face of temptation. After all, when her deciding mind is weighing what to do, all of the reasons to act as she's tempted will be part of the deliberative milieu. Furthermore, she can't just pull out the effective agency card and say to

[160] Furthermore, it's not totally clear that this is possible, although it seems to be possible to at least some degree. More importantly, it seems to be a violation of rational norms to wholesale discount some class of reasons. It's not as if tempting reasons are a totally different type of reason, a type that doesn't actually bear on our acting. Tempting reasons are still reasons, and so we will have to say something about why it's permissible to discount them if this is the route we take. I will briefly revisit the rationality of this below, at the end of the chapter.

herself, “ah, compelling though these are, I must rule out and ignore all of the tempting reasons because I want to be an effective agent.” You can be an effective agent without following through on every single one of your resolutions. So why shouldn’t the deciding mind choose to downgrade effective agency on this occasion and go for the pleasure of going out with friends instead?

Furthermore, it seems to me that this difficulty applies just as much to Liberman’s SODA account as it does to Holton’s two-tier resolutions. Although it’s possible to want conflicting things at the same time, it would be unusual to experience a tempting desire *and* a desire to not reconsider what you’ve previously decided to do because you want to be an effective agent. Why is this unusual? For the same reasons listed immediately above. Temptation makes us think that we no longer have good reason to do as we previously intended, and in the absence of these reasons, it will be hard to maintain a desire to not reconsider that prior intention. Furthermore, although I do think it’s possible for the instinctive mind to be conflicted and thus for us to have inclinations for two conflicting things at once,[161] I think this is a case in which it’s important to distinguish between desire in the placeholder sense and desire in the substantive sense, i.e. inclination. I suppose it is technically possible to be inclined to be an effective agent, but this is rather odd as an object of inclination. Consider how odd it is for instance to speak of “it being easier to be an effective agent rather than not.”[162]

Granted, as Liberman argues, there is independent reason for the desire to not reconsider,

[161] It seems clear that non-human animals can have conflicting inclinations: consider for instance a dog who is waiting in sit for a release command prior to eating her dinner. The dog clearly wants food and also clearly wants to please her humans. The same seems possible for human animals: we can want different and incompatible careers, want to spend our evening in different and incompatible ways, etc.

[162] Arguably the problem with this as an object of inclination is that it is too meta. “Being an effective agent” is not an action you can undertake. It’s the accomplishment of one’s actions, and the way to accomplish one’s actions is just to perform a particular action. You cannot generically perform the action of accomplishing one’s actions.

and that reason is the aim of being an effective agent. In other words, Liberman claims that Bryn's "first-order" reasons for desiring to not reconsider her first-order intentions may disappear in the face of temptation, but the desire to be an effective agent will persist and protect the first-order intention from reconsideration. This, however, seems tenuous. I don't see why the temptation's capacity to affect what we see as good reason extends only to first-order reasons and not also to second-order desires, including desires motivated by "additional" reasons like being an effective agent. Of course, it's possible for this reason to persist in the face of temptation, just as it's possible for first-order reasons for a resolution to persist in the face of temptation. But I think we cannot appeal to this reason as a special consideration that is somehow immune from pressure by temptation.

And, as discussed above, being an effective agent doesn't require *never* changing your mind or giving into temptation. Sometimes we make foolish resolutions and sometimes the situation changes so significantly as to make our original resolution inapt. It's also perfectly acceptable for a normally resolute and effective agent to occasionally give up on a resolution just because.[163] In other words, our values and goals regarding the kind of agency we wish to have extend beyond mere efficacy. This is a point Sarah Paul makes in "Diachronic Incontinence is a Problem in Moral Philosophy," writing

> Most of us do not care about *perfect* self-governance, even as an ideal. We also care about things like existential spontaneity, losing control, rolling the dice and letting the world decide, and other more Romantic ideals. For an agent with these multifaceted values, a life that is perfectly self-governed would not in fact be successful relative to her varied concerns.[164]

[163] The problem, of course, is that we are always tempted in each instance to conclude that this is one of those situations in which it's permissible to give up the resolution.
[164] 345.

Given this, we should not rely too heavily on the idea of effective agency as a solution for resisting temptation. Sure, sometimes we resist tempting desires in order to persist in our goals and be effective as agents, but in other cases we prefer the ideal of being a flexible, spontaneous, or even rebellious agent.

Instead, I think that the power of Liberman's positive account derives from her emphasis on desire. When we want to do what we've resolved to do, it's much harder for temptation to get a foothold in our consciousness.[165] When a desire to do as we've resolved is making our reasons to act on the resolution very salient, tempting reasons will have a harder time crowding them out. For instance, compare someone who enjoys running resolving to push themselves to run their first marathon in contrast to a self-identified "couch potato" who resolves to run a marathon only on a dare from a friend. Although it's certainly possible for the second person to successfully finish a marathon, it seems more likely that the first person will complete the race, and furthermore, it's likely that the first person will have an easier time with their training. This person identifies as someone who enjoys physical exertion and they view their training as something they want to do. The second person will relate to running the marathon as something they "have" to do. It's not an accident, in other words, that many people come to like and/or teach themselves to like something in order to fulfill a resolution to do that thing.[166]

---

[165] This idea is well-supported by empirical research on temptation that shows that having "want-to goals," goals that reflect our "genuine interest and values and are personally important and meaningful" helps us to focus on our goals and not get distracted by distracting and tempting alternatives (Marina Milyavskaya et al., "Saying 'No' to Temptation," *Journal of Personality and Social Psychology* 109, no. 4 (2015), 679): See also Edward L. Deci and Richard M. Ryan, "Facilitating Optimal Motivation and Psychological Well-Being Across Life's Domains," *Canadian Psychology* 49, no. 1 (2008): 14-23 and Kaitlyn M. Werner and Marina Milyavsksaya, "Motivation and self-regulation: The role of want-to motivation in the processes underlying self-regulation and self-control," *Social and Personality Psychology Compass* 13 (2019): 1-14.

[166] I am not a good runner, but I am a very committed runner, and in conversations with non-runners I often find that non-runners assume that I find running miserable and just have more will-power to push through miserable experiences than they do. I constantly have to reassure them that the key to being a runner is taking it easy at first

However, although I think desire is a powerful tool for achieving our goals and resolutions, it does not thereby follow that making non-reconsideration the object of our desire is the most effective and rational way to achieve our goals. Rather, note that in the example above, the object of desire was the goal itself: wanting to run the marathon. This is importantly different from desiring to maintain a resolution or desiring to not-reconsider. Thus, if desire will help us resist temptation, we will find the most support from desires to do what we intend to do, i.e. first-order desires, and not in the second-order desire to resist temptation to which Liberman appeals.

In short, it seems that we finally have the answer to the question I posed at the outset of the paper: resolutions understood as a special two-tier intention are not a reliably effective source of resistance to temptation. Neither Holton or Lieberman is committed to the claim that resolutions are effective in every instance, but one way to describe their project is the attempt to give an account of why we should expect resolutions to be effective at resisting temptation in a rational agent. But given the nature of temptation, I think we must abandon this aim. Temptation is not a force outside of rationality such that if we put up appropriate rational bulwarks, we will be free from temptation's pressure. Temptation occurs *within* our rational nature. Temptation is an inner conflict, not a conflict in which one party to the conflict assails us from without.

However, it is important to clarify at this juncture that resolutions nonetheless play an important conceptual role when it comes to temptation. We should not discard the idea of resolutions altogether, because resolutions mark the difference between intentions that are open

---

and learning to find running at least moderately enjoyable overall, even if not enjoyable at every single instance. I too would give up running if I found it miserable, but a key part of becoming a runner is becoming physically adapted to running such that it's not literally painful to run for more than a short distance and furthermore learning to enjoy other things about running, i.e. having the chance to clear my mind, getting out of the house and onto trails, the satisfaction of challenging myself, etc.

to easy revision and those for which there ought to be a high bar for revision. Furthermore, this conceptual difference does make a difference in what's most rational for us to do. If on vacation I intend to spend my afternoon watching TV but then change my mind and go for a mystery novel instead, there's no sense in which this is a poor decision or a moment of weakness of will. I just changed my mind about what I wanted to do. If on the other hand I give up on my resolution to spend my afternoon working in favor of reading a mystery novel, I have probably failed to act as I ought.

This might seem inconsistent with my claim that resolutions are not of special use in resisting temptation: how could is be that resolutions mark what is (normally, in the absence of other special reasons or notable changes in circumstance) most rational for us to do but are not of any special use in resisting temptation? Are we really that insensitive to what we have most or best reason to do? In some cases, yes. This just is the problem of weakness of will or *akrasia* or resisting temptation. If we always did what is most rational for us to do, there would be no need for this chapter.

However, having said that, I want to immediately walk the claim back to some extent. I am not saying that resolutions are of no use; that our understanding of what our reasons are is useless in the face of the force of tempting desire. I am just denying the strong claim that simply forming a resolution understood as either a two-tier intention or two-tier intention and desire pair is sufficient for resisting temptation. I will take for granted that very often the presence of a resolution means that we *ought* to resist temptation. But simply having a second-order intention or desire present is not sufficient as a strategy for rationally resisting temptation.

Insofar as resolutions help make us aware of the excellent reason we have to φ, yes, I do

think they probably help rational agents resist temptation. But this is not a function that is unique to resolutions. This is just a point about what it is to be a rational agent who is responsive to reasons that bear on practical questions. And certainly this claim does not show that forming a resolution is some special strategy that will take away the power of tempting desire altogether.

I think the point is even clearer in the context of the instinctive-deciding mind dichotomy. The instinctive mind is rational inasmuch as it guides our action according to our teleological understanding of the world, an understanding shaped by what pleasure picks out as good for us, but the instinctive mind is prone to error about what is all-things-considered best for human agents to do. So it may be rational, all things considered, for us to resist temptation and not pursue the action that yields an immediate reward, but our inclinations are not under the direct control of the deciding mind. We cannot change our inclinations at will, and so the instinctive mind doesn't just give up its conclusions when the deciding mind insists, "No, in fact there is not decisive reason to do that thing." Or, to use the language of the prior chapters, the instinctive mind may continue to insist "delicious-doughnut-to-be-eaten!" even when the deciding mind has concluded that the doughnut is fried and therefore to be avoided for the sake of lowering one's cholesterol. Neither does the instinctive mind cease to see the world teleologically and cease forming inclinations just because the deciding mind formed a second order intention or desire to not change its mind about a first order decision.

Where does this leave us, then? In order to give an account of how to rationally resist temptation, we need to broaden our perspective and move our focus beyond resolution and the discrete desire to be an effective agent to a practical virtue that encompasses our inclinations themselves.

3.3 MANAGING TEMPTING DESIRES

Our instinctive mind sees the world teleologically and instructs us to act in light of this understanding. In this sense our inclinations *are* rational. They're not just brute urges pressuring our deciding mind from without, but are a creaturely way of acting towards what's good. On the other hand, our inclinations are not rational if by "rational" we mean immediately responsive to what we judge our reasons to be. In light of this latter point, perhaps we should give up on the idea of resisting temptation in a uniquely rational way and focus on distancing ourselves from our inclinations in order to manage them.

In the penultimate chapter of his book, Holton introduces the idea that rationally resisting temptation requires a general policy of non-reconsideration, a policy of not reopening the deliberative question when temptation threatens a resolution.[167] What exactly Holton means by this is sometimes difficult to trace, but one natural thought is that reconsideration is a strategy of management: since we cannot get our inclinations to respond to our judgment about what we have best reason to do, and since furthermore our tempting desires can affect our judgment about what we have best reason to do, perhaps we should focus on simply ignoring tempting desires altogether.

However, Holton denies that non-reconsideration is to straightforwardly ignore one's tempting desires. He writes,

> In saying that agents do not reconsider, I do not mean that they do not think about the issue at all; as we have seen, some thought will typically be necessary for effective monitoring. Non-reconsideration only requires that they do not seriously reopen the issue of what to do, and seriously arrive at a new judgment … That judgment [that it would be best, all things considered, to abandon the resolution] involves not just an evaluative

[167] *Willing, Wanting, Waiting*, 140.

> judgment, but a comparison: a *ranking* of one option as better than the others … [Such a ranking] is not the kind of thing that simply arrives unbidden.[168]

In other words, for Holton, non-reconsideration is supposed to protect the rationality of resisting temptation: it's not a concept introduced to solve the problem of efficacy but is rather introduced in order to preserve the rationality of acting on one's resolutions in the face of temptation. If we were to reconsider in the face of temptation, Holton reasons, then we would end up making the judgment that it would be best for us to act on our tempting desire and thus we would have a sort of reverse *akrasia* problem in which it would be irrational for us to act as we initially resolved.

More practically speaking, in terms of what it actually looks like to adopt a policy of non-reconsideration, in the above quote Holton suggests that reconsideration does involve thinking about one's resolution, just not "seriously reopening the issue of what to do." There is something compelling about this reply, since there is an important difference between merely thinking about an alternative course of action as opposed to re-opening a deliberative question and actively ranking one's options. However, this reply also requires a delicate balancing act, and it's not clear that this balance is possible in practice. About this, Paul writes,

> [Holton] denies that what he is recommending is weathering temptation by making oneself irrational, or even arational; we are meant to be able to see ourselves as in rational control of our actions when implementing a prior resolution. At the same time, his proposal requires ignoring one's own evaluative ranking at the time of action and refusing to reconsider a resolution one knows it would be rational to revise if one did [reconsider in view of the tempting reasons]. Holton therefore needs a cognitive state to exist in which the agent takes her present action to be up to her, maintains awareness of her resolution and the considerations supporting it, undergoes a shift in evaluative judgment in the light of which those considerations appear comparatively weak, and yet sees no open practical question. This strikes me as a very difficult state of mind to consciously maintain, bordering on bad faith.[169]

---

[168] *Willing, Wanting, Waiting*, 150.
[169] Sarah Paul, "Review of *Willing, Wanting, Waiting*." *Mind* 120, no. 479 (2011): 890-91.

Furthermore, I think we can add to these concerns the idea that the more natural interpretation of how non-reconsideration would work in practice would be to say that non-reconsideration just is the policy of refusing to take tempting reasons into consideration, steadfastly ignoring them, refusing to engage with or think about them. This is after all a strategy people take with respect to temptation: say that I've resolved to forgo all fun purchases for the remainder of the month. Essential purchases only. I might reasonably refuse to engage with any tempting thoughts in the course of carrying out this resolution, delete or block all e-mails about sales, immediately dismiss proposals from friends to go out for the night.

However, in keeping with Paul's remarks, this is unhelpful as a singular long-term strategy, since it results in a kind of alienation from one's inclinations that I think cannot be sustained for long. Furthermore, this strategy is open to the objection that in so acting the agent is irrational, closing herself off to a set of perfectly good reasons to revise her earlier intention. And these points are ultimately the problem with adopting distance-and-manage-by-ignoring as our overall strategy with respect to tempting desires. Refusing to consider tempting reasons may be a way of managing one's tempting inclinations, but to adopt this position is to treat one's inclinations as something that happens to you that you must work around. It may be necessary or appropriate to do this in the short run or in certain extenuating circumstances, but it is not the overall outlook we should adopt towards tempting inclinations because it is not sustainable or appropriate to perpetually live in a state of divided agency. In other words, rather than skirt or bypass our inclining nature in order to forestall inclination from interfering with resolution, the better option is to recognize that inclinations are among the attitudes that constitute our agency, and as such our stance towards them should not be denial but rather cultivation.

3.4 RESOLVE AS PRACTICAL VIRTUE

An effective response to temptation must address temptation itself. One often hears that it is better to address to source or root of a problem, rather than simply try to manage or mitigate its effects, and the adage holds in the case of temptation as well as in home repairs. Rather than respond to temptation by counting up reasons, a process that will be prone to distortion by highly salient tempting reasons, or respond by permanently adopting an alienated stance towards one's tempting desires, we ought instead to strive to resist temptation by attending to what and how we generally desire.

In other words, we need a practical virtue to forestall temptation. If we are to rationally resist temptation and uphold our resolutions, we must focus on mitigating or even preventing temptation altogether, not resisting it. This however requires a practical virtue, since a practical virtue shapes the nature of agency, forming habits of deliberating and desiring.[170] I suggest that we call the practical virtue relevant to resisting temptation "resolve."[171] I will first sketch out

---

[170] By practical virtue, I mean an excellence of agency. On my view it would be possible to lack moral virtue and nonetheless have practical virtue, although I nonetheless think the goodness of resolve is substantive and is in that sense "moral."

[171] One might wonder if resolve is simply another name for continence. I do think these is a strong connection between resolve and continence, but I think resolve involves more than mere continence, since continence involves merely controlling how one acts in the face of one's desires as opposed to shaping them.
I also think there may be a connection between resolve and the concept of "trait self-control" in social psychology, which is self-control that involves automatic behaviors as well as conscious and effortful exercise of control. About it, psychologists Marleen Gillebaart and Denise T. D. de Ridder write "These findings give credit to the idea that people high in trait self-control make the desired choice in an automatized, effortless manner, suggesting that trait self-control does not so much involve effortful resistance of immediate urges on single occasions, but rather involves the ability of not being tempted or distracted by such urges at all" ("Effortless Self-Control: A Novel Perspective on Response Conflict Strategies in Trait Self-Control," *Social and Personality Psychology Compass* 9, no. 2 (2015): 90). For more on the idea of self-control as a trait or disposition, see June P. Tangney, Roy F. Baumeister, and Angie L. Boone, "High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success," *Journal of Personality* 72, no. 2 (2004), 271-322 and Denise T.D. de Ridder et al., "Taking Stock of Self-Control: A Meta-Analysis of How Trait Self-Control Relates to a Wide Range of Behaviors," Personality and Social Psychology Review 16, no 1. (2012): 76-99.

what I imagine the resolute person would look like before transitioning to address the question of how we cultivate resolve.

3.4.1 The resolute person is good at taking the long-view, good at remembering why they resolved to do the thing in the first place and good at anticipating how they'd feel in the future if they abandoned their resolution. Furthermore, such a person will have dispositions to desire that which they've resolved to do and minimize the effects of tempting desire. Of course, there are no absolutes here. Sometimes it is normal or even good to be tempted, and so I'm not claiming that a resolute person will never experience conflict over a decision or desire something opposed to their resolution. Rather, the resolute person is the kind of person whose inclinations are generally in keeping with her judgments about what is worth desiring and doing, and furthermore is good at delaying gratification, not getting easily distracted by desires for immediately available pleasant things, the pursuing of which will prevent her from following through on her other resolutions.

Return to the example of Bryn and consider what this might look like in practice. Above, I claimed that when Bryn is tempted to go out to eat with her friends, her temptation will make her reasons for going out very salient. One response to this temptation would be to tally her reasons for and against going out. The problem with this response, as previously delineated, is that a strong temptation to go out will make all of her tempting reasons very salient, and the initial resolution reasons will pale in comparison, and so her reasons on balance may favor going out to eat, even though this is opposed to her resolution. In contrast, on the view I'm advancing, Bryn's capacity to resist tempting desires will depend on her broader tendencies regarding desire and deliberation. Is she easily distracted by the inclination to go out with friends and does the

tempting desire swamp out all other relevant desires? Or does she remain mindful of her desire to be more financially disciplined, to increase her savings? When deliberating, does she account for the fact that her resolution was formed in order to resist temptations exactly like these, thereby downplaying the apparently good tempting reasons and refusing to actively reconsider her resolution?[172] Or does she take the tempting reasons as new pieces of information that call for full-scale reconsideration of her resolution?

Some of this description might seem relatively obvious, but notice how I'm *not* describing resolve: resolve is not forming a resolution that will form a "wall" around one's future deliberation and prevent temptation from taking hold. Neither is resolve strong will power, the ability to punch down or overcome any temptation that comes one's way. This way of conceiving resolve sees resolve as a virtue exclusive to the deciding mind, a strong capacity to resist the provisional decisions of the instinctive mind. On the contrary, I think resolve is a practical virtue of the whole agent, which means that it encompasses both the instinctive and deciding minds. So yes, resolve involves certain dispositions of the deciding mind relevant to temptation, but it also includes the instinctive mind. Resolve shapes our teleological understanding of the world as much as our deliberation.

Reading these descriptions, it may begin to seem as though the person of resolve just doesn't experience temptation. This is true in some sense. It is not true insofar as a resolute person should be able to feel or recognize the force of conflicting considerations. So Bryn, for

[172] However, it is clearly irrational to *never* reconsider one's resolutions or intentions. So any such policy will have to take this into account. Holton proposes the following as guidelines: "It is rational to have a tendency not to reconsider a resolution: if one is faced with the very temptations that the resolution was designed to overcome; if one's judgment will be worse than it was when the resolution was formed. It is rational to have a tendency to reconsider a resolution: if the reasons for forming the resolution no longer obtain; if circumstances turn out to be importantly different from those anticipated; if one made an important mistake in the reasoning that led to the resolution" (*Willing, Wanting, Waiting*, 160).

instance, might remain fully resolute, fully committed to her budget, and yet acknowledge the presence of conflicting reasons and even "feel" their force, by which I mean seeing them as compelling reasons on which she could act, as opposed to considerations that she merely recognizes as potentially reason-giving for another person although they have no draw on her, at least not in the current context. Furthermore, I think having the virtue of resolve is compatible with experiencing conflict (understood broadly) over a decision or commitment. However, it is true that the resolute person doesn't experience temptation in that the resolute will take pleasure in and thus be inclined to do that which she intends to do and has resolved to do.

Notice that idea of resolve as a practical virtue encompasses key ideas Liberman appealed to in her account. The best understanding of Liberman's SODA highlights the centrality of desire *for* that which we've resolved to do, and the resolute person will either naturally desire the object of his resolution or actively work to cultivate desire for it. What about the desire to be an effective agent, which played a prominent role in avoiding reconsideration for Liberman? Will this desire be present in the resolute person? Yes, although I think this desire will not play an especially prominent role in the psychology of the resolute person. If the resolute person was constantly being resolute in order to prove his effectiveness as an agent, this would be a desire to be effective for its own sake, and this hardly seems like an excellence of agency. However, I do think a resolute person would desire to be an effective agent, and this desire might be especially important for those in the process of cultivating resolve.

3.4.2 Having seen an outline of what resolve looks like, arguably the more difficult question is how one becomes a person of resolve. The reason tempting desire poses a problem in acting as we've resolved to act is because it is recalcitrant to our judgments about what we have best

reason to do in a given moment. So how can I then insist that the rational way to relate to temptation is to cultivate one's inclination in the right way?

First of all, perhaps surprisingly, we can use management to cultivate our inclinations. This is surprising because managing or acting on is a stance of alienation: when I take this stance towards my inclinations, I seem to be regarding them from a distance rather than inhabiting them. However, when we manipulate our own attitudes with the result that the attitudes themselves are changed, the end result is a change in the attitudes we inhabit and potentially a change in our own character. Take a specific example: say that I am trying to teach myself to love running, and so I decide to make running more fun by listening to my favorite music for dancing while running. I think this is best described as a case of manipulation or management because it is an attempt to shape one's inclinations, not by directly altering our teleological understanding of the world but by associating something I wish to be inclined to do with something I already take pleasure in and am inclined to do. But, if the result is that I develop the inclination to run, I think this strategy of management is a form of cultivation insofar as it results in me having cultivated a new inclination.

Still, the natural objection is that unless my new inclination involves some sense of the goodness of running – the inchoate or latent reason of inclination – it will not be an inclination proper but rather a brute urge. This then is where the role of attention comes in. If all we did to cultivate inclinations was bribe or trick ourselves into acting, cultivating inclinations would be no deeper than forming new associations. So for instance in the example above, I would simply form an association between something I was already inclined to do and something new, thus not

properly acquiring a new inclination.[173] In cases of especially dull or odious action, this may be the best we can hope for.

So if our ideal end result is instead a grasp of the goodness of running, then I think there are two primary ways in which I might come to find running good. One is of course through the practical cognition of judging something good. The problem is that in the running case, I probably do judge that running is good, and I'm simply trying to bring my inclinations in alignment which what I judge good. The problem in this case is that I do not take pleasure in running and thus do not practically cognize it as good *qua* instinctive mind, since pleasure is a form of practical cognition common to both humans and animals. What we take pleasure in (for its own sake) is not directly up to us, and so there is no way to guarantee or force one to have an inclination, although I suspect that the right kind of appreciative attention can at least encourage us to take pleasure in something. Again, take running: if I am attentive to my running and actively thinking about the good embedded in the activity, I might begin bribing myself to run with the promise of my favorite music but along the way, through attention to running and its goods learn to appreciate running for other reasons, for example the mental clarity, the fun of pushing oneself to run further or faster, the simple joy of movement.[174] Appreciating these new

[173] If the activity in question is one for which only certain motivations count as good motivations, then we can raise the further objection that merely association an existing inclination with this new activity will not generate the right kind of motivation. I have for instance in mind the chess-playing-child case raised by Alasdair MacIntyre in *After Virtue* (Notre Dame: University of Notre Dame Press, 2007, 188). There he imagines bribing a seven-year-old child to play chess with candy, but points out that we can reasonably hope that over time the child will come to appreciate the goods internal to chess, and so desire to play chess for its own sake. This has the further important effect that the child will no longer be willing to cheat, since if the child can cheat successfully in order to get candy, he has every reason to do so as long as candy is his only reason for playing chess. But if and when he begins to play chess out of appreciation of the goods internal to chess, he will no longer be willing to cheat because to do so would be to violate the goods he appreciates in chess. The example is supposed to be a metaphor for the acquisition of virtue.

[174] In the follow chapter I will among other things consider the kind of control we have over inclinations, asking whether we can exercise what Hieronymi calls "evaluative control" over inclinations. There I argue that we do not. However, I want to here raise the relatively speculatively point that we might via attention exercise a modified form of control over inclinations. The idea to focus on attention as a mode of cultivating inclination comes from Iris

reasons does not mean that I will have an inclination to run at every waking moment – one can after all appreciate reasons in a motivationally cold way – but it does make it possible for me to have the inclination to run at all, since it means that running is now an activity my instinctive mind takes pleasure in and regards as (at least sometimes) *to-be-done*.[175]

3.4.3 Although I have for the most part set aside questions about the rationality of resolution in

---

Murdoch. In "The Idea of Perfection," Murdoch writes that "[attention is] the characteristic and proper mark of the active moral agent" (On 'God' and 'Good,'" in *The Sovereignty of Good,* New York: Routledge Classics, 1970, 33). Attention sounds a lot like evaluative control. After all, one way in which we settle a question is by attending to the question. I might exercise evaluative control over my beliefs about the weather by careful attending to the sky outside my window, for example. So in light of this, it might be natural to read Murdoch as rejecting voluntarist understandings of agency and embracing an understanding of agency something like Hieronymi's. However, I think there is reason to carve out room for attention as a mode of agency or control separate from evaluative control. Although my thoughts on the matter are again, highly speculative, I think that we may be able to shape our inclinations through attention. It may be that we often exercise attentional control only after reflecting and coming to the judgment that one of our inclinations is wrong, but when we shape our inclinations through attention we are not reflectively or unreflectively settling a question. Consider a more mundane example: one of the things I have tried to do in the last few years is learn the local and migratory bird species of my area. I did not previously care very much about birds, but in the past few years I have tried to attend to the birds in the backyard at the feeder and the birds in the tree outside my office window. When out on a walk, I direct my attention to the trees and phone lines, looking for birds. And although I am dreadful at identifying bird calls, I have tried to remind myself to listen to the birdsong I hear when outside, to switch from dismissing it as background noise to zeroing in on it as a distinct call, even if I do not know who is calling. The result of all this attention has been a shift in what I automatically see as worthwhile and to-be-done, and furthermore a shift in my inclinations. If my husband yells, "Hey, there's a funny looking bird in the backyard!" I am eager to cross the house and identify it before the bird flies away. When the neighbors talked of cutting down trees containing multiple nests, I grieved. If I'm out on the walk and hear the cranky chatter of a red-bellied woodpecker (one of the few calls I can recognize), my instinctual reaction is delight of the sort I might feel if a friend were suddenly to call out my name. What began as a sort of intellectual enrichment – here is a set of interesting facts that someone who strives to be educated might undertake to learn – has in fact become a project involving my whole person. In part, the transformation has been my coming to take pleasure in identifying the birds, but it seems to me that attention has played a crucial role nonetheless. My inclinations and emotions would not have been so shaped were it not for the directing of my attention. (The inspiration to use attention to birds as an example of the kind of attention Murdoch might have in mind comes from Jenny Odell's book *How to Do Nothing: Resisting the Attention Economy* (New York: Melville House Publishing, 2019). Among other things, the book is an insightful meditation on attention and will, and although not a philosopher, Odell is philosophically astute and interesting. For instance, Odell writes, "If we think about what it means to 'concentrate' or 'pay attention' at an individual level, it implies alignment: different parts of the mind and even the body acting in concert and oriented toward the same thing. To pay attention to one thing is to resist paying attention to other things; it means constantly denying and thwarting provocations outside the sphere of one's attention" (80)).

[175] In many cases we will need to begin acting before our attention can grasp and appreciate the goodness of a particular activity, but I suspect it is also possible for attention in a more purely contemplative sense to shape our inclinations, i.e. for us to reflect on or ponder the worth of some activity and along the way acquire an inclination to do the thing in question. Often this involves inspiration. For example, I am not a good dancer and am relatively rarely inclined to dance, but watching an especially good dancer or dancers might inspire me and thus prompt me to be inclined to dance as well.

this paper, I want to at this point briefly address the rationality of resolve. Is it really rational for an agent to have habits that support resisting temptation and maintaining resolutions? One intuitive answer would claim that objectively speaking, it's normally rational to act as we've resolved, perhaps because resolution is the product of the cool and deliberative deciding mind in contrast to hasty and often misguided inclination. However, although I think there are cases in which there is decisive objective reason for an agent to act as she's initially resolved, I do not think these reasons undergird the rationality of resolve in general. After all, in other cases agents may have good reason to act as they are tempted. Rather, the rationality of resolve stems chiefly from the authority of our decisions. That is not to say that we can never rationally revise our decisions. But as Paul writes,

> We may see our decisions as to some degree up to us, but we must also see the act of deciding as a matter of relinquishing our authority to change them whenever we like. For, otherwise, they would not be the kind of thing that can do the job of settling an open practical question.[176]

That is, the rationality of resolve does not stem from balancing of reasons to see whether resolution or temptation has better reason on its side but is rather located in our authority to settle practical questions.[177] As cross-temporal agents we need to be able to make plans that settle current and sometimes future practical questions, and it is for this reason that we are often justified in refusing to reconsider our resolutions in the face of temptation.

But this answer simply pushes the question back. We can still ask why it's rational to regard our decisions as authoritative. One answer would hold that it is rational to view our

---

[176] Paul, "Diachronic Incontinence is a Problem in Moral Philosophy," 349-350.

[177] By this, I don't mean that intentions have a reason-giving force of their own, but rather that intentions play a particular role when it comes to how we relate to our reasons. Namely, settle our answer to a question and in so doing they block off reconsideration, or more accurately, once an intention has been formed, reconsideration must be justified and not undertaken on just any whim.

intentions as blocking overeager reconsideration because practical rationality demands it: perhaps practical rationality requires that we take a long-range view about our preferences, or perhaps we will fail to be instrumentally rational in achieving our ends if we are constantly re-opening the deliberative question. However, another answer maintains that granting past decisions special authority is not a strict requirement of practical rationality but is rather rational in the virtue-theoretic sense: it is a wise or good way to act and live in the world. This is the conclusion at which Paul arrives, writing,

> Diachronic continence [or resolve] is in many respects a virtue but not a rational requirement. That is, from the point of view of living well, treating one's intentions as having default stability and refusing to reconsider one's plans too frequently are highly recommended. The appropriate criticisms to make of someone who fails to be stable in this way will be that she is irresolute, flaky, wanton, always wondering where the better party is. But there are failures from the point of view of human excellence or virtue, not the philosophy of action.[178]

This is a substantive claim, and there may well be individuals or cultures who reject this vision of agency, but for my part I think Paul is right. Thus resolve is rational in two important senses. On one hand, it is uniquely rational as a method of resisting temptation. Efforts to become resolute might require non-rational (or arational) methods of resisting temptation, like refusing to buy a dozen donuts at the grocery store for fear that one will overeat them upon arriving home, but resolve consists in a set of dispositions of rational agency and is thus essentially rational in its function. However, resolve is furthermore rational as an excellence of agency. Being a resolute agent enables excellence in desiring, deliberating, acting, and ultimately, living. We should aim to be the kind of agent whose decisions are generally resistant to tempting desires

[178] Paul, "Diachronic Continence is a Problem in Moral Philosophy," 354. For a contrary view on which diachronic continence is a requirement of practical rationality, see Michael Bratman, "Time, Rationality, and Self-Governance," *Philosophical Issues* 22 (2012): 73-88.

because the alternative is to be fickle, to flit from project to project, commitment to commitment, and follow through on few or none of them.

4. Inhabiting Inclinations

"S" was newly married when he and his wife decided to spend their Friday night watching a movie containing depictions of violence.[179] Partway through the movie, S became overwhelmed with violent thoughts. As he describes it: "What if I were to brutally stab someone or shoot someone or harm my wife?" S eventually went to bed and hoped the thoughts would be gone, but they were not. On the contrary, violent images became a recurring feature of his life. Again, S describes:

> "So morning time, I would wake up and maybe the first thought in my mind was an image of stabbing an innocent person. From there, I would take a walk with my dog and, boom, there pops the thought – what if I brutally kill or rape someone and their family when their lights are on at their home?"

Importantly, S never acted on any of these thoughts, but he became terrified that he might. He began avoiding his friends and wouldn't use knives. Eventually, S sought help from a therapist. His first therapist regarded the violent images as expressions of what S actually wished to do, and thus regarded him with fear and alarm, which in turn only heightened his own fear about the violent thoughts. After a handful of visits, she refused to keep working with him. Eventually, S found a new therapist. But this new therapist had a very different opinion of S's violent thoughts: he found them utterly unimportant, "synapses popping off in our head" that we don't need to take seriously. This new therapist put S through exposure therapy and encouraged S to hold a knife to his throat in order to help S realize that he was not going to harm someone, even when explicitly given the opportunity to do so. Furthermore, this therapist helped S realize that his

[179] A true story drawn from "Dark Thoughts", reported by Lulu Miller and Alex Spiegel, *Invisibilia,* podcast audio, January 8, 2015, https://www.npr.org/2015/01/09/375928124/dark-thoughts

distress at the presence of the thoughts was actually a sign of his moral conscientiousness, since someone with less commitment to morality would likely be less distressed at their presence.

This case raises a number of interesting questions in philosophy and psychology, most centrally about who we are, and which actions, ideals, or mental states are robustly attributable to us. For instance, in this case, central to S's learning to live with his violent thoughts was his coming to accept that his violent thoughts were not attributable to him and so he need not be distressed over them. My own interest, however, is not specifically in our relationship to our thoughts understood simply as images or ideas, but rather in our relationship to our desires and inclinations. In particular I want to consider our relationship not to our most cherished desires, but rather our relationship to tempting desires or inclinations that we in some sense repudiate. However, I hope as a result to be able to say something more general about our relationship to our mental states.

Although I am interested in our relationship to our attitudes in general, I focus here on tempting desires because these desires are the most difficult and interesting cases. When it comes to our most cherished desires, it is no doubt easier to see ourselves as responsible for them because they are deeply enmeshed in an agential ecology for which we already regard ourselves responsible: they no doubt cohere with our explicitly endorsed commitments and carefully chosen actions. Thus it might be easy to conclude that we are responsible for such desires by letting actions, commitments, and judgments do the heavy lifting.

When it comes to tempting desires, however, the question of responsibility is made clearer because tempting desires may be at odds with the ideals we hold and the goals we set and the actions we do. The best cases for motivating my question here are cases in which an agent is

explicitly committed to one ideal or course of action but nonetheless is inclined to do something else.[180] For example, the following example serves to illustrate the kind of case I find most interesting:

> Amy is committed to ideals of anti-racism and is a vocal advocate for racial integration in her town. However, Amy's child is about to enter kindergarten, and Amy really wants to send her child to a private school attended almost exclusively by white children instead of the integrated public school. Amy tries to tell herself that she wants to send her child to the private school because the academic standards are better, but she also strongly suspects that she is rationalizing and that racial bias is at play in her desire to send her child to a private school even though she explicitly disavows racist beliefs.[181]

To reiterate, my primary question in this paper is whether Amy's tempting desire is attributable to her. Does it express who she is? Or is it something that simply happens to her? It is normally pretty easy to see how explicit judgments or endorsed commitments are attributable to a person, but in the above case the tempting desire is at odds with the agent's judgments and commitments. Furthermore, neither is it quite right to say that this is simply a case in which someone has conflicting commitments or judgments, which is surely possible, because Amy does not endorse her inclination in the way necessary for it to count as an explicit judgment or endorsed commitment.

Before diving in, I want to head off at the outset a concern that I am defending the view that we should blame and sanction others for their tempting desires. On the contrary, I am not

---

[180] I think it is possible to have a tempting desire that is not an inclination in the sense that one might see as good an action that one reflectively judges to not be good without really feeling like doing the action in question, but if such cases are possible, I think they are the exception and not the norm. One possibility is that such cases are just cases in which there is an inclination present but the inclination is just very weak.

[181] Huck Finn is also an interesting case of recalcitrant tempting desire, and his case presents an interesting twist because instead of being morally concerning, Huck's tempting desire is intuitively what makes him praiseworthy. (The case in question is: Huck Finn has run away from home with his friend Jim. Jim is formerly enslaved and Huck, in keeping with what he was taught to be morally correct, believes that he ought to report Jim to the authorities. However, Huck is tempted to protect his friend Jim and desires to stay quiet out of affection for Jim. Huck is disappointed with himself and regards his desire to protect Jim (and ensuring failure to actually turn Jim in) as a moral failing.) For more on Huck Finn, see Nomy Arpaly and Tim Schroeder's "Praise, Blame and the Whole Self" and Arpaly's "Moral Worth."

advocating for the view that we should hold others accountable and blame or sanction them on account of their temptation.[182] For what it is worth, my view is that although this may be permissible in the occasional exceptional and no-doubt far-fetched case, as a matter of day-to-day practice it's both inappropriate and ill-advised to blame others for their desires and especially tempting desires. Rather, my primary interest here is in whether our tempting desires are attributable to us – whether they are ours, properly attributed to us, perhaps expressions of us.[183] Actions we freely choose are paradigm cases of things that are attributable to us, whereas events that simply happen to us (being rained on) or in us (feeling nauseous) are presumably not attributable to us. As we shall see, there is some controversy in the literature on responsibility regarding how to understand the idea of attributability. On Angela Smith's account, attributability is coextensive with answerability. On other views, however, attributability and answerability break apart. Standardly such views understand answerability as a condition that makes reactive attitudes fitting while attributability is cashed out in terms of aretaic appraisal, such that although we cannot be blamed for states that are merely attributable to us, we can be morally evaluated in light of those states.[184]

---

[182] I will not in this chapter tackle the question of whether (or when) it is appropriate to sanction or blame someone for tempting desires. However, even if we assume from the outset that this is never appropriate, that it not thereby automatic reason to conclude that we're in no way responsible for tempting desires. As Angela Smith points out, "Our judgments about when it would or would not be appropriate to actively express moral criticism or to punish someone for some action or attitude are sensitive to a host of considerations that have little or nothing to do with that person's responsibility and culpability for the action or attitude in question" ("Control, responsibility, and moral assessment," *Philosophical Studies* 138, no. 3 (2008): 379).

[183] Responsibility as attributability is arguably an updated version of what Susan Wolf refers to as "real self" views, which ground responsibility for states in the fact that the states express who the agent truly is. See *Freedom Within Reason,* Oxford: Oxford University Press, 1990.

[184] The original distinction between responsibility as attributability and accountability (it is a separate question how answerability would fall into this mix) comes from Watson's "Two Faces of Responsibility" (in *Agency and Answerability: Selected Essays,* Oxford: Oxford University Press, 2004). Note that Smith has a slightly different view than Watson on this front. More accurately, she thinks that Watson's distinction between aretaic and accountability blame is confused and holds that if aretaic blame is to amount to anything more than superficial evaluations that something is good (or bad) to have, it will have to involve what Watson calls "reactive entitlement"

Part of my motivation for setting aside blame and accountability in this paper is that I approach our relationship to tempting desires from the first-person perspective.[185] My ultimate interest is in how the agent should relate to their own inclinations and especially tempting inclinations, not in how other agents should react to the inclinations of others.[186] For example, in the case of S above, arguably a crucial element in S's learning to live with violent images and thoughts was his coming to externalize them, to view them as meaningless if annoying and sometimes upsetting ideas that happen to him but say nothing about who he really is. Of course, it's possible that this was simply a therapeutic technique. It might be practical in some situations to come to view our thoughts or beliefs as things that simply happen to us as a coping mechanism or as part of an attempt to find greater peace and stability.

("Control, responsibility and moral assessment," 380). This is one of the things that motivates her to regard answerability as the fundamental and unifying concept of responsibility. See also footnote 236 below.

[185] In doing so I am following Schapiro's emphasis on considering inclinations from the standpoint of the participant. See *Feeling Like It*, Section 1.4, 19-26. However, I should clarify that in so doing I do not want to deny that the interpersonal perspective is entirely irrelevant to inclinations. Rather, I think that reactive attitudes, at least those that figure centrally in Strawson's "Freedom and Resentment," are probably too strong in the case of inclinations (it at best unusual to resent or be grateful towards or forgive someone purely on account of their inclination; more likely it is simply not possible to take up such attitudes towards inclinations) (In *Freedom and Resentment and other essays,* London: Methuen & Co. Publishing, 1974). However, neither do I think we should take up what Strawson calls the objective attitude towards our own or others' inclinations – or at the very least it should be an open question whether this attitude is the correct one (9). In a way, this is what makes inclinations so interesting when it comes to responsibility. It seems inappropriate to be reactive towards ourselves or others on account of inclinations. But it also seems deflationary to regard inclinations with the objective attitude. Thus we need to carve out a kind of middle ground in thinking about the sort of responsibility we have for our inclinations.

[186] This is not to deny that this this an interesting and philosophical question. It may for instance be that we have instrumental reason to hold others responsible for their inclinations even if they are not properly speaking responsible or fully responsible for them. By way of analogy, I have in mind here the way skilled teachers of young children "scaffold" in responsibility. So for example, a child in preschool who hits another child is probably not fully responsible, but a good teacher or parent would nonetheless hold a child responsible for this action by talking to the child about how their action harmed others or violated classroom or moral rules. About this, Michael Burroughs writes, "[holding a child responsible] focuses, instead, on educating the child about the classroom or moral value that was violated and, also, to draw out agency in line with this value. That is, the holding responsible, in this and cases like it, both serves to address the violation in question and, also, points the child toward consideration and internalization of the value that was violated (e.g., in this case, "respect the personal welfare of your peers"). …But the practice of holding responsible, via scaffolding, is focused on promoting understanding of what happened, on making connections between the child's action and its consequences, and, also, on resolving the issue at hand in a supportive manner" ("Navigating the Penumbra: Children and Moral Responsibility," *The Southern Journal of Philosophy* 58, no. 1 (2020): 91).

However, whether it's *beneficial* to view our mental states in this way is not the topic of this paper. Instead I am asking this question from the perspective of an agent who wants to know whether such mental states truly are attributable to them or not. So how are we going to determine which states are attributable to us? Although it may be that our moral selfhood is more robust than our agency, one plausible starting point for determining which states can be attributed to an agent is to consider agency. Lest we embrace a kind of intuitionism about the moral self and just point to certain states and attitudes and insist that they are indeed part of the moral self, we need some sort of rationale or procedure for identifying which states are robustly attributable to the moral self. Agency is one natural starting point for identifying such a rationale or procedure: we need simply determine those states that arise from agency or are otherwise related to agency in the correct way, and then we will know which states can be robustly attributed to us in the sense relevant to responsibility.

4.1 ATTRIBUTABILITY AND VOLITION

One natural thought is that states are attributable to us when they are under the control of our will. This view has a variety of expressions and wide-ranging support, but for the time being I will consider an expression of this view present in Schapiro's work.

The question of whether our inclinations are attributable to us does not arise explicitly in Schapiro's work, and I think understandably so. Any proponent of the Incorporation Thesis is unlikely to hold that inclinations are attributable to us in a strong sense given that they think that inclinations must be incorporated and transformed in order for us to act on them. However,

Schapiro does provide more insight than this into her views on whether inclination is attributable to us. Considering the view that inclinations are perceptions are judgments we make hastily or inattentively, she writes:

> this is why it makes sense to regard them as 'prima facie' attitudes. They are judgments we arrive at through a procedure that tends to be unreliable, and because of this, we have a standing reason to review or double-check them. One problem with this way of meeting the constraint, whether as an account of perception or inclination, is that it is in tension with non-voluntariness. Judgments you make hastily or inattentively are no less attributable to you than judgments you make more slowly and carefully. Your judgments, however unreliable, are your doings, in a way that your inclinations and perceptions are not.[187]

Granted, the point of this passage is to argue *against* the idea that inclinations are judgments of a sort. But in teasing out the line of argument, we can also see Schapiro's views on inclination. She reasons:

1. The agent's judgments are attributable to her.
2. So if inclination is a kind of judgment, inclinations are also attributable to the agent.
3. We cannot be inclined at will.[188]
4. If we cannot bring about a state at will, then it is not attributable to us.
5. Thus inclinations are not attributable to us, and so cannot be a kind of judgment.

The key assumption in this line of reasoning is 4. Number 4 is assumed and not explicitly stated in the passage quoted just above, but I think it must be at play in Schapiro's claim that if we say inclinations are attributable to us then we have violated non-voluntariness (which states that we cannot be inclined at will).

It's not difficult to imagine reasons why someone might hold 4. This is especially so when we're thinking about when sanctions are justified, since it is normally unfair to sanction

---

[187] *Feeling Like It,* 80. The objection is raised again later on, on page 102.

[188] "You can determine yourself to φ, and thereby φ, but you cannot determine yourself to be inclined to φ, and thereby be inclined to φ. You can, however, do something we call 'cultivating' your inclinations over time" (*Feeling Like It,* 31).

someone for a state they cannot control at will. However, again, in this chapter I'm focusing not on sanctions, nor on blame, but rather on attributability of the kind needed for aretaic appraisal. So why might someone think that states we cannot bring about at will are not attributable to us? The the motivation for 4 is not, I think, the idea that it would be unfair to morally evaluate someone for a state that is not wholly under her control. I think the motivation for Schapiro's view is the view that autonomy is the defining feature of our agency, and so states that do not stem from our autonomy are not states that can be attributed to us *qua* autonomous agents. In an earlier paper, she writes,

> The concept of full personhood, then, is for Kant the concept of one who has completed the task set for her at the dawn of her reflective consciousness. That "moment" is the beginning of her liberation from the jurisdiction of nature, and that liberation is only completed when she constitutes herself as the authority under whose jurisdiction she falls. Our everyday concept of an adult, I want to claim, just is that of a full person in Kant's sense. To treat someone like an adult is to treat her as the ultimate source of her words and deeds, as the final authority to whom those words and deeds are attributable. As such, the concept of an adult is an essentially normative concept; an adult is the *source* of her beliefs and actions in the sense that she *authorizes* them.[189]

Consider the final sentences: the words and deeds of an adult are attributable to her because she is their ultimate source. What does it mean for her to be the ultimate source? For her to authorize her words and deeds, and presumably in turn for them to stem from the autonomy definitive of her rational agency. Although this is not a full-fledged account of responsibility but rather a description of adulthood, I think this understanding corresponds to something like an endorsement view of our responsibility for attitudes.[190]

---

[189] "Childhood and Personhood," *Arizona Law Review* 45, no. 3 (2003): 588.

[190] Schapiro's theory of inclination is of course strongly influenced by Kant, and so it would not be surprising to find her skeptical about the idea that inclinations are attributable to us in light of Kant's views on the matter. He writes, "since it is there, as intelligence only, that he is his proper self (as a human being he is only the appearance of himself), those laws apply to him immediately and categorically, so that what inclinations and impulses … incite him to cannot infringe upon the laws of his volition as intelligence; indeed, he does not hold himself accountable for the former or ascribe them to his proper self, that is, to his will, though he does ascribe to it the indulgence he would

On the endorsement view, our responsibility is grounded in our endorsement or choice. Thus, for instance, we are responsible for our actions because to act is to choose and thereby endorse the action one performs. In choosing, we make ourselves, our free and autonomous agency, the source of our action. When it comes to attitudes, including emotions and inclinations, the endorsement view would hold that an attitude belongs to me and I can be held responsible for it if and when I endorse the attitude in question.[191] Again, in exercising my autonomy to choose or endorse I make myself the source of the attitude in question. Versions of the endorsement view have been advanced by Harry Frankfurt and Christine Korsgaard, among others.[192] So, if an agent finds himself feeling hateful towards a particular social group, what matters is whether he *endorses* his hateful feelings. If he is distressed by his attitude and explicitly disavows it, he is not responsible for it, but if he endorsed his feeling of hatred, then we could hold him responsible. Importantly, though, we would not be holding him responsible for the attitude in itself, but for his *endorsement* of it, since only the latter is an expression of his agency. Thus on the endorsement view, our responsibility for attitudes and other states we cannot control at will is only ever indirect: we are responsible for them insofar as we harness our agency to said states

---

show them if her allowed them to influence his maxims to the detriment of the rational laws of his will" (*Groundwork of the Metaphysics of Morals,* translated by Mary Gregor, Cambridge: Cambridge University Press, 1998, 4:457-458 (p. 62)). For more on Kant's theory of desire and in particular criticisms of Kantian views on the moral significance (or lack thereof) of our desire, see Tal Brewer's"The Character of Temptation," *Pacific Philosophical Quarterly* 83, no. 2 (2002): 103-130.

[191] That is, although on this view we do not start out with responsibility for desires and attitudes, the endorsement view holds that "[a person] can *acquire* moral responsibility for an attitude by choosing to endorse or 'identify with' it" (Angela Smith, "Attitudes, Tracing, and Control." *Journal of Applied Philosophy* 32, no. 2 (2015): 116). Schapiro would I think want to add that we must also incorporate our inclinations into a possible object of choice before it is possible to endorse them, but the basic point that endorsement grounds responsibility nonetheless holds. See for instance Schapiro's "What Makes Weak-Willed Action Weak?" (forthcoming).

[192] See e.g. Harry Frankfurt, "Identification and Wholeheartedness" in *The Importance of What We Care About,* Cambridge: Cambridge University Press, 1998 and Korsgaard's *The Sources of Normativity* and "Personal Identity and the Unity of Agency: A Kantian Response to Parfit" in *Creating the Kingdom of Ends,* Cambridge: Cambridge University Press, 1996).

.

through choice or active endorsement.

4.2 RESPONSIBILITY AS ANSWERABILITY

The primary problem with the endorsement view and other volitional views like it is that in our everyday moral practices, we regularly evaluate others for attitudes, including attitudes that are not under our direct control. Sometimes we criticize others or experience distress and shame in ourselves because of problematic desires and emotions, even when our volition is actively working to remove or alter the desire or attitudes in question. For instance, take a person who was raised in a deeply prejudiced moral community. Furthermore, say that he disavows his racist attitudes and has been working to overcome these attitudes. However, although he has made some progress, he finds himself regularly experiencing racist attitudes towards the groups he was raised to hate and distrust. In spite of all this, although we would be wrong to berate or sanction him for his attitudes, it would be appropriate for us to negatively evaluate his moral character on the basis of these racist attitudes (keeping in mind that this is one dimension of evaluation and we might regard other aspects of his character more positively). Furthermore, I think it would be right for such a person to be distressed about these attitudes, and it's not just that he should be distressed at these attitudes in the way that he might be distressed to find that he has a chronic illness, i.e. distressed to find out that something bad is happening to him.[193] Rather, the content of his distress should be that he himself is not who he might like to be, not as good as he hopes to be. We are not so neatly distinguished from our attitudes. We cannot recede

[193] See footnote 236 below regarding Smart's understanding of "grading evaluations." Distress over racist attitudes in oneself is not a grading evaluation.

and pretend that our attitudes and desires are irrelevant to who we are.

However, it's not just that the endorsement view fails to account for our existing moral practices. After all, we might have inappropriate or unjustified moral practices. Rather, the deeper problem with volitional views is that such views have a narrow and problematically control-centered understanding of agency and responsibility. They overemphasize what we do and endorse, defining our selfhood too exclusively in terms of autonomy.[194]

There has been a tendency in work on agency and responsibility to take bodily movement and bodily control as a paradigm. This is forgivable inasmuch as our ability to manipulate our own limbs and objects around us at will is remarkable, perhaps the strongest expression of freedom and agency we experience. However, a difficulty arises when we assume that this is the only form agency takes, or when we try to extend this model as a paradigm and apply it to cases of intentions, beliefs, desires and attitudes, none of which can be directly controlled at will. Careful attention to states like intentions and beliefs, on the other hand, suggests that it is answerability that is essential to responsibility, and indeed essential to agency itself, not control. The primary articulation of this view comes from Angela Smith, who uses the view to champion for the idea that we have responsibility for attitudes, emotions, and desires.

To be answerable for a state is to be the appropriate target of justificatory questions regarding the state in question. I cannot choose to believe just anything, but I can be asked to give my reasons for believing. In this sense, I am answerable for my beliefs. [195] The same holds

[194] For more on concepts of self that rely less heavily on autonomy and the historical context of our close association between selfhood and autonomy in the modern West, see Charles Taylor's *Sources of the Self: The Making of the Modern Identity*, Cambridge, MA: Harvard University Press, 1989.
[195] See e.g. Hieronymi's "Controlling Attitudes," *Pacific Philosophical Quarterly* 87, no.1 (2006): 45-74 and "Responsibility for believing," *Synthese* 161 (2008): 357-373.

true of intentions: I cannot choose to intend just anything, but I can be asked to give my reasons for my intentions.[196] Thus, on the view now being considered, we are not responsible for states like intentions and beliefs only insofar as we choose them or they can be connected to choice, but rather we are responsible for them because we are answerable for them, because they are responsive to reasons and so we can be asked to justify them.

However, one might object that the subject of this dissertation, inclinations and in particular tempting inclinations, are importantly different from intentions and beliefs. We can give reasons for intentions and beliefs, and we can also reason our way to a belief or an intention. But any adequate theory of inclinations has to take seriously the ways in which inclinations have lives of their own. We cannot reason our way to an inclination.[197] They seem to arrive passively in us and they are recalcitrant, persisting in the face of judgments that we don't have reason to act as we're inclined. They often make it difficult to act as we judge that we ought to act. Furthermore, I've given an account of inclination that introduces duality into our agency, claiming that inclinations are the products of our instinctive mind. So although answerability might appear to be a good route when it comes to responsibility for desires or emotions in general, it seems at first look a stretch to hold that our inclinations are attributable to us because we are answerable for them. All of this creates a difficulty when it comes to explaining our responsibility for desires on the responsibility as answerability view.

Still, at this point I have not done justice to Smith's views, so in the following I will

---

[196] See for example Gregory Kavka's "The Toxin Puzzle," *Analysis* 43, no. 1 (1983): 33-36. This theme in also explored in Hieronymi's "Controlling Attitudes," "The Will as Reason" and "Two Kinds of Agency" in *Mental Action*, edited by L. O'Brien and M. Soteriou, Oxford: Oxford University Press, 2009.

[197] Using states we can reason to as a litmus test for thinking about agency and relationship is a strategy drawn from Kyla Ebels-Duggan's thinking about attitudes. See for instance "Love and Agency," In *Routledge Handbook of Love in Philosophy,* edited by Adrienne Martin. Abingdon: Routledge, 2018.

summarize Smith's understanding of answerability before returning to the question of whether this account applies to inclination as well. Smith is a champion for the idea that we are responsible in the answerability sense even for apparently passive states, so her work is a good place to start with respect to inclinations. Note that since Smith speaks generically of attitudes and desires, in the following I will revert to speaking of desires in general and cease discussing inclinations in particular.

4.2.1 Clearly when it comes to our desires, we lack control over them of the sort we can exercise over for example our actions and middle-size objects around us. But one theme in Smith's work on responsibility is the idea that we need a more expansive understanding of agency and control. Thus Smith argues that we are responsible for desires because we have what she calls "rational control" over our desires.

> What makes an attitude 'ours' in the sense relevant to questions of responsibility and moral assessment is not that we have voluntarily chosen it or that we have voluntary control over it, but that it reflects our own evaluative judgments or appraisals.[198]

> According to [the evaluative model], it is those mental states which we regard as normatively connected to a person's evaluative judgments that we take to be attributable to her for purposes of moral appraisal.[199]

And then in a later paper

> To say that an agent has 'rational control' over his attitudes is to say two things: first, that he has the capacity to critically evaluate reasons; and second, that his attitudes are (generally) responsive to his assessment of reasons.[200]

Thus, unlike the view that attitudes are attributable to our agency when we couple the attitude in question to our agency by endorsing it, on this view attitudes are attributable to our agency when

---

[198] "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115, no. 2 (2005): 237.
[199] *Ibid,* 257.
[200] "Attitudes, Tracing and Control," 126.

they are generally responsive to our evaluative judgments or assessment of our reasons.

Smith's account is of course Scanlonian in origin. In particular, she clearly draws on Scanlon's idea of judgment-sensitivity. [201] Scanlon writes,

> "[Judgment sensitive attitudes] are attitudes that an ideally rational person would come to have whenever that person judged there to be sufficient reason for them. [202]
>
> Judgment-sensitive attitudes (including desires)] are 'up to us' – that is, they depend on our judgment as to whether appropriate reasons are present. Because of this dependence of judgment, these are things that we can properly be 'held responsible' for in several central senses of that phrase: they can be properly attributed to us, and we can properly be asked to defend them – to justify the judgment they reflect.[203]

The idea of judgment-sensitivity is slightly ambiguous and can be understood in two slightly different ways, and we need to clarify which sense is at play in Smith's account. The first quote above suggests that JSAs are sensitive to the judgment that we have reason for the attitude. Thus, for instance, if I have a desire to go swimming and I judge that I have sufficient reason to have the desire to go swimming, then the desire in question is judgment-sensitive. However, I take it that Smith uses the idea of judgment-sensitivity in a related but distinct sense, namely, that a desire is judgment-sensitive when it is sensitive to a judgment about the goodness of the desire's content. Thus in Smith's sense of judgment-sensitive, my desire to go swimming is (or ought to be) sensitive to my judgment *that swimming is good.* These two senses of judgment-sensitivity standardly go together, since in the normal case my primary reason for having an attitude is the judgment that the content of the attitude is apt, but they can come apart. I might for instance judge that I have sufficient reason to desire to go swimming (in Scanlon's sense of judgment-

---

[201] Although I above quoted her using the label "rational control," a different paper Smith explicitly uses the language of judgment-sensitivity, writing, "Our attitudes, by their very nature, are judgment-sensitive states" ("Attitudes, Tracing, and Control," 124).
[202] *What We Owe to Each Other*, 20.
[203] *Ibid*, 22.

sensitivity) because you have accused me of being dull and never wanting to swim and I want to prove you wrong and show that I am a fun, swimming type of person, and yet not judge that swimming is good, the judgment that would correspond to the content of the desire to go swimming. In the following, when speaking of judgment sensitivity, I will be referring to the idea that desires are sensitive to our evaluative judgments about the content of that desire unless otherwise specified.

The other idea that needs clarifying is Smith's idea of an "evaluative judgment." Attempting to clarify how she understands the term, she writes,

> […] the judgments I am concerned with are not necessarily consciously held propositional beliefs, but rather tendencies to regard certain things as having evaluative significance. These judgments, taken together, make up the basic evaluative framework through which we view the world. They comprise the things we care about or regard as important or significant.[204]

and shortly thereafter adds

> When I say that a mental state or condition 'depends upon and reflects a person's evaluative judgment,' therefore, I am not claiming that the state in question must have arisen from an explicit judgment of any sort; I am claiming, rather, that the state in question can reasonably be taken to reflect an evaluative judgment on the part of the person, a judgment, moreover, which it is appropriate, in principle, to ask her to defend.[205]

I find the first passage quoted somewhat confusing, since taken alone it suggests that when Smith talks about judgment she doesn't really mean "judgments" in the normal sense. Taken in conjunction with the second quote, however, I think that what Smith means to convey is that we need not have an *occurrent* or reflectively formed judgment in order for a desire to count as sensitive to that judgment. Furthermore, a desire need not arise from or have a causal origin in a

[204] "Responsibility for Attitudes," 251.
[205] *Ibid*, 252-3.

judgment in order for it to be under rational control. Beyond this, it seems to me that she must mean "evaluative judgment" to refer to something like our normal understanding of judgment, for how else would a desire be sensitive to an evaluative judgment? That is, if "evaluative judgment" is just a name for the evaluative content of a desire itself, then how could a desire be sensitive to an evaluative judgment? This would require saying that a desire is sensitive to itself, and this wouldn't generate an account of answerability for desires, which is exactly what Smith needs. Accordingly, in the following I will understand the sense of "judgment" at play in the idea of evaluative judgments in the standard sense.[206]

Returning to Smith's account of answerability for desires with these clarifications in hand, let's consider how it's supposed to work in greater detail. The first thing to note is that this account of responsibility clearly assumes a view of desires on which desires are evaluative. That is, for Smith, desires express and are sensitive to our evaluations of goodness. The exact details of this are left undeveloped in Smith's work, perhaps because she focuses on attitudes in general and not desires in particular.

However, at the very least it is clear that on Smith's view, desires are sensitive to our critical evaluation of our reasons.[207] This means that desires are under rational control, and as such they are states whose rationality or fittingness can be brought into question. In other words, they are states we can be asked to justify, and we justify them by citing the reasons we have for

[206] I actually think there's tension in Smith's account of responsibility on this point. On one hand, she speaks of evaluative judgements as actual judgments, separate from their corresponding attitudes. On the other hand, she sometimes slips into speaking of evaluative judgments as judgments implied by desires and attitudes, although we may not in fact have or endorse that judgement. There may have been a shift in her work on this issue over time, since the treatment of evaluative judgments as propositional forms of the content of the attitude itself occurs in her earlier work whereas she switches to the language outlined above in her later work. I will choose for the purpose of the dissertation to understand Smith to be saying the former, not the latter.

[207] "Attitudes, Tracing and Control," 126.

the relevant evaluative judgment. This appeal to justification or answerability offers a procedure for determining which states are connected to agency in the way necessary for responsibility. In other words, justification captures a rational or normative connection between a state and an agent, and when this link holds, the state can be attributed to the agent in the robust sense relevant for attributability.

This is a bit convoluted, so let's briefly walk through an example. Say that Maya has a desire to write a great book on Maimonides. When we ask whether Maya is responsible for this desire, we are (according to my understanding of responsibility) asking whether this desire is robustly attributable to Maya. We could just assert that desires are constitutive of the self, but that would be philosophically unsatisfying. We could take the route of endorsement and say that the self is autonomous and so states are attributable to the self when the autonomous self endorses or chooses the states in question. On Smith's model, the moral self is essentially constituted by its assessment of reasons, and so a given state is attributable to the self when it is sensitive to those reasons. So Maya is responsible for her desire to write a great book on Maimonides because the desire is a response to and is sensitive to her evaluative judgment that Maimonides is an important and worthwhile philosopher and that an excellent book on Maimonides would be a good thing. Were she to lose interest in Maimonides and come to believe that he is a second-rate philosopher, we would expect her desire to disappear.

4.3 RECALCITRANT ATTITUDES

The natural objection to this view is recalcitrant desires. Although I will continue to speak, as Smith does, of desires and not inclinations, this worry about recalcitrance is especially forceful when it comes to inclinations. What about all those desires that aren't sensitive to our evaluative judgment, namely, tempting desires? This issue of tempting desires is a problem to which Smith is sensitive. She writes,

> [Conflicts] between our desires and other attitudes, on the one hand, and our deliberative verdicts about what to do, on the other, are sometimes cited as evidence for the fact that our desires and other attitudes are not reflective of our moral agency.[208]

However, Smith thinks she can deal with the objection. How so? By appealing to the contingency of most conflicts between judgment and desire. Take the example she considers, a case where I desire to stay in bed while coming to the verdict that I ought to get up and go to work.[209] In other words, the conflict in this case is between a judgment about how I should act at a given moment ("I ought to get up right now and go to work") and a desire that inclines me towards a contrary course of action in the same moment ("I want to stay in bed right now"). The desire to stay in bed, however, is *not* in conflict with my evaluative judgment about the goodness of staying in bed. I can judge that it is good or pleasurable to stay in bed and also judge that right now it would better for me to get up.[210]

---

[208] "Conflicting Attitudes, Moral Agency, and Conceptions of the Self," 337.
[209] *Ibid,* 338.
[210] Note that the divergent senses of judgment-sensitivity are at play in this example. In Scanlon's sense of judgment-sensitivity, I would probably judge in this case that I do not have sufficient reason for the desire in question. In Smith's sense, my desire is sensitive to my judgment that this kind of object is worth desiring.

Smith's point goes deeper, though. It's not just that a verdict about the best course of action can coexist with an evaluative judgment without contradiction. Rather, she thinks this case is evidence *for* her view of responsibility, and not just an objection she can deal with. For instance, she writes, "it seems to me that my desire to stay in bed itself depends upon my evaluative judgment that doing so would be pleasant; were I to lose or to abandon that evaluative judgment (e.g., by becoming bored with lounging around), I would also cease to have the desire in question."[211] Thus, Smith thinks that even apparently tempting or conflicting desires are in fact desires sensitive to evaluation, and accordingly, desires for which we are answerable.

I think this is an important response, but I also think it only goes so far. For one, unlike Smith, I worry that there are cases in which the conflicts between desire and judgment are not merely contingent but are rather intractable. I will refer to such cases as "recalcitrant tempting desires." Arguably the case of Amy given at the beginning of the chapter is one such case, but consider the following example as well:

> Nick is a member of a conservative religious tradition that teaches that homosexuality is a sin. However, Nick is attracted to other men and quietly suspects that he is gay although he never talks about it. Nick is miserable about this conflict between what he wants sexually and romantically speaking and his religious beliefs and entertains entering a residential conversion therapy program.

I think it is not a stretch to say about this case that Nick's desires really do clash with his evaluative judgments. The question for our purposes then is whether Nick's desires are attributable to him. Nick himself seems to think they are, even though he simultaneously believes that he thereby needs to repent and be cured of them. On the other hand, many people are likely

[211] "Conflicting Attitudes, Moral Agency, and Conceptions of the Self," 338.

to think that they are attributable to him and for that very reason he ought to give up his conservative religious tradition.[212]

The Scanlonian line on recalcitrant desires would standardly be to either deny the apparent recalcitrance, as Smith appears to do in her discussion of contingent conflicts, or to emphasize that the recalcitrant state is the result of a rational malfunction but is nonetheless a

[212] Curiously, in one of her papers Smith herself makes several remarks that suggest that she might deny that Nick's desires are attributable to him. The remarks arise in the context of Smith's consideration of a person who "makes and accepts egalitarian pronouncements […] [while] his actual behavior reflects subtle attitudes of prejudice" ("Conflicting Attitudes, Moral Agency, and Conceptions of the Self," 341). Smith then imagines an objector responding to this case saying that what the case really shows "is not that these underlying attitudes 'belong' to the person [….] but rather that 'the person' is letting his behavior be directed by forces that are not fully his own" (342). The relevant passage for my purposes comes in her response to this objection. She writes, "This type of analysis seems most plausible when the desires in question fall into the narrow class of appetite desires I mentioned above, e.g. biologically or physiologically based desires of hunger, thirst, and addiction. Someone who unreflectively follows the lead of such desires does seem to be surrendering himself to 'natural' forces in a way that makes it tempting to describe his behavior as 'heteronomous.' Since such desires do not themselves reflect his evaluative judgment, following them without sufficient reflection can be regarded as negligently allowing his actions to be determined by forces 'outside' of his moral self" (342). (Note that Smith elsewhere seems hesitant to include sexual desires in the category of appetitive or biological desires ("Responsibility for Attitudes," footnote on page 248)). In the corpus of Smith's work this passage might seem like a minor aside, but I think it is actually revealing in helping understand what her account would say about recalcitrant tempting desires and inclinations. Of course, because her remarks are so brief, exactly what she means is ambiguous. For instance, if she is thinking of hunger and thirst purely as physical sensations, then yes, I think it's clearly right to say that these sensations are not judgment sensitive. Similarly, if by "natural forces" she means reflexes or the urge to sneeze, then I agree, these states have no bearing on our evaluative judgment and are not attributable to us in a morally interesting sense. But if we are thinking of for example hunger and thirst not simply as brute physical sensations but as mental states, i.e. inclinations to eat and drink, then I think it would be odd for Smith to so quickly conclude desires of this sort aren't sensitive to our evaluative judgment. Normally our inclinations to drink water and keep food are pretty in keeping with our judgments that such activities are good, even if these desires are often prone to the sort of contingent conflicts Smith describes above. In spite of this, in the above passage Smith appears to endorse something like the incorporation thesis. Again, she writes: "Someone who unreflectively follows the lead of such desires does seem to be surrendering himself to 'natural' forces in a way that makes it tempting to describe his behavior as 'heteronomous.'" ("Conflicting Attitudes, Moral Agency, and Conceptions of the Self," 342). This sits uneasily with her Scanlonian understanding of desire: she seems to be saying that on the one hand we have desires, which are expressions of our practical thinking and sensitive to our judgments. On the other hand, we have something else, "natural forces" that are not attributable to us and stand in need of incorporation if we are to autonomously act on them. However, I don't see a principled way to divide up desires into "practical thinking desires, responsive to reasons, states attributable to us" and "natural forces, biological desires, states not attributable to us." Once we start to press and especially if we start to consider our inclinations from an evolutionary perspective, it can quickly begin to seem that all or practically all of our inclinations have some kind of biological basis. In Smith's example where the agent desires to stay in bed, drowsiness and the body's need to rest are surely at play even more than judgments we can call consciously to mind about the goodness of sleeping. Arguably all or most inclinations to procrastinate ultimately aim to soothe anxiety and help the agent feel calm and at ease again, even if this motivation is relatively hidden from conscious awareness when we're deliberating about whether or not to procrastinate. The examples could go on.

state for which we are responsible. After all, Smith points out, if recalcitrant desires weren't under normative pressure to be sensitive to judgment, we would describe them as "arational" and not "irrational."[213] In keeping with this, both Smith and Scanlon are careful to emphasize that what grounds our responsibility for attitudes is not their actual adherence to judgment, but the fact that they are under normative pressure to conform to judgments. Scanlon writes,

> States of the kind I have in mind have [significance for moral appraisal] because they involve taking something to be a reason. Insofar as they are such states, they are the *kind* of thing that is in principle subject to control by the agent's judgment. *But they retain their significance even when this control is, in fact, absent*.[214]

4.3.1 One of the underlying problems here is that Smith's account relies too heavily on judgment and does not give enough attention to desire in its own right.[215] As a reminder of the heart of her view, she writes,

> … the rational relations view, makes rational judgment rather than choice or voluntary control the basic condition of moral responsibility. To say that an agent is morally responsible for some thing, on this view, is to say that the thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it.[216]

So on her account, we are responsible for desires including tempting desires because they reflect what we see, judge, believe – all cognitive verbs. The other verb central to Smith's account is of

---

[213] "Responsibility for Attitudes," 254-5.

[214] Scanlon, "Reasons and Passions*"* in *The Contours of Agency,* edited by Sarah Buss and Lee Overton, Cambridge, MA: MIT Press, 2002, 172. Emphasis added. In *What We Owe To Each Other,* he writes, "Akratic actions (and irrational thoughts) are cases in which a person's rational faculties have malfunctioned, not cases in which these capacities are overmasted by something else, called desire" (40). Smith makes a similar point in "Responsibility for Attitudes," writing, "The 'should' in question here is the should of rationality, and, therefore, marks a normative ideal which our actual attitudes may not always meet" (253).

[215] My assessment that Smith's account is highly judgment-centric is shared by David Shoemaker. In *Responsibility from the Margins*, he writes, "many theorists take judgment to be the key to a type of responsibility labeled as *answerability*. This is because they take responsibility to be about what is 'up to us' as practical agents, and what is most fundamentally up to us, it is thought, is our practical judgment. In addition, they take it that holding others responsible implicates their judgment too, insofar as it assume their ability to *answer* for, to defend, what they have done" (Oxford: Oxford University Press, 2015, 70).

[216] "Control, responsibility and moral assessment," 369.

course "evaluate," but even here, she writes that "To evaluate is to judge the value, worth, or significance of something, which requires the active engagement of our rational faculties."[217]

This is not to say that I want to wholeheartedly reject Smith's view. To the contrary, grounding responsibility in our answerability for our judgments as opposed to our ability to manipulate objects and actions at will is an excellent strategy. Furthermore, our answerability for our judgments *is* relevant to desires insofar as desires are not divorced from these epistemic states but are rather normally well-integrated into an agential ecology that includes judgments and beliefs. However, when it comes to recalcitrant desires, it starts to seem tenuous on Smith's account that we should have responsibility for desires when those desires conflict with the relevant evaluative judgment(s). Why not give priority to judgment, in such cases? Smith and Scanlon both seem to insist that even in recalcitrant cases, what matters is not whether the attitude is in fact sensitive to judgment but simply the normative requirement, the fact that it *ought* to be sensitive. But if this line of response is to be satisfying, we need a story about what's gone wrong. Why is it that recalcitrant states are judgment-sensitive (in the sense of being under normative pressure to conform to judgment) rather just being outside of the purview of judgment altogether?

The difficulty of grounding responsibility for desires in desires themselves, not their corresponding judgments, becomes even more pointed when we pivot to thinking about responsibility for inclinations rather than desire in the generic sense, and not just any inclinations but specifically tempting inclinations. As I understand them, inclinations are provisional decisions of the instinctive mind. Are we responsible for inclinations in this sense? Let's be more

[217] "Conflicting Attitudes, Moral Agency, and Conceptions of the Self," 339.

specific: on Smith's view, to be responsible for inclinations would be to be answerable for them. And what does it mean to answerable for one's inclinations? To be the appropriate target of demands for justification. But what would it mean to justify one's inclinations?

4.4 JUSTIFYING INCLINATIONS?

Intuitively speaking, the idea of justifying one's inclinations sounds implausible, but it's worth unpacking this implausibility. Intentions and beliefs can be conclusions of chains of reasoning, and so it is natural to ask someone to justify an intention or belief: they can simply provide you with the chain of reasoning that resulted in the intention or belief in question. But inclinations are not like this. Inclinations are not the conclusions of deliberation.[218] Granted, if Smith is right that inclinations are sensitive to judgment, then there is a sense in which we can give reasons for an inclination by giving the reasons for the judgment corresponding to the inclination. This still leaves us with a question about recalcitrant attitudes and inclinations, but even setting recalcitrant cases aside, one might begin to wonder whether judgment isn't doing all the work.

Moreover, I think that once we begin to press on the idea of justification, an ambiguity about what it means to justify an attitude or be the appropriate target of justificatory questions arises. On the one hand, justifying might refer to an assessment of the fittingness of one's

[218] This is not entirely correct, since our inclinations can become more specific as a result of deliberation. Say I have the inclination to eat something salty and fried as part of dinner but I'm not sure exactly what I want or exactly where I'll go. In the course of deliberation about what I want, perhaps including researching what options are available to me, I may form a new more specific inclination when I come across a restaurant offering fried cheese curds. Now I am inclined to eat these cheese curds in particular. However, it would not be right to say that this inclination resulted from deliberation, but rather that deliberation made a general inclination more specific.

attitude. Any attitude that manifests some kind of take on the world can be evaluated for justification in this sense.[219] My anger manifests my sense that I have been wronged; my inclination manifests my sense that a given action is to-be-done. Clearly I can evaluate both of these attitudes and conclude that they are justified – yes, I was wronged; yes, that action is to-be-done – or I might evaluate them and find that they are not justified – no, no one has wronged me; no, that action is not-to-be-done. Given that I've used inclination as an example, it should be clear that I think inclination can be evaluated for justification in this sense. This means that inclination does have a normative significance that, for instance, nausea does not. Nausea can be unexplained or surprising and thus "not make sense." But nausea cannot be unjustified. It is not intentional. The problem is that justification in this sense seems too weak for answerability. On Smith's view, answerability grounds not just attributability but a kind of accountability. When you are answerable for an attitude, we can hold you to account for that attitude and ask you to defend it.

This leads to a second sense of justification. We can also take justification to refer to the activity of justifying, or more aptly, settling a practical or theoretical question according to our assessment of our reasons. On this understanding of justification, attitudes that cannot be justified must be modified or abandoned. For instance, Smith writes,

> If a person's judgments, as manifested in her actions and attitudes, appear to violate certain normative standards (whether those be moral, philosophical, prudential, or whatever), it is appropriate (in principle) to ask her to reassess those judgments and to explain, justify, modify, and in some cases apologize for her actions or attitudes in light of this reassessment. Criticism, in this case, is not mere unwelcome description, but calls

[219] My phrasing and thinking here are both influenced by Hieronymi's account of non-voluntary attitudes in "Reasoning First." There, she writes, "The same is also true of a wide range of attitudes – of any attitude that manifests our take on the world, on what is true, important, worthwhile, insulting, wonderful, horrifying, trustworthy, impressive, and so on, for which we can be asked our reasons. Such attitudes must be non-voluntary, in the sense just explained, in order to play the roles they play and bear the significance they bear in our lives" (357).

> upon a person to re-evaluate the grounds of her attitudes and intentions and to modify them if those grounds seem faulty or insufficient.[220]

In this quote I take her to be emphasizing justification in this stronger sense: to justify is not just to evaluate intentional attitudes for fittingness but rather to undertake the activity of evaluating one's reasons and perhaps revising one's attitudes in light of that reassessment. Attitudes that can be justified in this further sense are attitudes for which we can be held to account and asked to defend. But the reason it's appropriate to respond with reactive entitlement and hold people to account in the case of these attitudes is because in this sense of justifying, as activity, justifying is a kind of control. In this sense of justification, when we say that a person is answerable we are saying that he was the one who settled what to do or believe and so we can ask him to defend or justify what he did or believed. Attitudes we can justify in this sense are attitudes over which we exercise what Hieronymi calls "evaluative control."[221]

Evaluative control is the control we exercise over attitudes that change as we form or revise our answers to particular questions. This concept of control of course depends on Hieronymi's prior assumption that attitudes like belief and intention "embody the subject's answer to some question or set of questions."[222] We control attitudes like belief and intention not at will, in the same way we can manipulate ordinary middle-sized objects in the world around us, but rather through reasoning about questions like "whether p" or "whether to φ." When we settle questions like these or revise our answers to questions like these, we control our beliefs and intentions by forming or revising them.

---

[220] "Control, responsibility and moral assessment," 386.
[221] With apologies to my readers for introducing yet another sense of control into this chapter. I will not consider here whether "evaluative control" is the same thing as Smith's "rational control," although both are clearly influenced by Scanlon.
[222] "Two Kinds of Agency," 138.

4.4.1 The key question for my purposes is whether we can exercise this kind of control over inclinations. It turns out to be rather difficult to say whether or not this is possible. Hieronymi, perhaps surprisingly, seems to think this is possible although she does not defend the view and furthermore is thinking of emotions, not inclinations. She writes,

> Far more controversially, I think the same sort of claims can be made about certain emotions – the e.g. one's resentment of S for phi-ing embodies one's answer to some range of questions about S's phi-ing.[223]

One might think that the obvious objection to the idea that emotions (and then possibly inclinations in turn) embody our answer to some question is that emotions and inclinations are famously recalcitrant to our consciously held views. So for example, I might ask the question, "Did S wrong me?," conclude that S did not, and yet continue to feel the emotion of anger.

However, I think the problem is deeper. Consider the kinds of questions we settle with belief and intention: "whether p" and "whether to φ." What sort of questions might emotions and inclinations embody answers to? One might naturally say "whether to feel [angry, sad, etc.]" and "whether to be inclined." But these questions are *about the attitudes themselves,* not about the content of the attitudes: to settle the question "whether to feel" is not to have the feeling in question any more than settling that we have reason to intend results in an intention.[224] Although these questions are normally related, in some cases they can come apart. We might have reason to hold a certain attitude (i.e. reason to believe p because believing p will help us calm down and

[223] "Two Kinds of Agency," 139.
[224] Hieronymi herself in some work describes the question relevant to belief as "whether to believe p," but I think the question is better put "whether p," a locution she uses in other work, since this is the question internal to belief.

get some sleep) even though this attitude-related reason is not reason to believe p because it has no bearing on the truth of p.[225, 226]

So emotions and inclinations cannot embody our answer to questions like "whether to feel" or "whether to be inclined," because these questions ask whether we have reason for the attitude itself and don't ask the question internal to these attitudes, the answer to which they embody. The problem is that it's not entirely clear what the questions internal to these attitudes would be. It is tempting to simply say that the question internal to, for instance, emotion is also "whether p," where p refers to for example some state of affairs about which we might be, for example, angry. But this cannot be quite right, because this turns emotion into a belief, and we can judge that, for instance, a certain sad event or angering event has occurred without feeling sad or angry. When it comes to inclination, similarly, it is tempting to say that the question internal to inclination is "whether to φ." But again, this cannot be quite right, because this assimilates inclination to intention. And it is clear that we can have inclinations without having the corresponding intentions *and* that we can have an intention without having the corresponding inclination.

One might object that beliefs and intentions can be similarly recalcitrant, although recalcitrance is I think somewhat more rare when it comes to belief and intention. A vaccine hesitant person might ask herself whether she should believe that the vaccine is safe, reason

[225] Hieronymi uses the label "attitude-related reasons" and "content-related reasons" in "The Wrong Kind of Reason." There she writes, "We can generalize the method: attitude-related reasons count in favor of an attitude by bearing on the question of whether the attitude is, in some way, good to have. Content-related reasons somehow manage to count in favor of an attitude by bearing on a different question – a question that typically does not mention the attitude itself, but concerns its content" (445).

[226] A related but separate distinction relies on the distinction between practical and theoretical reasoning. It seems clear to me that inclination is a practical attitude, and I will not hazard to guess here whether emotion is practical or theoretical. Noting the practical/theoretical distinction, however, opens up space for us to settle the practical question (what to do) or a theoretical question about action (what we judge ourselves to have best reason to do).

about the question and conclude that it is, and yet find herself still believing that the vaccine is not safe. Hieronymi argues that these cases are not evidence against the idea that our attitudes embody our answers to such questions but rather evidence that particular reasoning subjects can be conflicted, irrational, or easily swayed.[227] That is, she holds that it's possible for a person to both believe p and believe not p, settling the question of whether p in two different ways, or it's possible that such cases are cases in which the agent waffles back and forth, believing p and then believing not p and so on and so forth.

So perhaps we should say about the person who for instance judges that there is reason to be sad or angry but does not feel sad or angry that they are irrational. Smith appears to take exactly this line when it comes to truly recalcitrant attitudes. Imagining someone who judges that spiders aren't threatening but continues to be afraid of spiders, she writes,

> When this happens, I am open to a particular kind of rational criticism – namely, to a charge of irrationality. There is a direct inconsistency between the judgments I explicitly make and the judgments entailed by the attitude I in fact hold. … the very possibility of such criticism brings out the normative link between our attitudes and certain kinds of judgments in a particularly clear way.[228]

I think this response is too strong. A certain kind of criticism is warranted insofar as fear of spiders is unnecessary because spiders are not threatening under normal circumstances, but I think it is too strong to call this a kind of irrationality. Believing two contradictory propositions is irrational, but emotions are not straightforwardly beliefs, and so it is not irrational to have an emotion in conflict with one's judgment in the same way it is irrational to have two conflicting judgments. The same is true of inclinations. We should hope and work to integrate our emotions

[227] "Two Kinds of Agency," 143.
[228] "Responsibility for Attitudes," 253-4. As an aside, again, notice the importance of judgment: it is not that the attitude is itself irrational, but it is irrational because it entails a judgment that is in conflict with another judgment.

and inclinations into our agential ecology, but we should not expect them to behave like beliefs and intentions.[229]

The reason for this difference between beliefs and intentions, on the one hand, and emotions and inclinations on the other, is that inclinations are pre-theoretical (and perhaps emotions as well, although I will not stake a position). Beliefs and intentions are attitudes over which we exercise evaluative control, which is to say that when we settle questions about what to do and what is true our answers to these questions are embodied in the attitudes of intention and belief. This does not mean that beliefs and intentions are always the products of explicit deliberation. Hieronymi is very clear that we can settle questions unreflectively.[230] It is not an accident, I think, that she describes these attitudes as "embodying" our answers to questions which leaves open space for them to have been formed unreflectively. They are not consciously chosen or deliberated answers.

To put this in the terms used in Chapters 1 & 2, the deciding mind, understood in this context as our capacity for practical thinking (asking and settling questions), can work reflectively or unreflectively. We can exercise evaluative control reflectively or unreflectively. So what then do I mean by saying that inclinations are pre-theoretical? Most directly, that they are not attitudes over which we exercise evaluative control. Although we may exercise evaluative control unreflectively in many cases, it seems to me that one litmus test for our

[229] I also wonder if philosophical disdain for strong emotions may be doing some of the work here. Intuitions will vary wildly, but I worry that we are unfairly accepting of the person who judges that something sad has happened and does not feel sad – they are "stoic" or "brave" – while we criticize as "sensitive" or "hysterical" the person who feels strong emotions in the face of judgments that those emotions are unwarranted as representations of what has happened.

[230] Again, she writes, "Finally, reasoning is organized thought, not explicit deliberation. Explicit deliberation is a conscious activity that unfolds across time. Organized thought need not be. I can take reasons to bear on, or to settle, a question without explicitly deliberating about that question" ("Reasoning First," 355).

capacity to exercise evaluative control over an attitude is whether we can reflectively pose a question and reason about it in order to settle it. And although we can clearly deliberate about inclinations, I think in so doing we are either deliberating about whether to φ (i.e. reflectively deliberating about whether we should intend as we are inclined) or deliberating theoretically about whether we have good reason to be inclined. But neither of these questions is the question internal to inclination. To intend to φ or judge that we have reason to be inclined are not the same as having an inclination.

However, in saying that inclinations are pre-theoretical I also mean to capture that inclinations are a pre-theoretical counterpart to intentions. Inclinations are what would guide our action if we were the kind of creature who could not ask and answer practical questions. It is tempting to further say that inclinations are how we would settle practical questions if we were the kind of creature who could not ask and answer practical questions. This way of putting it is absurd, but it also serves to capture the basic idea: all creatures are faced with the problem of how to represent their environments and respond to it. They are faced with the problem of acting. In at least some non-human animals, inclination (or more precisely, the animal equivalent) is how they solve this problem. In the case of humans, although we have the further ability to solve the problem of action through deliberation by asking and answering practical questions, we don't thereby lose our animal nature. We are not a species of robot or alien that represents and responds to its environment in purely "rational" ways. We retain our animal, inclining nature even as our capacity to deliberate about and call into question this nature radically changes the shape of our overall agency.

Another way to put this point is to emphasize that, contra Smith and Scanlon, I do not think that an ideally rational agent's inclinations and emotions would be perfectly sensitive to her judgment. I think both emotions and inclinations have an irreducibly biological or animal origin. Thus, although an ideally rational human should have a very high degree of integration among her judgments, inclinations, emotions, etc., I think a "perfectly" rational being, a being in whom all emotions and inclinations were fully judgment-sensitive, would be a different species of being. I am tempted to say that such a being would no longer be an animal. However, in saying that an ideally rational being would not have perfectly judgment sensitive inclinations and emotions, of course I do not mean to deny what I have repeatedly emphasized about inclinations in this and earlier chapters; that inclinations do have rational content and are informed and shaped by our judgments in deep and significant ways.

Returning to the question at the heart of this chapter, then, are we responsible for our inclinations? At this point it should be clear that I think we are not answerable for our inclinations. Although there is a sense in which our inclinations can be justified or not – a sense to which I will return shortly – we cannot justify our inclinations in the same way we can justify our intentions and beliefs, because we do not exercise evaluative control over our inclinations. Although my interest in this paper is not in warranted responses to states for which we are responsible, my denial of our answerability for inclinations does mean that I think certain kinds of responses to inclinations are inapt. We should not blame, resent or demand otherwise purely on account of someone's inclinations (although we might reasonably blame, resent or demand otherwise on account of their judgments about or actions taken on their inclinations).[231]

[231] Furthermore, as discussed above in footnotes 182 & 186, it may be that certain practices of holding responsible are appropriate with respect to inclinations even if we are not in fact responsible for inclinations as such.

But are our inclinations attributable to us? And if they are, is there any moral or other significance in this assertion?

4.5 BEYOND RESPONSIBILITY AS ANSWERABILITY

In the beginning of this paper, I tried to head off from the outset that the concern that this chapter advances the no doubt absurd view that we should hold others to task for their inclinations by blaming and/or sanctioning them. However, clearly questions about when it might be appropriate to blame and/or sanction are deeply relevant to responsibility in general. This means that responsibility as a philosophical topic encompasses a huge range of questions: on the one hand, simple questions of attributability like "is this state mine or something that happened to me?" and on the other end, questions about when it's fair to sanction others. In light of this many people have proposed that we give up treating responsibility as one unified concept and instead embrace a pluralistic concept of responsibility. David Shoemaker notably divides responsibility up into three concepts (attributability, answerability and accountability) but for my purposes here I will consider the more common division of responsibility into accountability and attributability. [232] The exact details of each sub-concept as defined by the existing literature are not important for my purposes, but in brief, the idea is that responsibility understood as attributability concerns "conditions that must be met in order for an agent to be eligible for

[232] See *Responsibility from the Margins* and "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," *Ethics* 131, no. 3 (2011): 602-632. The distinction between responsibility as attributability and accountability comes from Watson's "The Two Faces of Responsibility," but it has been taken up by many others since then. For a summary of other theorists who embrace the division between accountability and attributability, see "Responsibility as Answerability" (*Inquiry* 58, no. 2 (2015): 100).

various forms of aretaic appraisal" whereas responsibility as accountability concerns "conditions that must be met in order for an agent to be eligible for moral responses that go beyond aretaic appraisal – such as the negative reactive attitudes and various forms of 'adverse treatment.'"[233]

On Smith's account, on the other hand, answerability is the fundamental concept of responsibility. Answerability grounds the fittingness of responding to someone with negative reactive attitudes, and she furthermore holds that answerability and attributability are coextensive. However, I side with those who think we need to broaden our concept of responsibility. Or, more precisely, I am at least open to the idea that we might stipulate that responsibility as such is defined in terms of answerability or accountability. It seems that we might reasonably wish to preserve the idea that to be responsible is to be eligible for negative reactions from others. However, regardless of whether we stipulate that responsibility proper concerns states for which we can be held accountable, I think we nonetheless need space for something that I will for the time being called attributability, separate from accountability.

Let me put the point in terms of Hieronymi's concept of evaluative control. In her work on control, Hieronymi argues that we can exercise a second kind of control, which she calls "managerial or manipulative control." We exercise managerial control over objects when we "take actions designed to affect them according to our purposes."[234] For example, I exercise managerial control over my coffee cup when I raise it to my lips and tip it just so in order for me to have a drink of coffee. We can also exercise managerial control over our beliefs and intentions, "by taking steps that you can predict will provide you with convincing reasons for the

[233] "Responsibility as Answerability," 100.
[234] "Two Kinds of Agency," 140.

answer embodied in the attitude."[235] So for example, if I am anxious and I cannot fall asleep because I keep wondering if I remembered to lock the door, my deciding to get up and double-check the doors is a way of exercising managerial control over my beliefs.

When it comes to inclinations, I have already argued above that I do not think we exercise evaluative control over our inclinations and as such I do not think we are answerable for our inclinations. Clearly we can exercise managerial control over our inclinations – there are a plethora of self-help books devoted to strategies for doing so. The question for my purposes is whether there is any space between evaluative and managerial control. If we deny that inclinations are states over which we have evaluative control, is the only thing left to say about our relationship to our inclinations that we can manage them in the same way we manage difficult relationships, nut allergies, and unruly pets? This space, if we can carve out such a space, is I think the proper sphere of states like inclinations. Inclinations are not actions, not things we do. Neither are they attitudes that embody our answer to a question, states that can be the result of reasoning. But neither are they simply states that happen to us, states simply to be managed.

Of course, we might just assert at this point that the mere existence of inclinations demonstrates the need for some additional fourth category beyond "things we do," "states for which we are answerable" and "things that happen to us," a category that captures some kind of agential or moral significance but does not yet rise to the level of answerability. Arguably this category just is the category of those things that are properly attributable to us. However, I think there is some reason to pause on embracing attributability as the category proper to inclinations.

---

[235] "Two Kinds of Agency," 140.

Most theorists who write about attributability emphasize that it grounds aretaic appraisals. If reactive attitudes like blame and resentment are properly directed only to states for which we are answerable or accountable (I will remain agnostic on this debate), then admiration and disdain are the attitudes proper to attributability in this sense.[236] In the case of Shoemaker, aretaic appraisals aren't just grounded in attributability; they also help us sort out the states which are attributable to us from those that are not. He writes,

> Among the excluded elements [i.e. the states not attributable to the deep self, i.e. states not attributable to us], we tend to think, are the following: (a) pains, itches, and reflexes; (b) what I call *psychic junk*, for example, the obsessional, compulsive, and addictive urges and cravings of many psychologically disordered people, as well as the random thoughts, impulses, and images that cross the consciousness of everyone from time to time; and (c) some attitudes formed on whims or as the result of altered states of consciousness. To discover the boundaries between the included and the excluded, I suggest, we can lean on our fitting aretaic sentimental responses.[237]

I have several concerns about this line of reasoning. One, it assumes that our patterns of admiring and disdaining accurately track the states attributable to us. As a matter of fact, I happen to share most of Shoemaker's intuitions about the kinds of states for which we can appropriately be admired or disdained, but I also think that a theory of attributability should have something to say about *why* certain attitudes and states are fittingly admired and disdained and others are not.

---

[236] My use of disdain here follows Shoemaker in *Responsibility from the Margins.* Disdain involves "feelings of superiority, thoughts about aretaic failure, and a motivation to be better than the disdained agent, or at least not to emulate him or her" (42). The idea that responsibility as answerability grounds aretaic appraisal if not blame and resentment was most notably defended by Watson in "Two Faces of Responsibility." The essay is among other things a response to Wolf's argument that real self views of responsibility ground only an ethically superficial sense of responsibility and thus cannot be correct as theories of responsibility (*Freedom within Reason*). To use the language put forth by J.J.C. Smart in "Free-will, Praise, and Blame," to say that a state is expressive of the agent's self (or attributable to the agent) only allows us to make "grading evaluations," i.e. grade a person as good or bad in some way, but a negative grading evaluation is not the same thing as blame and implies no fault or discredit (*Mind* 70, no. 279 (1961): 303). (Unlike Wolf, Smart embraces a revisionist account of responsibility and is happy to accept that responsibility ascriptions have no more moral depth than a mere grading evaluation.) For more on this debate and a defense of the view that mere attributability cannot ground responsibility ascriptions but we further need answerability (although in fact she thinks the concepts are coextensive), see Smith's "Control, responsibility, and moral assessment."

[237] *Responsibility from the Margins*, 44.

Why is it that I can be admired and disdained for my anger but not the mole on my ankle?[238] This case is relatively easy; moles are physical objects and as such cannot constitute our character. But what about nausea or a headache? These are trickier: they are internal. Furthermore these states can color my entire conscious awareness in the same way emotions can. Things become even muddier when we take into account the physiological roots of emotions and inclinations. Similar questions can be raised about compulsive or addictive thoughts, as described in the case of S at the beginning of the chapter. Again, I share Shoemaker's intuition that nausea and compulsive thoughts are not states for which we should be admired or disdained. But why? It is not helpful to just assert that these states are fitting objects of admiration and disdain, since that is exactly the point under consideration.[239]

> Smith's distinction between depth and significance is of some assistance here. She writes,
>
> We must distinguish between the *depth* of a form of appraisal, which concerns whether the person can legitimately be asked to justify that for which he is being appraised, and the *significance* of it, which concerns the kind of importance that attaches to his failing to meet the normative standards that apply in the particular case.[240]

So to use a more straightforward example, one of action, I may perform two different actions with equivalent depth – both chosen by me and both representing my assessment of what I had

---

[238] One might worry that there is a sense in which we admire and disdain people for things like moles: we evaluate such physical traits as beautiful and desirable or to-be-avoided all the time. For the time being, I will set aside this concern with the stipulation that the kind of admiration and disdain I'm referring to here are "agential," to use Shoemaker's phrasing – I refer to admiration and disdain of an agent's character (*Responsibility from the Margins*, 40-42).

[239] In fairness to Shoemaker, his account is of course more nuanced and developed than this. He holds that we are attributable for our cares, commitments, and care-commitment clusters, although not every care, commitment, or care-commitment cluster will be of moral significance (see *Responsibility from the Margins,* 50-59). However, my fundamental worry here remains: I worry that his primary argument for thinking that we are attributable for this list is just the intuition that these are fitting objects of admiration and disdain, while my interest is in the deeper question of *why* these states are candidates for admiration and disdain while others are not.

[240] "Control, responsibility, and moral assessment," (385). Scanlon makes a similar distinction in perhaps more intuitive terms, distinguishing content (itches, e.g. lack morally significant content) and control (itches are also not under our control, not even in an ideally rational agent) ("Reasons and Passions," 173).

reason to do, but one action (chasing down and returning the lost dog of my elderly neighbor) may have great significance than another (wearing my blue shorts instead of black ones because I wore black shorts yesterday), because the moral standards apply to the content of the first action but not the second. We can also imagine cases in which the content of two different attitudes is morally significant – S's violent images and the intention to harm, versus a case in which someone is actually harming another person – but the depth is very different, being very shallow in the former case.

Thus, one thing we might be saying when we express the intuition that nausea is morally insignificant, not a fitting object of admiration or disdain, is just that nausea is not significant in its content. There are no normative standards to fail. Emotions like anger on the other hand are clearly candidates for significance. Inclinations are perhaps less clear than emotions – the inclination to for example have a pickle with my lunch seems relatively thin on significance – but I think we can see how inclinations might have moral significance, even if not every inclination does.

However, although I think it is helpful to consider content and significance in trying to tease out whether inclinations are attitudes for which we are attributable, the trickier and more important question is depth. After all, if it turned out that, for instance, S's violent thoughts have no depth whatsoever, the fact that the images have significance would be of no relevance to our aretaic assessment of S. Smith of course understands depth in terms of answerability, but answerability is already off the table when it comes to inclinations.

At this point, then, I think it is time to set aside talk of attributability if by attributability we mean a feature that grounds aretaic appraisals. I am sympathetic to the idea that we should

carve out space for a concept responsibility that grounds aretaic appraisals, admiration and disdain, but is separate from answerability and/or accountability. However, although some inclinations do intuitively seem morally significant and reasonable objects of admiration or disdain, we do not want to let the content of these attitudes do all the work in convincing us that they are of moral significance and fitting objects of admiration and disdain. Rather, the deeper and more difficult question is whether inclinations have depth.

4.6. INHABITING OUR INCLINATIONS

In "Responsibility for Attitudes," Smith writes,

> [Theories that ground responsibility in voluntary control obscure] the special nature of our relation to our own attitudes: we are not merely producers of our attitudes, or even guardians over them; we are, first and foremost, inhabiters of them.[241]

Although I do not agree with her further claim, that we are answerable for such attitudes because we can be asked to justify them on account of their connection with our evaluative judgments, I think this description nonetheless provides a nice starting point to thinking about our relationship to our inclinations: we don't simply produce them (and if by "produce" we mean choose and bring about, then we do not produce them at all). We're not simply guardians of them, repositories of conscious experience in which they occur. We inhabit them. But above I objected to Shoemaker's account on the grounds that he does not have enough to say about *why* certain attitudes are states are attributable to the agent. So can I say anything in defense of this

[241]"Responsibility for Attitudes," 251.

conclusion? Why is it that we can inhabit our inclinations, but presumably not sneezing or a headache?

The first and most straightforward thing to say here is to point out that inclinations are intentional. They are imperatives, but they are imperatives whose force depends in part upon their content. In short, inclinations contain an inchoate representation of why something is worth doing. Because of this they can be assessed for justification in the first of the two senses outlined above, i.e. they can be assessed as accurate or inaccurate representations of our reason to act. So for instance, imagine I am climbing a tree and I have the inclination to jump off the branch because it's a sunny breezy day and I'm caught up in the joy of being up in the air and for a brief foolish moment I want to know what it's like to float – plummet – through the air. And then I laugh and dismiss the inclination as highly ill-advised. This marks an important difference between inclinations and physical sensations, i.e. nausea. Although I may have inclinations that stem directly from nausea – the inclination to run away from the liver frying in the kitchen, for example, nausea purely as a physical sensation is not about anything. And as such it is not an object of justification, although we might otherwise evaluate it as "unexplained" or "not making sense" in light of its causal origins or confusion about its causal origins. Because nausea is not intentional, it is not a component of agency in the most minimal sense, whereas inclinations are expressions of our agency on account of their intentionality.[242]

Since inclinations are intentional, they are not just agential in a minimal sense but furthermore candidates for moral evaluation when their content has moral significance. An inclination to get far away from the liver frying in the pan may be fitting, but it is not an

[242] By "minimal sense," I mean to convey that saying inclinations are agential in this way is not meant to entail anything further about our accountability or eligibility for aretaic appraisal on account of them.

inclination of moral significance. The inclination to harm someone else has morally significant content, even if the inclination itself is fleeting and is not grounds for negative assessment of the character of the agent in question.[243]

Furthermore, the intentionality of inclination makes them candidates for what I will call "uptake" or "integration" into one's agential ecology. Clearly we can exercise managerial control in Hieronymi's sense over both inclinations and nausea. We can act on both inclinations and nausea to manage them. However, there is nonetheless an important distinction in the results that come from exercising managerial control over inclinations and mere physical states like nausea. No matter how excellently we manage our nausea, we cannot manage our nausea into being itself part of our agency or more central to our character. One could take up or integrate nausea into agency by attaching an agential attitude to it, i.e. deciding to care deeply about nausea or make overcoming it a central aspect of one's identity, but this would not make nausea itself an expression of agency. It would simply make nausea the object of an agential expression, e.g. caring. When we exercise managerial control on our inclinations, however, it's possible to thereby make our inclinations themselves more or less integrated or enmeshed in our agency.

Say for instance that I have the desire to become a writer and I furthermore want to be the kind of person who enjoys writing and at least sometimes genuinely wants to write, i.e. is inclined to write. But in my current state, I only have the occasional flicker of an inclination to write. We can exercise managerial control with the aim of encouraging this inclination, for instance watching movies that feature inspiring writers or reading biographies of my favorite

[243] My inclination is that a fleeting, out-of-the-blue inclination to harm another person is not an inclination for which the agent should be negatively appraised, but I am trying to remain agnostic on this point for the time being since the nature of our responsibility for inclinations is the problem under consideration.

authors. I can write mantras like "I am a writer" and "Today I want to write" and stick them on cards around my house to encourage me to want to write. I can bribe myself to write with various rewards for writing. I can consciously direct my attention towards all the things that make me want to be a writer, trying to make these considerations so natural and familiar that my instinctive mind latches on to them and says "Write!" In this scenario, I am exercising managerial, not evaluative control over my inclination. But in successfully exercising managerial control over my inclination, in actually acquiring the inclination to write, I thereby making the inclination to write more central in my agency. I am making myself into the kind of person whose teleologically understanding of the world sees writing as *to-be-done* and the kind of person who cognizes writing as good by taking pleasure in it (in addition to being a person who cognizes it as good by judging it good). This marks an important distinction from cases in which we for instance exercise managerial control over objects.

4.6.1 To return to the case with which the chapter began, I think I am now prepared to give a more in-depth response about the case of S. The first thing to note about S is that as the case is presented, his thoughts appear to be just that, mere thoughts or ideas as opposed to inclinations or beliefs. This is significant because if his thoughts make no claim to represent the world or how he ought to act - if they have no force in Karl Schafer's sense of the word[244] - then they are not proper objects for justification in either sense.[245] But for the sake of argument, imagine that S's thoughts are not mere thoughts but are in fact inclinations: he doesn't just experience the mere

[244] See Schafer's "Perception and the Rational Force of Desire," *The Journal of Philosophy* 110, no. 5 (2013): 258-281.

[245] I will remain agnostic on the more controversial question of whether S's thoughts are intentional.

image of stabbing someone, but furthermore feels the fleeting inclination to stab someone. So what should we say about such an inclination? Is it a state attributable to S?

Perhaps unhelpfully, my answer is: it could be. Or more specifically, I would say that they are his in the sense of being attitudes that he can inhabit, but the mere fact that he could and occasionally has inhabited them tells us nothing of significance about S's deep self or moral character. Rather, I want to emphasize that because S's inclinations are attitudes he *can* inhabit, unlike for instance nausea, S is responsible for his inclinations in a forward-looking sense. He has the responsibility to exercise managerial control over the inclinations so that they do not become him, because they are the sort of state that he can inhabit and furthermore the sort of state that could constitute his moral character. Perhaps ironically, in this specific case the best response to his inclinations is to calmly proceed with life and ignore them.[246] In treating the inclinations as insignificant and unconcerning, they remain insignificant and unconcerning. But if on the other hand S failed to exercise managerial control, or exercised the wrong kind of managerial control, these inclinations are the kind of thing that could become attributable to his self in a deeper and morally significant sense.[247]

[246] It may be that this is generally a good strategy with respect to recalcitrant inclinations, but I will remain agnostic on this point. Intuitively it seems that other inclinations would need more active management. Regardless, one upshot of my view is that one significant function of therapy is to help us exercise managerial control over recalcitrant attitudes.

[247] This is after all why certain forms of mental illness, for instance schizophrenia, are morally tragic. A person with untreated schizophrenia may become someone else, someone unrecognizable and morally upsetting to their friends and family members. They are not responsible for becoming such a person, but schizophrenia is morally tragic precisely because it can drastically alter one's self and moral character if untreated.

5. Tempting Others

Until now the dissertation has primarily considered temptation from the first-person perspective. What kind of force do tempting inclinations exercise on us? Is resolution a unique rational method of forestalling temptation? What kind of responsibility, if any, do we have for tempting inclinations? In this chapter, I shift to considering temptation in the interpersonal sphere. Ultimately, my interest lies in whether there is any special wrong in influencing another person by tempting them as opposed to influencing them by more strictly rational means, but I must first begin with a brief puzzle: is it possible to tempt another person?

In one sense, obviously yes. We are tempted by other people all the time. We purposefully tempt other people all the time. Still, on the understanding of inclination I've adopted thus far in the dissertation, this is at least a little puzzling. For the time being, set aside the question of temptation and focus solely on inclination: is it possible to bring about an inclination in someone else? I ask the question from the perspective of our everyday practices, not that of science fiction. Say I am trying to get my sister to want to come to a play with me: can I do this? In the previous chapter I repeatedly emphasized that we do not have voluntary or even evaluative control over our inclinations. We can exercise managerial control over our inclinations and thus control them indirectly, over time, but we cannot bring ourselves to be inclined on the spot. So how then could I cause another person to have an inclination *sans* sci-fi interventions?

The obvious answer is to say that when we cause someone else to be inclined, we trigger a latent inclination. In some cases, this might only require exposing the person to the object of their inclination – a recovering alcoholic might only need to be around someone else drinking a

beer in order to have the inclination to drink – or it might require encouraging the person to link an existing inclination with a new action. This could happen through a bribe – I offer to buy you the dessert I know you always want to eat if you come with me to the play – or by reframing the other person's inclination for them – I convince you that what you *really* want is not to go to the club to go dancing, but just to get out and do something new, and what would be more new and adventurous for you than going to the theater with me! Of course, these tactics can be used to get others to act as we wish regardless of their inclinations, but what's important for my purposes is that we can use such methods in order to try to get other people to be inclined as we wish.[248,249]

5.1 TEMPTING PATERNALISTICALLY

Having briefly considered what it means to tempt another, I turn now to the wrong of tempting another. This means that I will be focusing in particular on trying to cause a conflicting, i.e. tempting inclination in another person. It may be that there is something suboptimal about trying to cause another person to have an inclination in general, but in this chapter, I will focus

[248] I say "inclined as we wish" rather than "act as we wish" because my emphasis here is on inclination. Furthermore, it might that I am successful in getting you to want to come to the theater although you do not in fact come. "Oh, my favorite dessert!," you might say. "Well, damn it, now I really want to come. But I can't – I already made plans."

[249] No doubt this accounts for many, if not most cases of causing and inclination in another person, and in turn tempting another person. After all, when we purposefully try to tempt another person, we often go for their known vulnerabilities. But the more interesting question is I think whether we can tempt someone to do something they are not already inclined to do without appealing to an inclination they already have. It seems clearly possible to do this, other people can be the impetus for us to have new inclinations, but I don't think it's possible to purposefully undertake this as an activity because there is no consistent way to bring about the intended result. It may be helpful to link back to the role pleasure plays in inclination at this point, and notice that we cannot control what others find pleasurable. We can do our best to redescribe objects or events in ways that they will find pleasurable, and we can do our best to bribe them according to what they already take pleasure in, but these are both ways of trying to form in inclination in someone by linking the new inclination to a pre-existing inclination. We cannot force someone to take pleasure in something they do not find pleasurable.

on trying to bring about tempting inclinations in particular. What, if anything, is wrong with influencing another by tempting them as opposed to influencing them through rational persuasion? Consider the following cases:

- **Attractive Advertisement**: A company hires attractive models to pose for photographs with their new headphones. When Greg sees the advertisements, he is drawn to the images of cool and attractive people and subconsciously wishes to be and appear more like them. He doesn't need a new pair of headphones and had intended to cut back on spending for the rest of the month, but he now feels tempted to buy a pair of new headphones.[250]
- **Paleo Mom**: Stacy devotes a lot of time and energy to reading about healthy eating and she concludes that a broadly paleo diet is the healthiest way to eat. In light of this, she and her partner raise their children on a generally paleo diet. However, as a pre-teen, one of her children decides to become a vegan. Stacy agrees to let him try a vegan diet for one week, but she plans to cook his favorite meats at the end of the week to tempt him back to the diet she sees as healthiest.

Both of these are plausibly cases in which one person (or entity) tempts another. But do these cases involve wrongdoing? Is it wrong to tempt another person?

Clearly temptation can be wrong when tempting another involves the intention to harm or humiliate or the object of temptation is morally wrong. However, these explanations are not especially helpful or relevant when it comes to these cases. Although most people dislike being manipulated by advertising, Attractive Advertisement describes an extremely common occurrence: any marketing or advertising employee who didn't employ tactics like using attractive models in ads would be promptly fired for being bad at their job. In other words, although certain advertising tactics may be straightforwardly manipulative and wrong and advertising in general may be morally queasy, it would require a very high bar to reject all

[250] If we asked Greg why he desires a new pair of headphones, the attractiveness of the models would not figure in his answer because he does not upon reflection endorse the claim that buying these headphones will make him more physically attractive, although something like this claim seems to be subconsciously at work in his tempted response to the advertisement.

advertising as such as morally wrong. In Paleo Mom, on the other hand, Stacy takes herself to be doing the right thing: she sees herself as acting to protect her son's health. Of course, from the perspective of ethical veganism, she tempts her son to do something morally wrong when she tempts him to eat meat, but intuitively there is something wrong with her action even if we grant for the sake of argument that it is morally permissible to eat meat.

However, although neither of the cases involves straightforwardly malicious intention, I also find it plausible that both cases are problematic in some respect. Thus, although I think that one way tempting another can be wrong is for the tempter to have malicious motives, I will not consider this category of wrong in this paper, in part because it is not unique to temptation. I can have malicious motives with respect to you - motives to control, harm or simply "mess with," or motives to entice you to do something wrong – and not use temptation in carrying out my motives. Tempting another person is one possible way of expressing or acting on malicious motives, but the wrong of have such malicious motives is in the motives themselves and is not intrinsic to tempting another as such.

Similarly, it seems that tempting another can be wrong when and because someone is tempted to do something morally or prudentially wrong. In such cases, the wrong is located primarily in the wrong of the action in question, and again, it's possible to try to get someone to do something morally wrong without tempting them. Thus, although malicious motives and/or morally wrong objects of inclination are no doubt common reasons for which it is wrong to tempt another person, in this paper I will set aside such cases and focus on temptation in its own right. Furthermore, I will set aside at the beginning the view that tempting another person is always wrong and focus on a more restricted question. Say that A correctly judges that B should

φ. All things considered, is it worse if A attempts to influence B to φ by tempting them as opposed to influencing them by rational means?[251]

5.2 UNDERMINING AUTONOMY

One intuitive response to the above question is that in attempting to influence someone by tempting them we dishonor or undermine their autonomy, whereas rational influence is compatible with autonomous governance. Appealing to autonomy to explain what's wrong with tempting someone in particular or manipulating others in general can take one of at least two forms. One, we might think that temptation undermines autonomy. Although most theorists seem to agree that manipulation of any form (including tempting another) does not literally remove one's autonomy, some hold that forms of manipulation like seduction (perhaps a species of tempting another) and deception prevent the victim of manipulation from exercising her own will and choice.

For example, in "Between Consenting Adults," Onora O'Neill suggests that the victim of seduction "lacks insight into what is proposed, [and so cannot] consent to it."[252] Paul Hoffman, writing about seduction, claims that "Seduction is another way in which our will is overpowered.

[251] Although I will in this chapter remain focused on the question "All things considered, is it worse to influence someone by tempting them rather than by rational persuasion?", in the following I will appeal to discussions of manipulation in general in trying to puzzle out what (if anything) makes tempting others wrong. It seems clear to me that tempting another is a species of manipulation, although manipulation includes many other forms of influence beyond temptation. For more on the many species of manipulation, see Anne Barnhill's "What is Manipulation?" in *Manipulation: Theory and Practice*, edited by Christian Coons and Michael Weber, Oxford: Oxford University Press, 2014, 51-72 and Robert Noggle's "The Ethics of Manipulation," *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.

[252] "Between Consenting Adults," *Philosophy & Public Affairs* 14, no. 3 (1985): 254.

When we are seduced, what we propose to do is not really up to us, it is up to our seducer."[253] The guiding metaphor behind this concept of manipulation is perhaps that of a puppet being jerked around: when we are manipulated and tempted, so the view goes, our actions stem not from our own will or choices but rather from the will of the one manipulating us.[254]

However, there are familiar problems with thinking of tempting others (along with other forms of manipulation) as literally removing or undermining their autonomy.[255] First of all, although manipulation and tempting another often involves some form of deception, we constantly make autonomous decisions in the absence of full information and in less than ideal circumstances.[256] Furthermore, the metaphor of "overpowering" is problematic because it implies that our freedom with respect to temptation is contingent: if the temptation is strong enough, it might take over our will, and we are helpless in the face of it. But this again implies that temptation is something that happens to us, since we cannot be overpowered by ourselves.

So perhaps we should rephrase the autonomy objection. Perhaps the problem with tempting another person is not that tempting them literally undermines or takes away their

---

[253] "Freedom and Strength of Will: Descartes and Albritton," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 77, no. 2/3 (1995): 254.
[254] Claudia Mills, "Politics and Manipulation," *Social Theory and Practice* 21, no. 1 (1995): 106.
[255] Claudia Mills offers a metaphor concerning the effects of temptation on autonomy that I find helpful. She writes, "It may be helpful here to compare autonomy with one's ability to drive. If someone deliberately sets out to make the road bumpier, there is a sense in which he does not interfere with my ability to drive: I am still the driver I am, whatever the weather or road conditions, with the same opportunities and responsibilities to exercise my skill. … Yet in some sense the person creating the road obstruction does interfere with my driving, for he certainly makes it harder to drive, harder to do a good job driving" ("Politics and Manipulation," 106-7).
[256] Although coercion is intuitively stronger than manipulation and I will not consider their relationship here, Stephen White makes the same point with respect to coercion, writing "However, though coercive threats may often involve distorting influences such as fear, they primarily work by altering the reasons the recipient has for pursuing certain options over others. But responding to changes in one's situation that make certain alternatives more attractive or reasonable is just what acting rationally and autonomously normally involves. The fact that the situational change that affects one's reasons is due to a *threat* is not in this respect relevant. To be sure, from the point of view of the victim, the overall situation is hardly desirable. But people frequently make free, rational decisions in less than ideal circumstances" ("On the Moral Objection to Coercion." *Philosophy and Public Affairs* 45, no. 3 (2017), 211).

autonomy, but that in so doing we fail to *respect* their autonomy. Although this seems closer to the mark, we run the risk of asserting a tautology if we do not flesh out the autonomy objection by defending a particular conception of autonomy and explaining why that conception is incompatible with tempting another person. Sarah Buss makes this point in her article "Valuing Autonomy and Respecting Persons," writing

> The belief in autonomy's moral significance presupposes a special understanding of what is involved in governing oneself. This means that we cannot derive moral conclusions from the mere concept of autonomy; we need to rely on a particular, substantive conception according to which governing oneself is at odds with being manipulated or deceived[257]

That is, it's intuitive that one condition of being autonomous is being free from interference, including interference in the form of being tempted by another person. But we cannot appeal to this conception of autonomy in order to explain why tempting another is wrong, because the conception already presupposes freedom from tempting influences and interference. Thus, although we might expect that any account of temptation's wrongness will appeal to autonomy in some way, any such account must first defend a view about the nature of autonomy and the nature of tempting another and identify the incompatibility between the two.

5.2.1 Although Schapiro focuses on weak-willed action and not manipulation via inclination, I think an extension of her account offers one interesting way to flesh out the autonomy objection to tempting others. Recall that Schapiro's view of weak-willed action holds that nothing can pressure a free will but itself, and so we are weak-willed when we fail to take responsibility for deciding for ourselves.[258] Schapiro furthermore holds that we can be weak-willed not just in

[257] "Valuing autonomy and respecting persons: Manipulation, seduction, and the basis of moral constraints." *Ethics* 115, no. 2 (2005): 196-7.
[258] *Feeling Like It,* 147-8.

response to inclination but in response to social pressure as well. Failing to decide for oneself and simply conforming to a social script, "[escaping] into our social environment," is also a form of being weak-willed.[259] Although Schapiro's discussion of weak-willed action is put in terms of freedom, not autonomy, I think it's a reasonable interpretation of her view to say that autonomy is not only a capacity but furthermore a *responsibility*, and thus to be weak-willed is to fail to live up to one's responsibility to govern one's own actions.

When it comes to what's wrong with tempting another, my extension of Schapiro's account would identify not one but two problems. One is the problem of influencing another person through inclination. Assuming tempting another means aiming to get them to acting directly on an inclination, without incorporating it (as Schapiro would say we must), tempting another would be wrong because it encourages the tempted to give up their responsibility to decide for themselves and just follow the lead of their inclination. We might also put this in terms of encouraging someone to act like an animal, directly on inclination, *sans* incorporation.

However, tempting another is social, and so I think this social element would be the second problem identified by Schapiro's extended account. Schapiro's account of weak-willed action already allows that we can be weak-willed in response to social pressure, and so I think it would be a natural further step to conclude that tempting another (and other forms of manipulation) are wrong insofar as they encourage others to give up responsibility to decide for themselves and instead just conform to the wishes of other people. Of course, this wrong could also apply to more uniquely "rational" forms of influence, but it is relevant to tempting another insofar as that also is a social exchange.

---

[259] *Feeling Like It,* 160.

The social element of the account is helpful and interesting, and I think there is something importantly right about the idea that manipulating another person is wrong when it encourages them to give up their responsibility to decide for themselves. However, focusing too much on the idea of being unduly influenced or giving up on deciding for oneself can also obscure the question under consideration here: is there something uniquely wrong with influencing another person by tempting them as opposed to using other forms of rational influence? Return to the example of Paleo Mom, and again stipulate that Stacy is right that her son should eat meat. Is there something wrong with Stacy trying to get him to eat meat by tempting him with a food he finds delicious as opposed to giving him a stack of literature on, for instance, the health dangers of removing meat from one's diet?[260] One thing Schapiro's extended account helpfully leaves room for is the idea that we can problematically manipulate (or influence, if we wish to reserve "manipulate" for arational or conative forms of influence) someone by pressuring them too much, even if the way we are pressuring them is characteristically rational. A parent who constantly sends their teenage child articles about the diet they regard as healthiest is placing undue pressure on the teenage child, even if the parent happens to be right and their chosen method of influence is "the facts."

Setting aside undue social pressure, if there's something uniquely wrong with tempting another, that wrong lies in the idea that tempting another is a kind of influence incompatible with rational self-governance whereas other forms (e.g. sharing facts) are not. Put in terms of autonomy, we might say that we need an account that explains why inclination in particular is a threat to autonomous self-determination. Schapiro's extended account does have an answer to

[260] If the example begins to seem stretched, simply flip the scrip and imagine the mother as a vegan or vegetarian. I use this case in particular because the case is real.

this question. Although she doesn't use this exact language, on Schapiro's account inclinations are always a bit of a lingering threat, since they offer a way out of the burden of freedom. She writes, "something about their nature [the nature of inclinations], independent of variations in content, makes it the case that when we are inclined, we are faced with a perfect opportunity to flee the burden of freedom."[261] Thus, on this account, the reason it's wrong to tempt another person is because in so doing we are trying to bring about a situation where a person is given an opportunity to abdicate responsibility and flee the burden of freedom.

The advantage of this account is that it has a story about the nature of inclinations such that it can explain why it's wrong to influence someone by trying to bring about a tempting inclination in them as opposed to influencing them rationally. However, I also think this account of weak-willed action and our relationship to our inclinations is just too strong. I won't rehearse the arguments from Chapter 1 against this view of inclinations and rational self-governance, but will rather just move on to consider other theories of inclination and manipulation that might better capture what's wrong with tempting another person.

5.2.2 One promising route suggested by this interpretation of Schapiro's views of weak-willed action is the idea that bypassing or subverting someone's reason is the key problem with tempting another person. We don't have to think that tempting another is a form of encouraging them to flee from the burden of freedom in order to think that tempting another in some sense interferes with or skirts our rational self-governance and thus is wrong. As a theory of manipulation in general, this position is what Moti Gorin refers to as a Bypass or Subvert View (BSV) of manipulation.[262] The view holds that manipulation

---

[261] *Feeling Like It*, 147.

[262] "Towards a Theory of Interpersonal Manipulation," in *Manipulation: Theory and*

> influences people's choices in ways that circumvent or subvert their rational decision-making processes, and that undermine and disrupt the ways of choosing that they themselves would critically endorse if they considered the matter in a way that is lucid and free of error. … Appeals to emotions, needs, or character flaws also count as manipulation because they, too, subvert the rational self-government of the person.[263]

Gorin unpacks this view in terms of two main tenets, as holding that manipulation is wrong because it

1. "fails to engage the rational capacities of the influenced agent" or
2. "engages these capacities in some way that undermines their function (that is, it subverts them)."[264]

I won't take a position here on whether BSV is correct as a general theory of manipulation, but I nonetheless think it's a helpful starting point for unpacking what's wrong with tempting another person as opposed to influencing them rationally. However, the view requires unpacking when it comes to tempting inclination in particular. Since I hold that inclination is reason-laden, more needs to be said about why tempting inclination is irrational or arational as a means of influence.

Examples of manipulation that bypass rational capacities altogether would include, at one extreme, hypnosis or coercion, and at the other end, sales tricks like baking something that smells delicious when showing a house to prospective buyers.[265] An example of manipulation that undermines or subverts rational faculties would be for instance feeding someone false information, which would manipulate them via belief by tricking them into a false belief. Where does tempting another person fall on this conceptual scheme? I think it is possible for us to undermine inclination in a way parallel to the undermining of belief in the case of deception.

Practice, edited by Christian Coons and Michael Weber, Oxford: Oxford University Press, 2014, 89.

[263] Allan Wood, "Coercion, Manipulation, Exploitation," in *Manipulation: Theory and Practice*, edited by Christian Coons and Michael Weber, Oxford: Oxford University Press, 2014, 35.

[264] "Towards a Theory of Interpersonal Manipulation," 90.

[265] See for instance Anne Barnhill, "What is Manipulation?", 58.

Assuming the function of inclination is to pick out goods and motivate us to act towards them as applicable, we undermine someone's inclination when we trick them into wanting something that is not in fact worth wanting. Such manipulation of inclination characteristically involves cognitive deception as well, although it may be possible to mislead someone's inclination without deceiving them in the standard, cognitive, sense.

However, insofar as this is wrong it is wrong as a form of deception. Undermining or subverting inclination is not what's uniquely wrong with influencing something through tempting them, because it may be that when we tempt someone we are tempting them to want something that is worth wanting, or more precisely, is not a mistake to want even if it's permissible to lack a desire for the thing in question. Again, think of the paternalism in Paleo Mom: if Stacy is correct that eating meat is morally permissible, that eating meat is healthier for her son, and furthermore her son really does enjoy the taste of meat, then there's nothing wrong in her son being inclined to eat meat. The inclination is not itself misguided. The question is rather whether Stacy is doing something wrong in trying to persuade her son through the means of inclination rather than for example a discussion or presentation of facts.

This then leaves the first tenet of BSV: perhaps it's wrong to influence another person by tempting them because tempting inclination fails to engage their rational capacities. But this immediately requires clarification: what counts as a "rational capacity"? Our will? Our belief? Are rational capacities the capacities characteristically associated with autonomous self-governance? Or are rational capacities any capacity that is sensitive to our assessment of our reasons? Any capacity that expresses rational content itself? I think what proponents of BSV have in mind are probably capacities characteristically associated with autonomous self-

governance, like will and belief. Still, my account of inclination views inclinations as states that are reason-laden and furthermore as states that aim towards settling our action. This makes it odd and inappropriate to assimilate influencing someone via tempting with influencing them by hypnosis or coercion. Take an easy and obvious case: if I grab your hand and punch it through a window, there's an obvious sense in which I'm failing to engage your rational capacities. I didn't engage your capacity to consent, your capacity to decide for yourself, your capacity to form judgments about what is best to do. I literally forced you to act. The same could be said about hypnosis.

But influencing someone via inclination, by tempting them, does not bypass their rational capacities in this sense. A tempting inclination does "bypass" our rational capacities insofar as it attempts to direct our action without input from the deciding mind. However, I think there's reason to be dissatisfied with this view. As I've emphasized in previous chapters, part of why temptation is hard to resist is because temptation is reason-laden. Temptation is not on my view a pure imperative. If it were, it would no doubt be much easier to resist, unless it were a pure imperative with physical manifestation that we literally, bodily could not resist. Being tempted involves seeing something as *to-be-done*, but it also involves being struck by the reason(s) one has for so acting.

However, the fact that tempting inclination is reason-laden is why it can so pervasively interfere with our deciding mind. In this sense, I don't think tempting another so much bypasses rational faculties altogether as it undermines one's will. Think of the individual case, setting aside for the time being interference from another person: in the individual case, temptation often undermines our capacity to sort through what reasons bear on the question at hand. Say that I

commit to being a vegan for moral reasons, but I really love cheese. At home, I just don't keep cheese around, but now I'm out at a party and the host has prepared a tray of delicious cheeses, and I am very much inclined to eat some cheese, to cheat "just this once." In this situation, I don't think it's right to say that my tempting inclination bypasses my rational faculties altogether. To the contrary, the tempting inclination will affect me *through* the rational capacities in question. When my deciding mind is trying to decide whether I'm going to break my commitment to veganism and eat some cheese, my inclination to eat the cheese will direct me to eat the cheese, leading me to see cheese as *to-be-eaten*, but it will also lead me to focus on the reasons I have for eating cheese: the cheese is delicious; if I don't eat it, the leftovers might be thrown away, and how does that help the mammals whose milk was used to make it?; it's a special occasion and it's just this once.

Of course, these reasons sound very suspiciously like rationalizations, or at least the latter two do. But that is the point. If veganism is morally correct, then I really should not eat the cheese. (Some consequentialist views might hold that I can eat the cheese if it's on the verge of being thrown away, but certainly I can't know that at the beginning of the party, and it might be prudentially unwise to give myself the taste for cheese again.) And if my mind were "clear," I would be able to see that this is the case, settling the question of what to do in light of what I take to be the many very good moral considerations, and sadly setting aside the fact that I find cheese delicious. But when I am tempted, my mind is not clear in this sense. As discussed in Chapter 3, when we are tempted it is very difficult to think through our reasons from a neutral deliberative space.

5.3 UNDERMINING THE WILL

I will not commit to the strong claim that a tempting inclination always undermines our will, but a weaker claim seems true: that tempting inclination often undermines our will in the sense sketched out above.[266] We can apply this account to the case of interpersonal temptation and say that the reason it is worse to influence someone by tempting them as opposed to influencing them by rational means is that in so doing, we are or risk undermining their will. More specifically, tempting another person can undermine their will in one of two ways:

1. Presenting them with bad reasons
2. Making it difficult for them to act as they intend

5.3.1 In the veganism case discussed above, I focused on temptation's capacity to draw our attention to bad or inadequate reasons. In some cases, temptation works by overemphasizing one or a few of the relevant reasons. For instance, in the Paleo Mom case, one problem is that in tempting her son to eat meat, Stacy is encouraging him to settle the question of what to eat with respect to only one consideration, namely, what he finds tasty. What we like to eat *is* relevant to the question of what we should eat, but this consideration must be balanced against others and insofar as Stacy is encouraging her son to allow this consideration to swamp out other important considerations, she is wronging him.

[266] I will remain agnostic on the strong claim for the time being. I hesitate to endorse the stronger claim because it seems like there might be some cases in which being tempted to do a morally virtuous thing in fact "perfects" the will insofar as it directs the will towards good reason. Or, in a less moralized sense, imagine that we are trying to tempt someone to not commit suicide, to tempt them to want to live and we are furthermore specifically trying to engage their desire as well as their belief. Is this really an instance of undermining their will? I suspect not.

The Attractive Advertising case is somewhat more complex. In this case as with Paleo Mom, part of the problem is that tempting inclination is going to focus Greg's attention on considerations that count in favor of buying new headphones, thus making it difficult for him to deliberate carefully about his reasons. However, the further problem in this case is that Greg's inclination is itself undermined insofar as it's responding to something that isn't really a good-making feature of the action in question. There may be some minimal reason to be a cool and attractive person, and it may be reasonable for Greg to have some minimal admiration or liking towards such persons. But in this case Greg's tempting inclination responds with pleasure to the attractiveness of the models and thereby leads him to desire a new pair of headphones even though the attractive and cool models have almost nothing to do with the goodness of buying a new pair of headphones. If on the other hand the advertiser tried to get Greg to desire their headphones by drawing his attention to good-making features of the headphones, like their noise-cancelling feature or their crystal-clear sound quality, this would not count as undermining inclination on my view, although the advertising could very well be manipulative in some other way. Thus in this case I think Greg's rational capacities are corrupted by bad reasons on two levels. On one level, his inclination itself is undermined or deceived, and then once he has the inclination to buy the headphones this further undermines his ability to deliberate clearly about what to do.

However, presenting someone with bad reasons cannot be the sole reason it is worse to influence another person by tempting them as opposed to influencing them rationally.[267] One, I

[267] Although my focus here is on tempting another in particular, this holds true of manipulation in general as well. For instance, Moti Gorin endorses the view that "it is manipulation's failure to track reasons that renders it morally suspect" ("Towards an Interpersonal Theory of Manipulation," 92). The actual view is more detailed than this, but we need only the basic point for my purposes, which is that we can wrong another person by "bullying them" with

want to allow for the possibility that there are cases in which tempting another person is a way of attempting to put them in touch with good reasons. Consider the following concrete case:

> Say that Josh resolves to run a marathon without training, in order to test his physical limits. I on the other hand have run several long-distance races and am fairly certain that Josh will incur injuries in attempting to run the distance without preparation, and furthermore, I believe that there would be more effective and safe ways to test his physical limits (e.g. to train for a marathon and then see how quickly one's body can finish such a long race).

In this case, I take it that there's nothing outright irrational about Josh's resolution. It's not that Josh is inconsistent in his reasoning or utterly clueless about the relevant dangers. But I also take it that this resolution is a bad one, and this fact might make it seem permissible for me to try to tempt Josh to stay home from the marathon. In other words, it seems like I could in this case tempt Josh to give up on the race without presenting him with bad reasons. Of course, this would have to be a particular kind of temptation; I would have to be sure that I really was giving him good reasons for desiring to stay home and skip the race. But I find it plausible to think that it's possible for me to tempt Josh to quit without preventing him from deliberating clearly about the situation and undermining his will with bad reasons, and thus tempting another does not necessarily involve presenting them with bad reasons.

On the other hand, it's possible to present someone with bad reasons without tempting them. Above I've described the presentation of bad reasons that happens when we're tempted in terms of our attention being overwhelmingly focused on the reasons we have to act as we're tempted to act. We can replicate the kind of "repeated attention" to bad reasons that happens in temptation in a case that involves no tempting desire. Consider:

---

good reasons. One might object that this behavior is wrong but does not deserve the label of manipulation, however. Certainly it is very difficult to define "manipulation" precisely because the word is used to refer to such a wide variety of things.

> Bella really wants a dog, and she's trying to convince her partner to agree to getting a dog. She thinks dogs are cute, and she imagines all the adorable Instagram content she would be able to get if she had a dog to photograph. Every chance she gets, she reminds her partner of the fact that she wants a dog and her reason for wanting a dog: over dinner, when driving to the grocery story, before he leaves for work, multiple times per day, every day.

On account of Bella constantly bringing it up her desire for a dog and her reason for wanting a dog, his attention will be constantly drawn to these reasons for getting a dog (the fact that his partner wants one, and furthermore her reason for wanting one). I think it is not too controversial to say that these are clearly poor reasons for getting a dog, and so we would hope that Bella's partner is not taken in by these reasons. Say that he isn't: he hears the reasons, and they register, but he does not in any want to get a dog. But is Bella's partner's will undermined in this case? Is he going to have difficulty deliberating clearly about whether or not they should get a dog together because his attention is repeatedly directed to some bad reasons for getting a dog? I think not. He may be annoyed or worn down by Bella's repeated requests, but her repeated requests do not amount to an inclination.[268] His deciding mind is being directed to consider these reasons by Bella, not by his instinctive mind, and so for this reason he will not see getting a dog as something *to-be-done.* In short, what this highlights is that tempting another is not just the equivalent of yelling one or several reasons repeatedly at a person such that their attention keeps being drawn back to that reason. When you tempt another successfully, you furthermore make it difficult for them to act against the tempting inclination. Thus, although tempting another is often

[268] To be clear, I think this is a case of manipulation, or if we prefer to call it something other than manipulation, perhaps bullying or badgering would do. Furthermore, what's going wrong in this case is not just the fact that the reasons are bad. Even if we change the case and make the reasons better and more numerous – the partner likes dogs, Bella likes dogs, they like to walk and dogs thrive on regular walking, they can afford a dog, there are many dogs in need of homes in their area – it would still be possible for Bella to inappropriately badger or manipulate her partner with these good reasons if she is constantly bringing them up.

wrong because it presents the person's will with bad reasons, presenting bad reasons, even persistently, is not unique to temptation.

5.3.2 In light of the fact that tempting someone can direct their attention towards good reasons (Josh and the marathon) and we can persistently direct someone's attention to bad reasons without thereby tempting them (Bella's desire for a dog), I think we need a further explanation of what's problematic about tempting another person. Chapters 1 & 2 focused on the idea that when we are tempted to do something, it is easier to act on our tempting inclination than to act against it. Tempting inclination has motivational pressure, in other words. Motivational pressure is the second way in which temptation can undermine the will. When we tempt someone successfully, they acquire an inclination to act as we've tempted them. Their inclination is a provisional decision, their instinctive mind's attempt to direct their action, and this means that they will see the object of their tempting inclination as *to-be-pursued*. But since they are tempted and not merely inclined, the presence of the tempting inclination will also create a motivational conflict, and it will be difficult for them act against their tempting desire. Thus tempting another also undermines the will insofar as it makes it difficult to act as one originally intended.

However, the conflict is important here since it marks the difference between influencing someone via inclination in general and specifically tempting them. On my view it is, all things considered, permissible to influence others via inclination.[269] However, when I specifically tempt you and don't just try to make you generally inclined, I create a conflict in your agency. Tempting another generates a conflict between intention and what they implicitly judge they

[269] This is especially true insofar as having certain inclinations is a sign of moral maturity. For instance, we try to influence children not just to share but furthermore to *want* to share. Having certain inclinations is also important in the context of some relationships. One spouse might reasonably be disappointed if the other does not want to spend time together and they might appropriately try to influence their spouse to want to connect.

ought not to do or a conflict between intention and what they already decided to do. Furthermore, it's not just that temptation introduces a conflict in the sense that we can be conflicted between two good courses of action. If I successfully tempt you, I make it difficult for you to act as you intended. It is as if I am purposefully standing in your way as you're trying to go about your day and run your errands, creating a nuisance for you that slows you down as you go about your life. Just like physical interference of this sort with others' actions is wrong unless justified by extenuating circumstances (perhaps your errand is to go and harm someone), interfering with someone else's intentions by trying to tempt them to do something else is wrong unless specially justified by extenuating circumstances.[270]

It is perhaps easiest to see these in cases where the action in question is clearly a matter of reasonable disagreement. Consider a case like the following:

> Jae and Griffin are friends. Griffin decides to spend his Friday evening working. He has a big project to accomplish, and although it's not strictly necessary for him to work on it Friday night, it's definitely the most convenient time for him to get the work done. Jae plans to go to the movies Friday night, and although he knows that Griffin intends to work and furthermore knows that Griffin regards Friday night as the best time for him to tackle this project, he thinks Griffin should come with him. He tries to tempt Griffin to come along, mentioning for instance the proximity of the movie theatre to Griffin's favorite restaurant, the amazing reviews the movie has been getting, and so on and so forth.

There are several things going on in this case, not all of them relevant to temptation, so let's set aside for the time those things. One, it's possible for Jae to conduct himself in a rather bullying

[270] I realize that this analogy is somewhat fraught insofar as one of the fundamental principles underlying this dissertation is the rejection of brute force views of desire. Thus, to be clear, I am not in making this analogy saying that desire literally is a brute force. Furthermore, I do not think our ability to resist inclination is contingent in the same way our ability to resist brute forces like winds and currents is contingent. However, it is also a key tenet of my view of inclination that inclination *is* forceful, i.e. that it has motivational or asymmetric pressure and it is easier to act on an inclination rather than not. Thus when it comes to understanding how this affects the moral status of tempting another person, it is easy to capture what's wrong with this by drawing a comparison between physically interfering with another person's action and tempting them.

way, repeatedly texting or calling Griffin to come along even when Griffin has already declined. This is a sort of manipulation via bullying or badgering, and although I think it is generally wrong to try to influence others in this way, it is not the same as tempting them. So let's assume that Jae is not excessively badgering Griffin. Two, it's also possible for Griffin in this situation to be unduly influenced by social pressure. We are all familiar with the person who constantly tries to accommodate the preferences of others, and so another way this case might go wrong is if Griffin immediately changes his plan upon hearing the preferences of Jae.[271]

My interest however is specifically on Jae's attempt to *tempt* Griffin. It may be that Jae tries to tempt Griffin and is unsuccessful. But for the sake of identifying what's wrong with even attempting to tempt another person, consider the case in which Jae is successful and Griffin is tempted: in this version of the case, Griffin's agency is now conflicted with respect to what to do. On the one hand his instinctive mind will represent going to the movies as *to-be-done* and he will be attracted to the idea of a great movie and his favorite restaurant. On the other hand there is his standing intention to work and furthermore his assessment that Friday night really is the best time to work on the project.

To be clear, I do not view the case as one in which it's obvious that Griffin must stay home and work on Friday night. Instead I mean for the case to be about an action over which we could reasonably disagree. Perhaps Griffin is right that Friday night is the perfect time to get this important project done. Perhaps Jae is right that Griffin is being a little dull and needs to get out

[271] In such cases, I think Schapiro's analysis is helpful: a person who is constantly accommodating other people's preferences is failing to take responsibility for their own decisions. This is not to deny that in many cases, being excessively accommodating towards others is a coping mechanism designed to protect the accommodating person from, for instance, abuse. However, such a person arguably *is* responsible for their decisions; they are making the decision (consciously or unconsciously) to accommodate others in order to protect themselves. The problematic case is one in which someone simply exports decision making to others for no particular reason beyond that they find it easier to let others decide for them.

and do something fun. But this is exactly my point: we don't want a case in which the badness or goodness of what we're tempted to do sneaks in and does the work. What goes awry in this case is that Jae tries to interfere in the intention of his friend, making it more difficult for his friend to follow through on that intention, and the bar for permissible interference with another person's intention is relatively high.[272]

Returning to the animating question of the paper, then, the reason it is generally worse to influence someone by tempting them rather than influence them by some other means is two-fold. One, tempting another always involves the attempt to undermine their will by making it difficult for them to act against temptation. This is generally wrong in the same way it's generally wrong to put obstacles in other people's path. Furthermore, to return to autonomy, we are now prepared to get a better sense of why tempting someone is very often disrespectful to their autonomy. Except in extenuating circumstances where their intended action is very bad or your relationship is such that a high degree of interference is permitted, it is wrong to tempt other people because in so doing you disrespect their discretion and right to decide for themselves when you purposefully make it more difficult for them to act as they intend.

Second, although less central because it's possible to tempt another person without presenting them with bad reasons, tempting another often undermines the will in a second way, by presenting it with bad reasons. Tempting inclination is not just a pure imperative, it is a reason-laden imperative and as such focuses our attention on the reason(s) we have to act as we are inclined. This very often means that when we are tempted our attention focuses on our

[272] It is probably more permissible to interfere in the decisions of our friends than the decisions of strangers, but even so, to my mind this case is one where Jae should not interfere.

tempting reasons and makes it easy to forget about the reasons we have for acting against the tempting inclination.

Finally, notice that on this account, influencing someone with inclination is not in-itself manipulative. I think it would be a straw man to say that some accounts of manipulation treat all influence via inclination as automatically manipulative insofar as they are attempts to influence someone outside of "reasoning with them," but it is nonetheless easy to treat inclination as an automatically a manipulative form of influence because we often assume that inclination is essentially irrational.[273] For example, a more rigorous version of this straw man account might hold that inclination is always suboptimal as a form of influence and although it may be permissible under some circumstances, it needs to be specially justified.[274] However, on my account, what's wrong with tempting another person is not that it involves inclination, a state that is essentially irrational and bypasses one's rational capacities and poses a threat to one's ability to self-govern autonomously. Rather, temptation is wrong if and when it undermines one's will. This can happen when temptation makes it difficult to deliberate clearly by drawing our attention to tempting reasons, or when temptation makes it difficult to act by sending the instinctive mind in a totally different direction than the deciding mind.

[273] As Allan Wood notes with respect to emotion, "*Reason* and *emotion* are not opposites: emotions – even irrational ones – always have some degree of rational content, and healthy emotions are indispensable vehicles of rationality. We rationalists have always known this; its tediously predictable denial is one of the sad errors of those who reject rationalism" ("Coercion, Manipulation, Exploitation," 37).

[274] It is not clear if this is Schapiro's view since again, she considers weakness of will and not manipulation, but it may be for instance that her account is committed to something like this claim.

Bibliography

Allison, Henry. *Kant's theory of freedom.* Cambridge: Cambridge University Press, 1990.

Andreou, Chrisoula. "Temptation, Resolutions, and Regret." *Inquiry* 57, no. 3 (2014): 275-292.

Anscombe, G.E. *Intention.* Cambridge, MA: Harvard University Press, 1957.

Aristotle. *Nichomachean Ethics.* Translated by C.D.C. Reeve. Indianapolis: Hackett Publishing Company, 2014.

Arpaly, Nomy, and Timothy Schroeder. *In Praise of Desire*. Oxford: Oxford University Press, 2014.

Arpaly, Nomy, and Timothy Schroeder. "Praise, Blame and the Whole Self." *Philosophical Studies* 93, no. 2 (1999): 161-188.

Arpaly, Nomy. "Moral Worth." *Journal of Philosophy* 99, no. 5 (2002): 223-245.

Augustine. *The City of God.* Translated by Henry Bettenson. New York: Penguin Books, 2003.

———. *Confessions*. Translated by Henry Chadwick. Oxford: Oxford University Press, 1991.

Barnhill, Anne. "What is Manipulation?" In *Manipulation: Theory and Practice.* Edited by Christian Coons and Michael Weber. Oxford: Oxford University Press, 2014. 51-72.

Barrett, Lisa Feldman. *How Emotions Are Made.* New York: Mariner Books, 2017.

———. *Seven and a half lessons about the brain.* New York: Houghton Mifflin Harcourt Publishing Company, 2020.

Bratman, Michael. "Toxin, Temptation, and the Stability of Intention." In *Faces of Intention*. Cambridge: Cambridge University Press, 1999. 58-90.

———. "Temptation Revisited." In *Structures of Agency*. Oxford: Oxford University Press, 2007. 257-282.

———. "Time, Rationality, and Self-Governance." *Philosophical Issues* 22 (2012): 73-88.

———. "Temptation and the Agent's Standpoint." *Inquiry* 57, no. 3 (2014): 293-310.

Brewer, Talbot. "The Character of Temptation." *Pacific Philosophical Quarterly* 83, no. 2 (2002): 103-130.

———. *The Retrieval of Ethics.* Oxford: Oxford University Press, 2009.

Brown, Peter. *The Body and Society.* New York: Columbia University Press, 1988.

Bryan, Gharad and Dean Karlan and Scott Nelson. "Commitment Devices." *Annual Review of Economics* 2, no. 1 (2010): 671-698.

Burnyeat, Myles. "Aristotle on learning to be good." In *Essays on Aristotle's Ethics*. Edited by Amélie Rorty. Berkeley: University of California Press, 1980.

Burroughs, Michael. "Navigating the Penumbra: Children and Moral Responsibility." *The Southern Journal of Philosophy* 58, no. 1 (2020): 77-101.

Buss, Sarah. “Valuing autonomy and respecting persons: Manipulation, seduction, and the basis of moral constraints.” *Ethics* 115, no. 2 (2005): 195-235.

“Dark Thoughts.” Reported by Lulu Miller and Alex Spiegel. *Invisibilia.* Podcast audio. January 8, 2015. https://www.npr.org/2015/01/09/375928124/dark-thoughts

Davidson, Donald. “How is Weakness of Will Possible?” In *Essays on Actions and Events.* Oxford: Clarendon Press, 2011. 21-42.

Deci, Edward L. and Richard M. Ryan. “Facilitating Optimal Motivation and Psychological Well-Being Across Life’s Domains.” *Canadian Psychology* 49, no. 1 (2008): 14-23.

De Ridder, Denise T.D., Gerty Lensvelt-Mulders, Catrin Finkenauer, F. Marijn Stok, and Roy F. Baumeister. “Taking Stock of Self-Control: A Meta-Analysis of How Trait Self-Control Relates to a Wide Range of Behaviors.” *Personality and Social Psychology Review* 16, no. 1 (2012): 76-99.

Ebels-Duggan, Kyla. “Bad Debt: The Kantian Inheritance of Humean Desire.” Forthcoming in *The Idea of Freedom: New Essays on the Kantian Theory of Freedom.* Edited by Dai Heide and Evan Tiffany.

———. “Beyond Words: Inarticulable Reasons and Reasonable Commitments.” *Philosophy and Phenomenological Research* 98, no. 3 (2019): 623-641.

———. “Dealing with the past: responsibility and personal history.” *Philosophical Studies* 164, no. 1 (2013): 141-161.

———. “Love and Agency.” In *Routledge Handbook of Love in Philosophy.* Edited by Adrienne Martin. Abingdon: Routledge, 2018.

Frankfurt, Harry. “Identification and Wholeheartedness.” In *The Importance of What We Care About.* Cambridge: Cambridge University Press, 1998. 159-176.

Frede, Dorothea. “Pleasure and Pain in Aristotle’s Ethics.” In *The Blackwell Guide to Aristotle’s Nichomachean Ethics.* Edited by Richard Kraut. Oxford: Blackwell Publishing, 2006. 255-275.

Gillebaart, Marleen and Denise T. D. de Ridder. “Effortless Self-Control: A Novel Perspective on Response Conflict Strategies in Trait Self-Control.” *Social and Personality Psychology Compass* 9, no. 2 (2015): 88-99.

Gorin, Moti. “Towards a Theory of Interpersonal Manipulation.” In *Manipulation: Theory and Practice.* Edited by Christian Coons and Michael Weber. Oxford: Oxford University Press, 2014. 73-97.

Harte, Verity. “The *Nichomachean Ethics* on Pleasure.” In *The Cambridge Companion to Aristotle’s Ethics.* Edited by Ronald Polansky. Cambridge: Cambridge University Press, 2014. 288-318.

Hearne, Vicki. *Adam’s Task.* New York: Skyhorse Publishing, 2016.

Herman, Barbara. “Making Room for Character.” In *Aristotle, Kant and the Stoics: Rethinking Happiness and Virtue.* Edited by Stephen Engstrom and Jennifer Whiting. Cambridge: Cambridge University Press, 1996. 36-60.

Hieronymi, Pamela. “Controlling Attitudes.” *Pacific Philosophical Quarterly* 87, no.1 (2006): 45-74.

———. “Reasoning First.” In *The Routledge Handbook of Practical Reason*. Edited by Ruth Chang and Kurt Sylvan. New York: Routledge, 2020. 349-365.

———. “Responsibility for believing.” *Synthese* 161 (2008): 357-373.

———. “Two Kinds of Agency.” In *Mental Action*. Edited by L. O'Brien and M. Soteriou. Oxford: Oxford University Press, 2009. 138-162.

———. “The Will as Reason.” *Philosophical Perspectives* 23, no. 1 (2009): 201-220.

Hoffman, Paul. “Freedom and Strength of Will: Descartes and Albritton.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 77, no. 2/3 (1995): 241-260.

Holton, Richard. *Willing, Wanting, Waiting*. Oxford: Oxford University Press, 2009.

Hursthouse, Rosalind. *On Virtue Ethics.* Oxford: Oxford University Press, 1999.

Kahneman, Daniel. *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux, 2011.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals.* Translated by Mary Gregor. Cambridge: Cambridge University Press, 1998.

Kant, Immanuel. *Religion Within the Bounds of Mere Reason.* Translated by Allen Wood and George Di Giovanni. Cambridge: Cambridge University Press, 1998.

Kavka, Gregory. “The Toxin Puzzle.” *Analysis* 43, no. 1 (1983): 33-36.

Korsgaard, Christine. *The Constitution of Agency.* Oxford: Oxford University Press, 2008.

———. *Fellow Creatures.* Oxford: Oxford University Press, 2018.

———.“Morality and the Distinctiveness of Human Action.” In *Primates and Philosophers: How Morality Evolved.* Edited by Stephen Macedo and Josiah Ober. Princeton: Princeton University Press, 2006.

———. “Personal Identity and the Unity of Agency: A Kantian Response to Parfit.” In *Creating the Kingdom of Ends.* Cambridge: Cambridge University Press, 1996. 363-397.

———. *Self-Constitution.* Oxford: Oxford University Press, 2009.

———. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

Liberman, Alida. “Reconsidering Resolutions.” *Journal of Ethics & Social Philosophy* 10 (2016): 1-26.

MacIntyre, Alasdair. *After Virtue.* Notre Dame: University of Notre Dame Press, 2007.

———. *Dependent Rational Animals.* Chicago: Carus Publishing Company, 1999.

Marušić, Berislav. *Evidence and Agency.* Oxford: Oxford University Press, 2015.

McDowell, John. “Virtue and Reason.” *The Monist* 62, no. 3 (1979): 331-350.

Mills, Claudia. “Politics and Manipulation.” *Social Theory and Practice* 21, no. 1 (1995): 97-112.

Milyavskaya, Marina, Michael Inzlicht, Nora Hope and Richard Koestner. “Saying ‘No’ to Temptation: *Want-to* Motivation Improves Self-Regulation by Reducing Temptation

Rather Than by Increasing Self-Control." *Journal of Personality and Social Psychology* 109, no. 4 (2015): 677-693

Moss, Jessica. *Aristotle on the Apparent Good: Perception,* Phantasia*, Thought, and Desire.* Oxford: Oxford University Press, 2012.

———. "Was Aristotle A Humean?" In *The Cambridge Companion to Aristotle's Ethics.* Edited by Ronald Polansky. Cambridge: Cambridge University Press, 2014. 221-241.

Murdoch, Iris. *The Sovereignty of Good.* New York: Routledge Classics, 1970.

Nagel, Thomas. *The Possibility of Altruism.* Oxford: Oxford University Press, 1971.

Noggle, Robert. "The Ethics of Manipulation." *Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Last revised March 22, 2020.

Oddie, Graham. *Value, Reality and Desire*. Oxford: Oxford University Press, 2005.

Odell, Jenny. *How to Do Nothing: Resisting the Attention Economy.* New York: Melville House Publishing, 2019.

O'Neill, Onora. "Between Consenting Adults." *Philosophy & Public Affairs* 14, no. 3 (1985): 252-277.

Paul, Sarah K. "Review of *Willing, Wanting, Waiting*." *Mind* 120, no. 479 (2011): 889-92.

———. "Diachronic Incontinence is a Problem in Moral Philosophy." *Inquiry* 57, no. 3 (2014): 335-355.

Pickard, Hanna. "Psychopathology and the Ability to Do Otherwise." *Philosophy and Phenomenological Research* 90, no. 1 (2015): 135-163.

Plato. *Republic.* In *Plato: Complete Works*. Translated by G.M.A. Grube, translation revised by C.D.C. Reeve. Edited by J.M. Cooper. Indianapolis: Hackett Publishing Company, 1997.

Quinn, Warren. "Putting Rationality in its place." In *Morality and Action*. Cambridge: Cambridge University Press, 1994. 228-272.

Scanlon, T.M. "Reasons and Passions*."* In *The Contours of Agency.* Edited by Sarah Buss and Lee Overton. Cambridge, MA: MIT Press, 2002.

Scanlon, T.M. *What We Owe to Each Other.* Cambridge, MA*:* Harvard University Press, 1998.

Schafer, Karl. "Perception and the Rational Force of Desire." *The Journal of Philosophy* 110, no. 5 (2013): 258-281.

Schapiro, Tamar. "Childhood and Personhood." *Arizona Law Review* 45, no. 3 (2003): 575-594.

———. *Feeling Like It*. Oxford: Oxford University Press, 2021.

———. "The Nature of Inclination." *Ethics* 119, no. 2 (2009), 229-256.

———. "On the Relation Between Wanting and Willing." *Philosophical Issues* 22, no. 1 (2012): 334-350.

———. "What are Theories of Desire Theories of?" *Analytic Philosophy* 55, no. 2 (2014): 131-150.

———. "What Makes Weak-Willed Action Weak?" Forthcoming in *Normativity and Agency: Themes from the Philosophy of Christine M. Korsgaard.* Edited by Kyla Ebels-Duggan, Tamar Schapiro, and Sharon Street.

Shoemaker, David. “Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility.” *Ethics* 131, no. 3 (2011): 602-632.

———. *Responsibility from the Margins.* Oxford: Oxford University Press, 2015.

Smart, J.J.C. “Free-will, praise, and blame.” *Mind* 70, no. 279 (1961): 291-306

Smith, Angela. “Attitudes, Tracing, and Control.” *Journal of Applied Philosophy* 32, no. 2 (2015): 115-132.

———. “Conflicting Attitudes, Moral Agency, and Conceptions of the Self.” *Philosophical Topics* 32, no. 1/2, (2004): 331-352.

———. “Control, responsibility, and moral assessment.” *Philosophical Studies* 138, no. 3 (2008): 367-392.

———. “Responsibility as Answerability.” *Inquiry* 58, no. 2 (2015): 99-126.

———. “Responsibility for Attitudes: Activity and Passivity in Mental Life.” *Ethics* 115, no. 2 (2005): 236-271.

Stampe, Dennis. “The Authority of Desire.” Philosophical Review 96, no. 2 (1987): 335-82.

Stohr, Karen. “Moral Cacophony: When Continence is a Virtue.” *The Journal of Ethics* 7, no. 3 (2003): 339-363.

Strawson, P.F. “Freedom and Resentment.” In *Freedom and Resentment and other essays.* London: Methuen & Co. Publishing, 1974. 1-20.

Tangney, June P., Roy F. Baumeister, and Angie L. Boone. “High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success.” *Journal of Personality* 72, no. 2 (2004), 271-322

Taylor, Charles. “Responsibility for Self.” In *The Identities of Persons*. Edited by Amélie Rorty. Orlando: University of California Press, 1976. 281-299.

———. *Sources of the Self: The Making of the Modern Identity.* Cambridge, MA: Harvard University Press, 1989.

Trianosky, Gregory. “Rightly Ordered Appetites.” *American Philosophical Quarterly* 25, no. 1 (1988): 1-12.

van der Kolk, Bessel. *The Body Keeps the Score.* New York: Penguin Random House, 2014.

Wallace, R. Jay. “Addiction as Defect of the Will: Some Philosophical Reflections.” *Law and Philosophy* 18, no. 6 (1999): 621-654.

———. “Three Conceptions of Rational Agency.” In *Normativity and the Will: Selected Papers on Moral Psychology and Practical Reason.* Oxford: Oxford University Press, 2006. 43-62.

Watson, Gary. “Two Faces of Responsibility.” In *Agency and Answerability: Selected Essays.* Oxford: Oxford University Press, 2004.

Weil, Kari. *Thinking Animals: Why Animal Studies Now?* New York: Columbia University Press, 2012.

Werner, Kaitlyn M. and Marina Milyavsksaya. “Motivation and self-regulation: The role of want-to motivation in the processes underlying self-regulation and self-control.” *Social and Personality Psychology Compass* 13 (2019): 1-14.

White, Stephen. “On the Moral Objection to Coercion.” *Philosophy and Public Affairs* 45, no. 3 (2017), 199-231.

Wolf, Susan. *Freedom Within Reason.* Oxford: Oxford University Press, 1990.

Wood, Allan. “Coercion, Manipulation, Exploitation.” In *Manipulation: Theory and Practice*. Edited by Christian Coons and Michael Weber. Oxford: Oxford University Press, 2014.