

THE NORMATIVE POWER OF RESOLUTIONS  
Angela Sun

---

ABSTRACT

This paper argues that resolutions are reason-giving: when an agent resolves to  $\phi$ , she incurs a normative reason to  $\phi$  over and above the reasons that led her to resolve to  $\phi$  in the first place. On my view, resolution-making—like promising, forgiveness and consent—is a *normative power*: with it, agents have a remarkable ability to alter their normative circumstances through sheer acts of will. I compare the reasons we incur from forming resolutions to the reasons we incur from making promises. One upshot of my view is that it offers a ready response to the bootstrapping problem for mental attitudes, on which if mental attitudes gave us reasons, we could bootstrap any action into rationality simply by acquiring the relevant mental attitude.

1 RESPECTING RESOLUTIONS

In *Being and Nothingness*, Jean-Paul Sartre (1956) introduces us to a gambler who, fearful of financial ruin and the disappointment his losses will cause his family, resolves to quit gambling. But the next day, as he approaches the game table, he "suddenly sees all his resolutions melt away" (p. 69). He realizes, in other words, that his resolution to stop gambling is totally inefficacious; it has no binding effect on him. He thinks:

[Y]esterday I even had a synthetic appreciation of the situation (threatening ruin, disappointment of my relatives) as *forbidding* me to play. It seemed to me that I had established a *real barrier* between gambling and myself, and now I suddenly perceive that my former understanding of the situation is no more than a memory of an idea, a memory of a feeling (Sartre, 1956, p. 70, emphasis in original).

So, the gambler takes a seat at the table and asks to be dealt in.

There is a strong intuition that Sartre's gambler has acted wrongly—but why?<sup>1</sup> One might say that the gambler has acted against his better judgment, that

---

For helpful discussion and comments on earlier drafts, I am grateful to Abdul Ansari, Chrisoula Andreou, Sarah Buss, Scott Hershovitz, Jim Joyce, Ishani Maitra, Peter Railton, Connie Rosati, and audiences at Central European University and the University of Michigan.

<sup>1</sup> I use "wrongly" not in the moral sense, but simply to indicate that there is some way that the gambler's behavior has gone awry, or is non-optimal. (However, someone with Kantian intuitions and strong views about what we owe to ourselves might believe that the gambler *has* acted

this is a classic case of *akrasia*. But that diagnosis wouldn't be quite right. In this case, the gambler doesn't act against his better judgment, since he unambiguously changes his mind about what he has all-things-considered reason to do. The right thing to say in this case seems to be that the gambler's change of mind, though "genuine," is *unjustified* or *illegitimate* in some way. There is therefore still a sense in which he acts irrationally.

Not all of us are trying to overcome gambling addictions. But we all find ourselves in situations that are like the gambler's in relevant ways. The projects we undertake in our lives are spread out in time and require planning for us to complete. Think of writing dissertations, fighting injustice, going for hikes, and cultivating relationships with family and friends. If every urge or inclination caused us to change our plans, we would be unable to make meaningful commitments and pursue the kinds of projects that enrich our lives. Every time in the past year I resolved to work on my research but changed my mind when it came time to sit down and do it (there were, I am sad to admit, many such occasions), I acted wrongly in a smaller but similar way to the gambler. I want my planning agency to be effective; I suspect many of us do. So we should have an interest in figuring out where we go wrong when we behave like the gambler.

Some philosophers have tried to explain what agents like the gambler do wrong by appealing to the benefits of planning agency that they forgo in failing to respect their resolutions. Michael Bratman (1983, 1987, 2000, 2007) has argued extensively that without the ability to follow through on our resolutions, we would be unable to pursue projects that are spread out in time. Moreover, because we are creatures with limited cognitive resources, we cannot always have the time, information, or clarity of mind to deliberate at the moment of action. It can be extremely useful for us to divide deliberative labor intertemporally, delegating decisions about what to do to our past selves who are better situated to make those decisions than our present selves are (Ferrero, 2010).

While I think it's true that agents who fail to respect their resolutions forgo the valuable benefits of planning agency, I don't think that this provides a satisfying explanation for the intuition that Sartre's gambler acts wrongly *right then*, when he feels his resolve melt away. There are certainly advantages to being the

---

wrongly in the moral sense, for if we have obligations to cultivate capacities that are distinctive of our human nature—including, perhaps, our *planning* capacities—we may wrong ourselves when we fail to respect resolutions that we made in our own best interest (Kant, 2017.)

kinds of creatures who follow through on our resolutions. But this doesn't explain why we are wrong to disregard our resolutions in *any particular instance*.

In this paper, I explain what's wrong with failing to respect resolutions. I argue that agents who unjustifiably act against their resolutions fail to be responsive to reasons in the way that well-functioning agents ought to be. On my view, resolutions are *reason-giving*: in resolving to  $\phi$ , one incurs a normative reason to  $\phi$  over and above the reasons that led one to resolve to  $\phi$  in the first place. In other words, the fact that one resolved to  $\phi$  ought to count as a consideration in favor of  $\phi$ ing at the time of action. An agent who unjustifiably acts against her resolutions therefore lacks sensitivity to her normative circumstances—specifically, to reasons she has to act in certain ways.

There is evidence in natural language and natural patterns of reasoning that resolutions are reason-giving. Suppose my partner asks, "Why are you doing yoga at six in the morning?" "Because last night I told myself I would" seems like a perfectly satisfactory answer. More often than not, the mere fact that one resolved to  $\phi$  is enough to explain one's  $\phi$ ing. Of course, my partner might ask why I made the resolution in the first place; I would tell him that I want to develop a habit of exercising first thing in the morning and get better at clearing my head. But, to be perfectly honest, at six in the morning, those reasons will probably be inaccessible to me. Like the gambler, "my former understanding of the situation [would be] no more than a memory of an idea, a memory of a feeling." What *will* be accessible to me is the memory of forming a resolution, and I have to trust that my past self made the right decision in forming it.

I suspect that one reason why philosophers have been hesitant to explore the possibility that resolutions are reason-giving despite evidence that they are is worries about *bootstrapping*: if resolutions gave us reasons, the worry goes, then we could bootstrap any action into rationality simply by resolving to do it. On the view I advance in this paper, the reasons we incur from forming resolutions are similar in important ways to the reasons we incur from making *promises*. A comparison between resolution-based reasons and promissory reasons will reveal that resolutions, like promises, may bootstrap, but not problematically. If the comparison is apt, the account of resolutions I offer in this paper has a ready response to a major barrier preventing philosophers from accepting that resolutions are reason-giving.

The paper proceeds as follows. In §2, I defend the value of our power to

form resolutions, explaining why we need it to shield our self-governance and interest in living authentically against temptation. In §3, I develop my thesis that resolutions are reason-giving, developing a comparison between resolution-based reasons and promissory reasons. On my view, resolution-making turns out to be a *normative power* alongside phenomena like promising, consent, and forgiveness. With the ability to make resolutions, we have a remarkable ability to alter our normative circumstances through sheer acts of will. In §4, I use the comparison between resolutions and promises to show that my account of resolutions does not lead to problematic bootstrapping.

## 2 THE THREAT OF TEMPTATION

Resolutions are often thought to be a special kind of intention formed to stand firm in the face of future contrary inclinations (Holton, 2009, p. 10). Sartre's gambler does not merely intend but *resolves* to quit gambling because he anticipates that he will be tempted to return to the game table. When we make New Year's resolutions—to call our parents more, save money, or pick up a new hobby—we aspire to do something we wish to do, but where barriers threaten our success. (The barriers in question might be substantial, but sometimes they are trivial. The barrier preventing us from picking up a new hobby might be laziness or the allure of watching another episode of TV.) Think, too, of showing resolve in the face of an obstacle. When one resolves to jump off the diving board, one decides to stop deliberating and *just do it*. We show resolve when we commit to a course of action *and* to refraining from further deliberation about the matter.

Intentions more generally are known to resist reconsideration and revision; they are characterized by what Bratman (1987) calls *stability*. Unlike resolutions, however, intentions are not always formed specifically with the aim of resisting temptation that threaten to cause us to deviate from our plans. This is why it seems silly to call a decision to, say, make pizza rather than pasta for dinner a resolution, rather than simply an intention. In cases like this, there is no temptation that threatens to throw us off course; we simply need to choose between pizza and pasta and for the intention we form to be stable so we have something to eat for dinner.

Another difference between resolutions and mere intentions has to do with the processes by which we form them. As Richard Holton (2009, p. 53) points out,

many intentions are formed automatically. When you see slow traffic ahead you form an intention to switch lanes and scan the mirror for an opportunity to do so without giving any conscious thought to the matter. We do not only form intentions automatically in mundane cases. One feature of expertise is the ability to form intentions automatically in familiar scenarios; a professional hockey player might recognize a play her teammate is starting and get in formation without explicitly deciding to do anything. Resolutions, by contrast, must be formed explicitly; one cannot resolve to do something without giving it conscious thought.

Why do we need a special kind of intention to shield our plans from the forces of temptation? In this section, I argue that, as agents with temporally extended projects, we need a capacity like the one to form resolutions in order to make self-governance and authenticity possible. Without something like the capacity to form resolutions, we would not be able to carry out plans to pursue projects that matter most to us and make us who we are when temptation threatens to throw us off course.

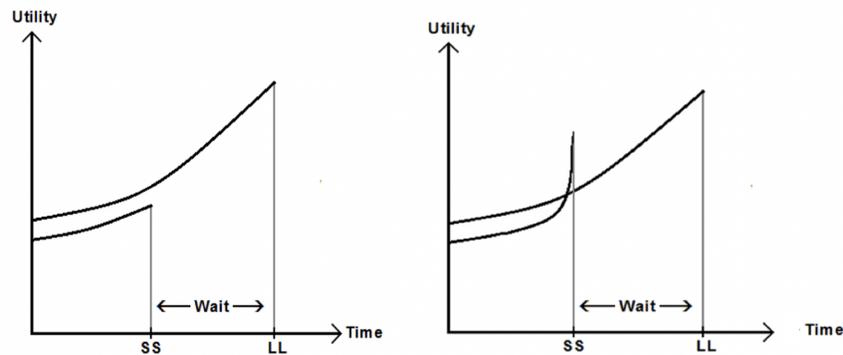
\*\*\*

Temptation is widely regarded to cause our desires—and in some cases our judgments about what we have most reason to do, all things considered—to shift. Suppose that you set some time aside this evening to watch TV. Because you would like to get some work done tonight, you resolve to limit yourself to one episode. But, cozy on the couch, you are tempted to watch a second, and when the streaming service automatically starts playing the next episode, you succumb to temptation and decide that in light of your heightened desire, you now have most reason to watch more TV rather than work.<sup>2</sup> As discussed in the previous section, the judgment shift induced by temptation makes it difficult to explain why your behavior is irrational. If temptation heightens your desire to watch TV and causes your all-things-considered judgment about what to do to shift, then it

<sup>2</sup> I am introducing a new example of temptation, rather than continuing to use Sartre's gambling addict, because addictive temptation may be different in important respects from standard cases of temptation. According to Holton (2009, pp. 103-111), while standard temptation should be characterized in terms of judgment shifts, addictive temptation should be understood in terms of a disconnect between an agent's desires and what they judge best. I think that Sartre's analysis of the gambler—especially his description of the gambler's thought process when he abandons his resolution—is illuminating, and I wanted to include it as a central case in the paper. But I also want to respect the differences between addiction and other forms of temptation.

appears as though it is rational for you to continue watching TV, even though it goes against your resolution not to.

The same problem arises for other forms of preference shifts, including temporal discounting and our ability to delay gratification. It is normal for one's desire for a good to increase as one approaches the time of consumption; strictly speaking, there's nothing wrong with preferring to have, say, a chocolate bar *now* to having a chocolate bar in four hours. But not all forms of temporal discounting are rationally acceptable. Consider the figures below, which contrast *exponential* and *hyperbolic* discounting:



Exponential vs. hyperbolic discounting

Suppose that you are deciding between receiving \$30 in one week and \$40 in one month; \$30 is therefore a "smaller but sooner" good (SS) and \$40 is a "larger but later" good (LL). If you are an exponential discounter, your preferences for SS and LL are represented by the figure on the left. Although your desires for the two goods grow as you approach the time of consumption, they grow at the same rate. Because LL never becomes more desirable to you than SS—because your *comparative ranking* of the two goods remains fixed over time—you will be patient and hold out for LL despite your increasing desire for SS. If you are a hyperbolic discounter, on the other hand, your preferences are represented by the figure on the right. In this case, your desire for SS *spikes* immediately before consumption, surpassing your desire for LL. Therefore, you will succumb to impatience and choose SS when it is made available to you, rather than hold out for LL.

Experiments conducted by Ainslie and Haslam (1992) suggest that most people are hyperbolic discounters, not exponential ones. The allure of having something *now* causes our preferences and our judgments about what to do to

shift, even though we would maximize utility by waiting. There is a strong intuition that hyperbolic discounting is irrational. However, if one prefers SS to LL and judges that one has most reason to pursue SS rather than LL *at the time of choice*, it is difficult to say why that is the case.

To explain why it's irrational to give into temptation and impatience despite the judgment shifts it causes, we can begin by appealing to a distinction drawn by Bratman (2007) between "policies" and "singular valuing." A policy is a rule that one imposes on one's practical reasoning; you might, for instance, have a policy of limiting yourself to one episode of TV on evenings you also have work to get done, and you ought to respect this policy even if, faced with the vivid and immediate prospect of a second episode, you temporarily come to value two episodes "just this one time." According to Bratman, we ought to prioritize our policies over our singular valuing in our practical reasoning because of their "agential authority" (2007, pp. 265-268). An agent's policies constitute a point of view that is, in an important sense, the *agent's* point of view: they establish her practical framework and have a privileged status as being peculiarly the agent's "own." Bratman argues that, crucially, agential authority enables our cross-temporal organization. As agents who persist over time, we need to prioritize the policies that structure our ongoing practical reasoning and action above singular valuing that structure only our *present* reasoning and action.<sup>3</sup> This isn't to say that you cannot have a policy of watching only one episode of TV on evenings you need to work *except on certain special occasions*, on days that really are different from other days. You would, of course, need to be careful to make sure that today *really is* different from other days and that you are not acting on the basis of a singular valuing disguised as a policy.

The idea that some perspectives or attitudes better represent "who we really are" than others has philosophical precedent. Consider Frankfurt's (1971) idea that our second-order volitions are more "our own" than unendorsed first-order desires. Frankfurt introduces an unwilling addict who hates his addiction but struggles to overcome it. The unwilling addict has conflicting first-order desires: he wants to take the drug and wants to refrain from taking it. But he is not neutral in that conflict: he wants the latter desire to be his *will*. Because the unwilling

<sup>3</sup> Bratman (2007) also provides a rich analysis of "valuing" and the role it plays in practical reasoning and motivation. For the purposes of this paper, it's not so important to get into those details. All that's needed is a distinction between general policies and fleeting valuing.

addict takes the drug in spite of his higher-order desire—because he cannot will what he wants—his will, on Frankfurt's account, is not free. In Bratmanian terms, the addict's desire to refrain from taking the drug has *agential authority* in virtue of its higher-order endorsement. This helps explain why the addict is in an important sense "not himself" when he takes the drug.

Another literature that makes use of the idea that some perspectives or attitudes are more our own than others is that on moral responsibility. On the deep self view of moral responsibility, we are responsible only for actions that emanate from our "deep selves": the attitudes in our psychology that are fundamental to who we are, and therefore have a privileged status in our metaphysical and moral lives (e.g., Arpaly & Schroeder, 1999; Sripada, 2016). On this view, one is not morally responsible for actions that emanate from one's "shallow self"—like the unwilling addict's drug use or a kleptomaniac's compulsive shoplifting—because those actions are not, in an important sense, *ours*, but the result of external forces acting upon us. As with all philosophical views worth taking seriously, these positions are controversial. But I've provided a sketch of them here to show that there is nothing new about the idea that the perspectives and attitudes we identify with are privileged, practically speaking, even if we do not always act on or in accordance with them. This is the threat of temptation. When temptation causes a spike in desires we do not identify with, we act in ways that are *inauthentic*, that are not representative of who we really are.

One question that arises at this point is why we should care about being authentic, about acting on or in accordance with the perspectives and attitudes that we identify most strongly with. What's wrong with giving into temptation? Although you may resent your desire to watch a second episode of TV, you do get *something* good out of giving into your temptation: the pleasure of a second episode you so desire to watch. What reason do you have to prioritize your "deep" desire to get to work over your "shallow" desire to watch another episode of TV?

The answer to this question lies, I think, in our interest in being self-governing and living authentically. We *want* to run our own lives, be in charge of what we do, and act in ways that represent who we really are. Some philosophers have even argued that self-governance is valuable regardless of whether we desire it, that it is a necessary component of a good life.<sup>4</sup> Because the desires we act on

<sup>4</sup> Philosophers who have pursued this line of thought argue that there is something deeply problematic about ceding our self-governance to others, even a panel of experts who would make the

in the grips of temptation are in an important sense not our own, we fail to be self-governing and authentic when we act on them. Temptation is therefore a threat to self-governance and authenticity. If we want to be in charge of our own lives and live in a way that is true to who we are, we need a way to resist temptation when we confront it so that we can execute our plans to achieve the ends that are most important to us.

How can we resist temptation? According to Sartre (1956), in order for the gambler to resist the temptation to gamble, he must "recreate" his thoughts of financial ruin and of disappointing his family as "experienced fear" (p. 69). This suggests that knowing the reasons that lead us to form a resolution isn't enough for the resolution to be effective; we must also feel the *force* of those reasons. But that can't be right. Although feeling the force of the reasons that led us to form a resolution might make it easier to be *motivated* to follow through on the resolution, there are plenty of ways we see our resolutions through even when the reasons that led us to form them are inaccessible to us. Sometimes we introduce external constraints or motivations to help us stick to our desired course of action. (A professor of mine once told me that she would buy herself a pair of earrings and leave them wrapped on the mantle, not to be opened until she finished the paper she was working on. We can also think here of Odysseus, who respected his resolution to resist the bewitching song of the Sirens only by tying himself to the mast of his ship and having his crew plug their ears with beeswax.) Sometimes we find ways to divert our attention away from temptation (as someone seeking to curb a gambling addiction might avoid the game table altogether).<sup>5</sup>

As effective as these methods for resisting temptation are, we have another that works *directly* to fight off contrary inclinations: the ability to form resolutions. Resolutions aren't always enough to get us to stick to a course of action, but as Holton (2009, p. 10) points out, very often they are. People frequently wake up on cold, dark mornings to go for runs and kick bad habits using no other mechanism than a resolution to overcome their contrary desires. Resolutions are important; they are a power we have to execute our plans in the face of temptation.

---

same (or *better*) decisions as us. For these authors, there is intrinsic value in making decisions oneself. See Crisp (1997); Griffin (1986); Wall (1998).

<sup>5</sup> George Ainslie (2001) elaborates on these methods and others for overcoming temptation.

## 3 RESOLUTIONS AND PROMISSORY REASONS

Having defended the value of resolutions, I now present a view of what resolutions are and how they work. On my view, resolutions are reason-giving: when one resolves to  $\phi$ , one incurs an additional normative reason to  $\phi$ . The reasons we incur from forming resolutions are similar in important ways to the reasons we incur from making promises. In this section, I compare resolutions to promises (in particular, *self*-promises) and identify crucial similarities between them.

\*\*\*

A person promises to  $\phi$  when she communicates an intention to be under an obligation to  $\phi$ . Although there is disagreement over the source of the normativity of promises—some think that promises are binding because they generate an expectation that the promisor will perform the promised act (Scanlon, 1990); others attribute the normativity of promises to the interests we have in forming intimate relations with others (Shiffrin, 2008) or in having practical authority over others in some instances (Owens, 2006)—there is agreement that promises *bind*. When you break a promise to someone, you wrong them.

I think that resolutions are similar to self-promises in important ways. Compare the statement "I resolve to watch less TV" to the statement "I promise myself I'll watch less TV." The invocation of a promise in the latter statement may make it appear as though there is something *moral* rather than merely prudential on the line. But I hope to show in what follows that these statements are, for practical intents and purposes, equivalent.

Historically, philosophers have been dubious of promises we make to ourselves. In particular, they have doubted that self-promises have the normativity—or binding force—that interpersonal promises obviously have. In an interpersonal promise, a promisor (the party bound by the promise) can only be released from her obligation to fulfill the promise by the promisee (the party to whom the promisor is bound). However, because in a self-promise the promisor and promisee are one and the same, it appears as though a promisor can release herself from an obligation to fulfill a promise *at will*. Therefore, the argument goes, the promisor was never bound by the promise at all.<sup>6</sup> Call this the *objection from self-release*.

<sup>6</sup> Thomas Hobbes (1894) articulates this objection in *Leviathan*, when he discusses why it is impossible for a sovereign to be bound by laws when the laws are up to him: "Having power to make

Recently, the philosophical consensus has shifted; there is now widespread agreement that self-promises and duties to oneself more generally are possible despite the possibility of self-release (e.g., Dannenberg, 2015; Fruh, 2014; Muñoz, 2020; Oakley, 2017; Schaab, 2021; Schofield, 2021). Connie Rosati (2011) offers a representative response to the objection from self-release. Rosati argues that self-promises are possible because we can meaningfully distinguish between *releasing* oneself from a self-promise and *breaching* a self-promise. A promisee releases a promisor from her obligation to fulfill a promise when the promisee recognizes herself as having communicated a genuine change of mind to the promisor about the promise. (The reasons for the promisee's change of mind need not be *good* reasons; we release people from promises all the time for reasons that aren't particularly compelling, and that we even come to regret.) According to Rosati, there's no reason to think that promissory release operates any differently when it comes to self-promises. There is an important difference between changing your mind about a promise you made to yourself and acting against that promise. When you fail to fulfill a promise to yourself, you might recognize that you are compromising your values, letting yourself down, or acting self-destructively, even though it was in your power to release yourself from the promise.

\*\*\*

There are striking similarities between resolutions and promises. Consider first the similarities between the *functions* of promises and of resolutions. Promises play an important role in our social lives; they allow us to form *expectations* about others' future conduct and, relatedly, provide us with *assurance* that others will act a certain way. If I promise I'll pick you up from the airport, then you can reasonably expect that I'll be there when your flight lands and you can plan the rest of your day accordingly.<sup>7</sup>

---

and repeale Lawes, he may when he pleaseth, free himself from that subjection, by repealing those Lawes that trouble him, and making of new; and consequently he was free before. For he is free, that can be free when he will: Nor is it possible for any person to be bound to himself; because he that can bind, can release; and therefore he that is bound to himself onely, is not bound" (p. 124).

<sup>7</sup> As I explain later in this section, some philosophers have argued that the expectations and assurances generated by promises aren't merely useful functions of promises; they ground the *normativity* of promises. But even if you disagree with the claim that expectations and assurances ground the normativity of promises, it's hard to deny that promises play the role of facilitating interpersonal coordination.

Analogously, an agent forms a resolution when she wants assurance that she will follow a course of action, and her resolution generates an expectation about her future conduct. Suppose, knowing that you have writing to do tonight, you are trying to decide whether to watch any TV at all because you suspect that you'll be tempted to watch more than one episode. In resolving to restrict yourself to one episode, you assure yourself that you will watch only one episode, and that your decision to watch TV at all is sensible. Moreover, resolving to watch only one episode generates an expectation that you will watch only one episode. Having made the resolution, you can reasonably plan ahead for nighttime writing. If you hadn't resolved to limit your TV consumption, it might seem foolish or overly optimistic to expect yourself to get any writing done in the evening. (It might, of course, be foolish and overly optimistic for you to expect yourself to get any writing done *even if* you resolve to limit yourself to one episode of TV. If you have a track record of failing to follow through on your resolutions, you might not have reason to trust yourself to follow through on this one. But the same thing is true of promises. If I promise I'll pick you up from the airport, but I have a bad track record of keeping my promises, it might be foolish and overly optimistic of you to form any expectation that I will actually pick you up when your flight lands.)

Because they create expectations and assurances about how others will act, promises facilitate *interpersonal coordination* and allow us to reap the benefits of social cooperation. Promises (and contracts, equivalents to promises in the legal world) give us a basis for trusting others. Analogously, the expectations and assurances generated by resolutions facilitate *intrapersonal coordination* and allow us to enjoy the benefits of planning agency.

One way to illustrate this similarity between promises and resolutions is to consider them as solutions to prisoner's dilemmas. In the standard prisoners' dilemma (illustrated in the payoff matrix below), two players have the option of either defecting against or cooperating with their opponent. Each player does better no matter what their opponent does by defecting; defecting is therefore a dominant strategy. But if both players cooperate, they each do better than they would have if they both defected; cooperation therefore yields a Pareto optimal outcome.

	Cooperate	Defect
Cooperate	3, 3	0, 5
Defect	5, 0	1, 1

### The prisoner's dilemma

The challenge of the prisoner's dilemma is figuring out how to get players to cooperate and achieve the Pareto optimal outcome despite the fact that defecting is a dominant strategy. A potential solution to this challenge is for the players to *promise* to cooperate, to assure each other that they intend to band together in order to secure the best overall outcome. There is, of course, a question of how such a promise could be credible, especially in a one-off game. If the players don't know each other before the game and won't play against each other again in the future, then they appear to have little reason to follow through on any promise to cooperate that they make. However, most prisoner's dilemmas in real life are iterated versions of the game. Because we can expect to encounter our opponents again and again, we have incentive to follow through on our promises to cooperate so our opponents will continue to trust us in the future.<sup>8</sup> But even in one-off prisoner's dilemmas, there is (I hope) *moral* incentive to follow through on any promises we make to cooperate. In making a promise to cooperate, an agent who cares about morality *changes the payoffs in the game*. Since defecting would involve breaking a promise, the cost of defecting increases.

In addition to prisoner's dilemmas between two parties, we can recognize *intrapersonal* prisoner's dilemmas, where the two players are versions of oneself at different points in time and with different motivational states. George Ainslie (2013) has suggested that recovering from addiction can be understood as an intertemporal version of an iterated prisoner's dilemma. The present-addict faces a choice between relapsing or abstaining. If he relapses now, he fulfills his present desire for a high, but imposes a cost on his future self, who must start from scratch tomorrow. If he abstains now, but relapses tomorrow, then he might as well have chosen to relapse now, so his present self incurs a cost. Relapsing therefore appears to be a dominant strategy. However, if the present- and future-addict both abstain, they attain a Pareto optimal outcome: recovery. Moreover, because

<sup>8</sup> Robert Axelrod (1980) has shown that winning strategies in iterated prisoner's dilemma tournaments are *trustworthy*: they earn the trust of their opponents, and do not betray that trust by defecting for their own advantage.

the addict must engage *repeatedly* in this intertemporal bargaining, every current defection diminishes her credibility and jeopardizes cooperation in the future.

Just as promises provide a potential solution to the interpersonal prisoner's dilemma, *resolutions* provide a potential solution to the intrapersonal prisoner's dilemma. When we respect our resolutions, we build up credibility that we will respect our resolutions in the future, giving our future selves reason to engage in intertemporal bargaining with our present selves. Just as there are benefits to cooperating with others, there are benefits cooperating with *oneself*, by respecting one's resolutions. When one cooperates with oneself, one makes it easier to attain outcomes that are better for oneself overall, even if they might not appear to be best for oneself at the moment.

I have argued that promises play the role in our interpersonal lives that resolutions play in our intrapersonal lives. I now discuss a second striking similarity between promises and resolutions: the same considerations that can be used to ground the normativity of promises can be used to ground the normativity of resolutions *for the same reasons*. My aim is not to take a stand on which account of the normativity of promises and resolutions is correct (to be honest, though I have my biases, I'm not sure which is), nor to survey every view of the normativity of promises in the literature and show how they also explain the normativity of resolutions. Rather, my aim is to examine a sample of popular theories of promising—Thomas Scanlon's expectation theory and David Owens' authority interest theory—and to show that if one of these successfully grounds the normativity of promises, it would also ground the normativity of resolutions.<sup>9</sup>

I have already shown that promises generate expectations about how others will act. For some philosophers, the value of the expectations created by promises explains why promises bind. Scanlon (1990), for instance, argues that negligently causing someone to expect that you will  $\phi$  and then failing to  $\phi$  amounts to deception and is therefore morally wrong. There is also a case to be made that the value of the expectations that resolutions generate explains why resolutions bind. An agent who negligently causes herself to expect that she will  $\phi$  and then fails to  $\phi$  engages in *self-deception*. Although the wrong of self-deception

<sup>9</sup> In an extended version of this paper, I also examine Seana Shiffrin's (2008) theory that grounds the normativity of promising in the interests we have in forming intimate relationships with others and John Rawls' (1955) theory that promises derive their normativity from the value of the conventions that support them. I have omitted these analyses in this version of the paper for length and readability, but would be happy to discuss them.

may not be the same as that of the deception of others (c.f. Baron, 1988), it is hard to deny that there is something objectionable about self-deception, especially if an agent seeks to be self-governing and authentic. Like promises, the normativity of resolutions may be grounded in the wrongness of deception, whatever that wrong amounts to.

On another view of promising advanced by David Owens (2006), promises derive their normativity from our authority interest: our desire, in some situations, to have authority over what others do. To understand authority interest, it's helpful to think about the difference between communicating an intention to  $\phi$  and promising to  $\phi$ . Suppose I tell you that I intend to pick you up from the airport, but I don't promise that I will. According to Owens, I refrain from making a promise because I don't want to be *bound* to that course of action. If something else comes up, I want to be free to alter my plans without wronging you. I want, in other words, to retain *authority* over my actions. If had I promised you I would pick you up from the airport, I would have ceded my authority to you, granting you the power to require me to give you a ride. There are many reasons why I might want in some circumstances to grant you this authority, deriving from your interests and needs as well as my own. Promises exist and derive their power, Owens argues, to serve our authority interest.

It's very plausible that *resolutions* exist and derive their power to serve our authority interest *over ourselves*. If resolutions allow us to overcome the threat of temptation to our self-governance, then they clearly serve our authority interest over ourselves. (Here I diverge slightly from Owens' analysis. In his discussion of authority interest, Owens asks us to consider an akratic agent who judges that she ought to give up smoking but knows perfectly well that she is unlikely to exercise self-control (p. 70). Owens argues that she might nevertheless have an interest in retaining her *right* to exercise self-control. She might, for instance, reasonably refuse to cede her decision-making authority to someone else, even if she would be better off doing so. I agree with Owens that the akratic agent may reasonably wish to retain her right to decide what she shall do even if she knows that she is unlikely to act in accordance with her better judgment. But I think that when an agent is *unable* to resist temptation, she fails to exercise authority over her actions even if she retains her right to exercise self-control. Our authority interest is closely tied to our ability to resist, and therefore to our ability to form and respect resolutions.)

I have argued that resolutions play the same role in our intrapersonal lives that promises play in our interpersonal lives, and that the normativity of resolutions can plausibly be explained in the same way as the normativity of promises. This should lead us to think that resolutions exert their normativity in the same way promises do: one incurs a reason to  $\phi$  by resolving to  $\phi$  in the same way that one incurs a reason to  $\psi$  by promising to  $\psi$ .

What kinds of reasons does one incur when one makes a promise? First, one incurs a reason to perform the promised act. Second, if one fails to perform the promised act, one incurs a reason to mitigate the effects of that failure. If I promise to pick you up from the airport, I have a reason to be at the airport when your flight lands. And if for some reason I can't make it, I have a reason to mitigate the effects of breaking my promise. At a minimum, I have a reason to apologize to you. But depending on why I failed to fulfill my promise, I might also have reason to help you find another ride home. Analogously, when one makes a resolution, one incurs a reason to perform the resolved act. And if one fails to perform the resolved act, one incurs a reason to mitigate the effects of that failure. If you resolve to limit yourself to one episode of TV so you can get some writing done in the evening, you incur a reason to do just that. And if you fail to fulfill your resolution and watch more than one episode, making it impossible for you to get writing done in the evening, you incur a reason to mitigate the effects of your actions by, say, doing extra writing tomorrow.

If resolutions give us reasons for action, then our ability to form them is a *normative power*: a power, that is, we have over our normative circumstances. Other phenomena regarded as normative powers include promising, consent, and forgiveness. When we make a promise, we give ourselves reasons to perform the promised act and make ourselves accountable to the person to whom we made the promise. When we give consent, we waive certain rights we have, making it permissible for others to perform acts that would otherwise be impermissible. And when we forgive, we relieve others of their debts to us. Alongside these phenomena, resolutions are a power we have to create reasons for action through sheer acts of will. That it is possible for our reasons to be "up to us" in this way is a remarkable thing.

## 4 THE BOOTSTRAPPING PROBLEM

Broadly speaking, there are two ways understanding the kind of normativity resolutions have. On one approach, our capacity to make resolutions is a power we have to create normative reasons by our willing. Call this the *reasons view*. On a second approach, we are rationally required to follow through on our resolutions as a matter of means-end coherence. On this approach, although we have reasons to become the kinds of people who are disposed to follow through on resolutions, resolutions themselves do not give us reasons. Call this the *requirements view*.

There are important differences between the reasons and requirements views, between thinking that we have *reasons* to follow through on our resolutions and that we are *rationally required* to follow through on our resolutions. Reasons have weight in practical reasoning, but are slack. They can be overruled by other, weightier reasons, but they always stay on the scene. Here is an example to illustrate how the two views work. If resolutions are reason-giving, then when you resolve to limit yourself to one episode of TV so you can get writing done in the evening, you incur a reason to limit yourself to one episode of TV so you can get writing done in the evening. But new information might come to light that overrides your previously all-things-considered reason to watch only one episode. If a friend calls you and asks you to binge-watch a TV show with her to take her mind off a recent breakup, you incur a reason to watch more than one episode of TV that overrides the reason you incurred from your resolution. By contrast, requirements have no weight and are strict; you either satisfy them or you don't, and they cannot be outweighed by other requirements. If we are required to follow through on resolutions as a matter of means-end coherence, then if you resolve to limit yourself to one episode of TV so you can get writing done in the evening, you must adopt the means to achieve that resolution and refrain from watching a second episode. But if your recently single friend calls you and asks to binge watch TV, you can abandon your resolution and adopt a new end of watching more than one episode, for which you are now rationally required to adopt the means.

The view I have offered in this paper is a *reasons view*: I have argued that resolutions are reason-giving. However, philosophers tend to be skeptical of the claim that mental attitudes (including intentions and resolutions) give us reasons because of concerns about *bootstrapping*. If mental attitudes gave us reasons, the

worry goes, then we could bootstrap any action into rationality simply by acquiring the relevant mental attitude. John Broome (2001) summarizes the bootstrapping problem clearly and forcefully:

The view that intentions [and therefore resolutions] are reasons is implausible. If you have no reason to do something, it is implausible that you can give yourself a reason just by forming the intention of doing it. How could you create a reason for yourself out of nothing? Suppose, say, that you have no reason either for or against doing some act, and you happen to decide to do it. So now, if intentions are reasons, you have a reason to do it. Since you have no contrary reason not to do it, the balance of reasons is in favour of your doing it. But [...] [i]t is implausible that just deciding to do something can make it the case that you ought to do it, when previously that was not the case (Broome, 2001, p. 87).

There are two kinds of cases where resolutions appear to lead to problematic bootstrapping. The first is *Buridan's donkey cases*, cases where one has the exact same reasons to pursue multiple courses of action. Suppose a donkey finds himself halfway between two bales of hay. He has no reason to prefer one to the other. But he needs to eat, so he decides to head toward the one on his right. Suppose now that a gust of wind blows the bale of hay on his right away and he is now closer to the bale of hay on his left. If his resolution to pursue the bale of hay to his right gave him a normative reason to head toward it, then it appears as though he still ought to walk toward it even though he would be better off pursuing the one to his left. He has problematically bootstrapped his pursuit of the bale of hay to the right into rationality. The second kind of case where resolutions appear to bootstrap problematically involves resolutions to perform immoral or self-destructive actions. If morality demands that we give to charity, it seems implausible that we could rationalize *not* giving to charity simply by resolving and thereby giving ourselves a reason to do so.

The requirements view easily sidesteps the bootstrapping challenge. To see why this is the case, we must distinguish between narrow and wide scope versions of means-end coherence:

*Narrow scope:* If you adopt the ends, then rationality requires of you that (you adopt the necessary means).

*Wide scope:* Rationality requires of you that (if you adopt the ends, then you adopt the necessary means).<sup>10</sup>

<sup>10</sup> The parentheses indicate the scope of the rational requirement.

Because means-end coherence is taken to have a wide scope, you can choose to fulfill the requirement either by adopting the necessary means *or by giving up the end*. No bootstrapping occurs; if you abandon an end and adopt a new one, you are no longer rationally required to take the means to your previous end.

I suspect that the requirements view's avoidance of problematic bootstrapping is one reason why philosophers have been hesitant to pursue the reasons view. But if I am right that the reasons resolutions give us are akin to promissory reasons, then we can respond to bootstrapping concerns by thinking about why promises are not susceptible to them. After all, on all accounts, promises give rise to reasons: when you promise to  $\phi$ , you incur a normative reason to  $\phi$ . Therefore, if promises don't bootstrap problematically, resolutions shouldn't either.

Consider, first, Buridan's donkey cases. Does the donkey's resolution to pursue the bale of hay on the right give him reason to continue pursuing it even if it blows away? We can explain why it does not by thinking about how changing circumstances render promises (and the reasons they generate) irrelevant. Suppose I promise to pick you up from the airport when your plane lands on Tuesday. I now have a reason to drive to the airport on Tuesday. But if your flight is cancelled and you decide to rent a car and drive home instead, I no longer have a reason to go to the airport on Tuesday; the changing circumstances have rendered my promise irrelevant. Similarly, even if resolutions give us reasons, changing circumstances can render those reasons irrelevant. Buridan's donkey resolves to walk toward the bale of hay to his right, but if that bale of hay is blown away, his resolution becomes irrelevant. He now has most reason to walk toward the bale of hay to his left, since it's closer.

Let's now turn to the second kind of case where resolutions appear to lead to problematic bootstrapping: cases involving immoral or self-destructive resolutions. Do immoral *promises* bind? If I promise someone I will rob a store, do I now have an obligation to rob the store, despite the demands of morality? The answer to this question is controversial. Joseph Raz (1977) argues that immoral promises *do* bind. For him, promises give rise to obligations, and fulfilling obligations will sometimes require agents to perform acts even if they should not be performed on the balance of reasons (p. 224). Holly Smith (1997) disagrees. On her view, it would be astonishing and troubling if the power to promise allowed us to dictate the moral status of actions. If resolutions give rise to reasons that are akin to promissory reasons, then we can respond to concerns about immoral and

self-destructive bootstrapping by appealing to either of these arguments. We can follow Raz and argue that resolutions of this kind bootstrap, but not problematically, or we can follow Smith and argue that these resolutions do not bootstrap at all.

I suspect that we can appeal to a comparison between resolutions and promises to respond to other scenarios where resolutions-based reasons appear to lead to unacceptable bootstrapping. That promises are not susceptible to problematic bootstrapping suggests that resolutions shouldn't be, either.

#### REFERENCES

- Ainslie, G. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.
- Ainslie, G. (2013). Intertemporal Bargaining in Addiction. *Front Psychiatry*, 4, 63.
- Ainslie, G., & Haslam, N. (1992). Hyperbolic Discounting. In G. Loewenstein & J. Elster (Eds.), *Choice Over Time* (pp. 57–92). New York: Russell Sage Foundation.
- Arpaly, N., & Schroeder, T. (1999). Praise, Blame and the Whole Self. *Philosophical Studies*, 93(2), 161–188.
- Axelrod, R. (1980). Effective Choice in the Prisoner's Dilemma. *Journal of Conflict Resolution*, 24(1), 3–25.
- Baron, M. (1988). What's Wrong With Self-Deception? In B. P. McLaughlin & A. O. Rorty (Eds.), *Perspectives on Self-Deception* (pp. 431–449). Berkeley: University of California Press.
- Bratman, M. E. (1983). Taking Plans Seriously. *Social Theory and Practice*, 9(2-3), 271–287.
- Bratman, M. E. (1987). *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Bratman, M. E. (2000). Reflection, Planning, and Temporally Extended Agency. *The Philosophical Review*, 109(1), 35–61.
- Bratman, M. E. (2007). *Structures of Agency: Essays*. New York: Oxford University Press.
- Broome, J. (2001). Are Intentions Reasons? And How Should We Cope with Incommensurable Values? In C. W. Morris & A. Ripstein (Eds.), *Practical Rationality and Preference: Essays for David Gauthier* (pp. 98–120). New York: Cambridge University Press.
- Crisp, R. (1997). *Mill on Utilitarianism*. New York: Routledge.
- Dannenberg, J. (2015). Promising Ourselves, Promising Others. *The Journal of Ethics*, 19(2), 159–183.
- Ferrero, L. (2010). Decisions, Diachronic Autonomy & the Division of Deliberative Labor. *Philosophers' Imprint*, 10(2), 1–23.

- Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5–20.
- Fruh, K. (2014). The Power to Promise Oneself. *The Southern Journal of Philosophy*, 52(1), 61–85.
- Griffin, J. (1986). *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford: Clarendon.
- Hobbes, T. (1894). *Leviathan*. London: Routledge.
- Holton, R. (2009). *Willing, Wanting, Waiting*. New York: Oxford University Press.
- Kant, I. (2017). *The Metaphysics of Morals* (L. Denis, Ed. & M. Gregor, Trans.). Cambridge: Cambridge University Press.
- Muñoz, D. (2020). The Paradox of Duties to Oneself. *Australasian Journal of Philosophy*, 98(4), 691–702.
- Oakley, T. (2017). How to Release Oneself from an Obligation: Good News for Duties to Oneself. *Australasian Journal of Philosophy*, 95(1), 70–80.
- Owens, D. (2006). A Simple Theory of Promising. *Philosophical Review*, 115(1), 51–77.
- Rawls, J. (1955). Two Concepts of Rules. *Philosophical Review*, 64(1), 3–32.
- Raz, J. (1977). Promises and Obligations. In P. Hacker & J. Raz (Eds.), *Law, Morality, and Society: Essays in Honour of H.L.A. Hart*. Oxford: Clarendon Press.
- Rosati, C. S. (2011). The Importance of Self-Promises. In H. Scheinman (Ed.), *Understanding Promises and Agreements* (pp. 124–155). New York: Oxford University Press.
- Sartre, J.-P. (1956). *Being and Nothingness* (H. E. Barnes, Trans.). New York: Washington Square Press.
- Scanlon, T. (1990). Promises and Practices. *Philosophy & Public Affairs*, 19(3), 199–226.
- Schaab, J. D. (2021). On the Supposed Incoherence of Obligations to Oneself. *Australasian Journal of Philosophy*, 99(1), 175–189.
- Schofield, P. (2021). *Duty to Self: Moral, Political, and Legal Self-Relation*. Oxford: Oxford University Press.
- Shiffrin, S. V. (2008). Promising, Intimate Relationships, and Conventionalism. *Philosophical Review*, 117(4), 481–524.
- Smith, H. M. (1997). A Paradox of Promising. *The Philosophical Review*, 106(2), 153–196.
- Sripada, C. (2016). Self-Expression: a Deep Self Theory of Moral Responsibility. *Philosophical Studies*, 173(5), 1203–1232.
- Wall, S. (1998). *Liberalism, Perfectionism and Restraint*. New York: Cambridge University Press.