

*Forthcoming in Noûs*  
(please cite published version—early view located [here](#))

## **Self-Prediction In Practical Reasoning: Its Role And Limits**

Stephen J. White  
Northwestern University

### Abstract

Are predictions about how one will freely and intentionally behave in the future ever relevant to how one *ought* to behave? There is good reason to think they are. As imperfect agents, we have responsibilities of self-management, which seem to require that we take account of the predictable ways we're liable to go wrong. I defend this conclusion against certain objections to the effect that incorporating into one's practical reasoning predictions concerning one's voluntary conduct amounts to evading one's responsibility for that conduct. There is, however, some truth to this sort of objection. To understand the legitimate role of self-prediction in practical reasoning, we need to distinguish instances of coping responsibly with an anticipated failure to behave as one ought, on the one hand, from mere acquiescence in one's flaws, on the other. I argue that, to draw this distinction, we must recognize certain limits on the use of self-prediction as a ground of choice.

Imagine a friend of yours has just come into a large sum of money. You ask him what he will do with it. "I'm probably going to take it to the track and gamble it away," he says. You try to dissuade him. It would be better to put it into savings, or perhaps spend it on some needed home repairs. He agrees that these things would be better, but says that you're missing his point. He's not saying gambling away the money is what he thinks he *should* do. He is simply predicting this is what he *will* do. And he might insist that he has good grounds for making such a prediction. "Ask anyone who knows me what I'm likely to do with a windfall such as this and they'll tell you the same thing."

There is, intuitively, something wrong with your friend's attitude toward his future behavior. He treats his own conduct as if it were merely something to speculate about rather than a matter for him to decide. In doing so, he displays a kind of Sartrean bad faith, a failure take responsibility for himself and his actions (Sartre, 1956).

This conclusion might at first incline us toward the view that agents ought to adopt a strong anti-predictive stance toward matters that are subject to their will. If someone is

considering how she herself will choose to behave in various circumstances, she should not be attempting to make a prediction about what she is likely to do. Rather, she should view the choice as hers to make, and thus as confronting her only with the question of what the right choice would be under the circumstances.

And yet, this seems to go too far. A realistic account of our responsibilities as imperfect moral and rational agents needs to include what we might call responsibilities of self-management. We may know of ourselves that, at least in some situations, we are prone to procrastination, inconstancy, errors of judgment, bouts of anger or depression, and so on. Managing ourselves responsibly in light of these things often requires making predictions about when and in what ways we are liable to make regrettable choices.

In fact, there appears to be a tension between two intuitively attractive ideas concerning the role self-prediction might play in an agent's practical reasoning. On the one hand, there are the requirements of self-management just mentioned. As flawed creatures, it seems we should to some extent try to anticipate the ways in which we are likely to behave badly and do what we can to preempt or mitigate those moral or prudential failures.

This may involve doing things that, ideally—if we were better or stronger people—we would avoid doing. For example, suppose you recognize it would be nice for you call your mother to wish her a happy Mother's Day—though only if you will be kind to her and don't say things to hurt her feelings. However, given the way your conversations with your mother usually go, you have good reason to predict you'll end up saying something cruel or cutting once you're on the phone with her. Arguably, the right thing to do under these circumstances is to refrain from making the call.

The other idea is that is that there is something objectionably evasive in appealing to predictions of our own voluntary conduct in order to justify doing what we would otherwise have reason not to do. In the example just given, you could, after all, call your mother without saying anything hurtful.<sup>1</sup> Isn't *that* what you should do? To deny this on the basis of a prediction that this is not something you're likely to do, the thought goes, amounts to treating the choices you will make as though they were merely features of your situation you need to deal with and respond to. And this seems to be a way of dissociating from the exercise of your own agency.

---

<sup>1</sup> Or so I am assuming. I return to this point below.

The worry is that a person cannot dissociate herself in this way from her future choices and conduct without in some sense denying her responsibility for how she chooses to behave.

These two ideas would not be in tension if the requirements of proper self-management concerned only behavior that would not itself count as an expression of one's agency for which one is non-derivatively responsible—for example, if they concerned only unintentional or compulsive behavior. If that were so, one could only be expected to take responsibility for the predicted behavior insofar as one had some indirect control over its manifestation—much as one might be responsible for coming down with an illness, having failed to take the relevant precautions. In that case, the second idea expressed above simply wouldn't get a foothold. But in fact, it seems much of what we need to anticipate and preemptively respond to is conduct that we are or would be more immediately responsible for. A person might have reason to expect that in certain situations she is likely to behave poorly in the face of some temptation or provocation, without this implying that she must therefore not be directly responsible for what she does in such situations. The issue is whether it's possible for the person herself to take it as a given—something that needs to be dealt with—that she is likely to make certain choices in response to a particular situation without in some way failing to acknowledge that those choices amount to an exercise of her agency and that she is directly (non-derivatively) responsible for them as such.

One aim of this paper is to examine and clarify the two ideas just outlined and to consider the extent to which they are in tension. To preview, although I think there is something to the second, anti-predictive idea, I will argue that, properly understood, it does not preclude the deliberative use of self-prediction altogether. After describing the problem in more detail (§1), I will consider several arguments against self-prediction and show that they do not support a blanket ban on its deliberative employment (§§2-4). What these arguments do suggest, however, is that the normative assessments of conduct that would ideally govern one's behavior—assuming one were fully rational, virtuous, strong-willed, and so on—retain some independent normative significance even in the non-ideal case. The other aim of the paper is to spell out this last claim and thus argue for a particular constraint on the role of self-prediction in practical reasoning. In particular, I will argue in §5 for what we might think of as an anti-opportunism constraint, one motivated by the need to rule out forms of reasoning that would enable a person to take advantage of anticipated wrongdoing or imprudence to pursue certain interests she would be required to set aside if she were more virtuous, rational, or strong-willed. I then consider, in

§6, the special role of what I call agency-supporting acts in cases where self-management is called for. I sum up my conclusions in §7.

## 1. The Core Predicament

I'll begin by laying out the structure of the sort of deliberative predicament that is my main focus in this paper. The problem of self-prediction is clearest when the following conditions hold:

*First*, given the relevant values at stake in your situation, the best thing for you to do, among the options available to you (in the sense described below) would be to perform a course of action that involves doing several things over a span of time—say, doing  $\phi$  at  $t_1$ , then  $\psi$  at  $t_2$ , and so on.<sup>2</sup> Call this optimal course of action *A*.

*Second*, it is up to you whether you perform *A*, at least in the following sense. First, there are no external obstacles that will prevent you from doing *A*, so that as long as you exercise your will appropriately at the relevant times, you will do *A*; nor are you suffering from some pathological compulsion that would prevent you from exercising your will appropriately at the relevant times. Second, you know (or are in a position to know) what is involved in doing *A* under the circumstances, as well as how to do each of the things involved.<sup>3</sup> Given this, if you do not do *A*, this will be because you yourself have freely chosen to act in a way you were at least in a position to know would be inconsistent with completing this course of action.<sup>4</sup>

---

<sup>2</sup> In referring to the “best” course of action, I mean the course of action that is best supported overall by whatever values, interests, and standards are relevant under the circumstances. I do not assume that “best” in this sense means “impartially best,” or that the best thing to do is the thing that yields the best total outcome.

<sup>3</sup> So, for instance, we can't say that it is up to you, in the relevant sense, whether to tell your German-speaking cook, in German, that you find his Spätzle delicious merely because there is some series of sounds that, strung together, amount to pronouncing the German phrase, “die Spätzle sind köstlich,” and you know how and are able to produce each of those sounds with your mouth and moreover have the opportunity to do so (your mouth isn't taped shut; the cook is right in front of you). Though these prerequisites are in place, you may not know that this is precisely what is involved in telling the cook, in the language he understands, that you find the Spätzle delicious. If, in this perfectly ordinary sense, you do not know how to say that in German, then paying your cook this particular compliment is not option for you.

<sup>4</sup> This sense in which it is up to you whether you do *A* is diachronic, in that it concerns the possibility of a sequence of opportunities to exercise your agency in certain ways over time. The idea is that whether you do *A* or not depends on how you exercise your will over time. It's important to distinguish this from the notion of what one has intentional or deliberative control over at a particular time. Thus, it may not be the case that you are in a position, at

*Third*, under the circumstances and given the values at stake, it would be better for you not to  $\phi$  at  $t_1$  than it would be for you to  $\phi$  at  $t_1$  but fail to do the other things involved in *A*.

*Fourth*, you accept, or at least have good reason to accept, the above three claims about your situation.

Now, from your point of view as an agent considering a situation of this kind, how, if at all, might facts about what you are in a position to predict, as far as your future behavior is concerned, be relevant to what you ought to do? Suppose, for instance, you have grounds for predicting that when the time comes you'll decide against  $\psi$ -ing at  $t_2$ . Does this have any legitimate bearing on your decision about whether to  $\phi$  at  $t_1$ ?

Before considering possible answers, a note about what this question is asking: Because I am interested, here, in the deliberative standpoint of the agent, the question should be understood as employing the sense of “ought” (or “should”) that plays a distinctive deliberation-concluding role. An agent’s judgment that she ought to do something is normally sufficient to bring her deliberation about whether to do it to a close. If she is rational, then from her point of view there will be nothing further to consider in deciding whether to actually do the thing. By contrast, were a person to take seriously the question of whether to do what she ought to do, this would indicate a form of practical irrationality.<sup>5</sup> The question, then, is whether an agent’s predictions about how she is likely to behave in some situation are relevant to her thinking about what she *ought* to do, in this deliberatively relevant sense.

---

any point prior to  $t_1$ , to ensure that you will do *A* by forming the intention, at that point, to do *A*. The difference between these notions will be relevant especially for the discussion of Marušić’s argument, taken up in §3. On the difference between these notions see, for example, Portmore (2019). Portmore argues there that whether a person, at a particular time, objectively ought to do something (i.e., ought, given to the facts of her situation, regardless of what she believes or should believe about her situation) depends on whether she can ensure she will do it by way of the intentions and other attitudes she has at that time.

<sup>5</sup> Contrast this with one’s taking seriously the question of whether to do what one legally ought to do. This by itself does not indicate any failure in relation to standards of rational agency. Thus, as Broome (2013) has noted, the deliberation-concluding sense of “ought” is the one relevant to our understanding of *akrasia* as a form of practical irrationality—the rational failure to intend to do what one believes one ought to do..

It will be useful, as a starting point, to contrast two broad approaches to understanding the role of self-prediction. The first approach draws a sharp distinction between the perspective the agent takes toward her future conduct, which is essentially opposed to making predictions or estimating probabilities, and the perspective of an outside observer, who may regard the agent's conduct as a more or less predictable feature of his circumstances. On this view, the question of what a person is likely to freely do in the future—as opposed to what she has good reason to do—is, or should be, irrelevant to the person's own practical deliberation and provides no justification for her choices one way or another.

The second approach holds that the question of how probable it is that one will act in this or that way in the future is quite generally relevant to one's current decision as to what one should do. Roughly, on this approach, what one should do at any given moment is take whatever action has, of the alternatives under consideration, the greatest *expected* value, calculated in light of the evidence available to one. Because the expected value of choosing a given option partly depends on the probabilities one has reason to assign to the different possible actions one will be in a position to perform in the future, given the choices one makes now, such probability assignments are, on this view, directly relevant to determining how one should behave in the present (Jackson, 2014).<sup>6</sup>

As I claimed in the introduction, the idea that we have some responsibility to manage our flaws as agents, and that doing so requires taking account of our own predictable bad behavior, has a good deal of initial plausibility. As temporally extended agents, we are vulnerable to the possibility that we will lose our nerve or change our minds or for some other reason fail to carry out our current plans. We need to grapple with this in determining what our plans should be.<sup>7</sup>

---

<sup>6</sup> The contrast between the two approaches just outlined has clear parallels with the debate between “actualists” and “possibilists,” with respect to the question of what role an agent's future conduct plays in determining what is, objectively, the morally right thing for that agent to do. The debate between actualists and possibilists has for the most part been concerned with what is *objectively right* for a person to do—what a person ought to do given the objective facts of the situation, abstracting from what he or she believes or is in a position to know. The question I am interested in, by contrast, is what an agent ought to do given that she has certain information relevant to how she is likely to act under various conditions. I think that this will lead more directly to some of the central questions that are at issue between possibilists and actualists—questions about complacency, recklessness, and responsibility that require consideration of the agent's point of view on herself and her situation, and not merely consideration of what her objective circumstances are. For defenses of actualism, see Goldman (1976); Jackson & Pargetter (1986); Jackson (2014); Sobel (1976). For possibilism, see Goldman (1978); Feldman (1986); Humberstone (1983); Thomason (1981); Zimmerman (1996); Zimmerman (2017).

<sup>7</sup> Because this sort of vulnerability depends on the fact that we exercise our agency over time, I will set aside “synchronic” versions of the core predicament, which are often discussed in the debate between actualists and possibilists (see the previous note). That is, it might be best for you to do  $\phi$  and  $\psi$  simultaneously, but better to do

Nevertheless, there are several interesting arguments in favor of the strict anti-self-prediction approach that are worth considering. Though I will argue that they do not fully succeed, the issues they raise are sufficient to cast doubt on the idea that the justifiability of choosing to perform an action at a particular time is simply a matter of the expected value of that act given the probabilities the agent has reason to assign to the various things she might subsequently choose to do.

## 2. “Deliberation Crowds Out Prediction”

I begin with an argument due to Wolfgang Spohn (1977) and developed by Isaac Levi (1997). The discussion is in a way preliminary to the main issue, because the argument is not really meant to rule out all reliance on self-prediction in the service of self-management. But it is helpful to consider, if only to show clearly its limitations.

The argument is that, in Levi’s memorable phrase, “deliberation crowds out prediction,” (Levi, 1997, p. 81). Why does deliberation crowd out prediction? Levi and Spohn both begin with the claim that the assignment of any unconditional subjective probability to the agent’s choosing a particular option plays no intelligible role in the agent’s own deliberation about whether to choose that option. And this seems entirely correct. Suppose, for instance, you’re considering whether to get up or stay in bed. And suppose you judge that you will more likely than not decide to stay in bed. How could this speak either in favor of or against staying in bed?

It’s easy to see the problem if we imagine that you are initially inclined to take the fact that you will probably choose to stay in bed to count in favor of getting up. To the extent that this convinces you that you should decide to get up, you should revise the probability you assign to choosing to stay in bed, to the point where you will have, in choosing to get up, undermined your

---

neither than to do  $\phi$  without at the same time doing  $\psi$  (perhaps  $\phi$ =ducking your head under water, and  $\psi$ =holding your breath). However, the possibility that you won’t  $\psi$  doesn’t raise the kind of issue it raises where there is a temporal gap between doing  $\phi$  and doing  $\psi$ . In the diachronic case, even if you are motivated to pursue the best course of action, you may have to reckon with the possibility that by taking action in pursuit of it, you will end up changing your mind and doing something worse than had you pursued some alternative course of action instead. This is not something you have to worry about if there is no possibility that your intent to both  $\phi$  and  $\psi$  will be effective in bringing you to  $\phi$ , but be abandoned before it comes time to  $\psi$ . The only real danger of  $\phi$ -ing without simultaneously  $\psi$ -ing (where these things are under your control) would come from the possibility of your deciding to  $\phi$ -without- $\psi$ -ing. The question I’m interested in, here, however, concerns the role of predictions that are not themselves based on one’s decisions about how to behave.

basis for making that choice. For this reason, it seems clear that you cannot coherently take a high probability that you will choose to stay in bed to count against making this very choice.

Taking unconditional probabilities regarding which choice you are about to make to be relevant to your deliberation concerning that choice is no more sensible, however, on the assumption that a high probability of deciding on a certain alternative counts in favor of that alternative. In general, insofar as you are deliberating between various options, you regard these options as available for your choice. That is, you view it as up to you which to choose. But then you must also regard it as up to you to overturn any prior prediction as to which option you'll choose. Such a prediction therefore won't place any normative constraint on the decision it is reasonable for you to make.<sup>8</sup>

This argument seems to me correct, as far as it goes.<sup>9</sup> Notice, however, that the argument only demonstrates the irrelevance of considering the probabilities as to whether you will choose this or that option as the conclusion of your current deliberation concerning those very options. It doesn't tell against considering predictions as to how you're likely to respond to some future situation in the course of deciding, not about that, but about how to respond to your more immediate situation and range of options.<sup>10</sup> Thus, for example, nothing in the argument rules out appealing to the prediction that you will speak harshly to your mother, once you have her on the phone, in the context of deliberating about whether to get her on the phone in the first place. Since the prediction that you'll end up speaking harshly is not a prediction about how you'll come down on the practical question you're currently considering, the Spohn/Levi argument doesn't apply. But this is generally how self-prediction enters into deliberation in typical cases of strategic self-management. So, even if we were to accept Levi's stronger claim that the point of view of the deliberating agent is incompatible with assigning any probabilities at all to possible conclusions of one's current deliberation, this would not itself cast doubt on the idea that there is a legitimate role for self-prediction when it comes to managing our imperfections as agents.

---

<sup>8</sup> Thus, Spohn: "The decision maker chooses the act he likes most—be its probability as it may." (Spohn, 1977, p. 115).

• <sup>9</sup> It is a further, and more controversial question, whether Levi is correct that an agent *cannot*, while deliberating, assign probabilities to hypotheses about what she will decide as a result of her deliberation. For discussion see Rabinowicz (2002); Hájek (2016); Liu and Price (2018).

<sup>10</sup> Levi himself emphasizes this point. See esp., Ch. 4 of (Levi, 1997).



### 3. Epistemic Evasion

Although the Spohn/Levi argument applies only to self-predictions concerning the outcome of the agent's current deliberation, some have held that there is a more general tension between a person's view of herself as an agent and the sort of attitude she takes toward her conduct in attempting to predict what she will do on the basis of evidence. In this section, I'll consider the claim that insofar as a person attempts to predict how she will behave in some situation—on the basis of anything other than a decision about what to do in such a situation—she thereby treats the matter as though it were not up to her.<sup>11</sup> If this claim were it true, it would ground an objection to self-prediction in instances of what I called the core predicament. For in these cases, the conduct in question *is* assumed to be up to the agent. To engage in self-prediction would then constitute a failure to acknowledge this—hence, a failure to accept responsibility for her conduct.

The most sophisticated recent development of this line of thought is due to Berislav Marušić (2015). According to Marušić, if it is up to the agent whether or not she will  $\phi$  in some situation, then for her to predict that she will  $\phi$  on the basis of evidence—as opposed to making a decision to  $\phi$  on the basis of the practical reasons she takes to count in favor of  $\phi$ -ing—would be for her to engage in what Marušić calls “epistemic evasion.” Here is his official formulation:

*Epistemic Evasion:* We epistemically evade our decision if and only if we treat a practical question as a theoretical question; that is, if and only if, when it is up to us to  $\phi$ , we settle the question of whether we will  $\phi$  by considering our evidence about whether we will  $\phi$  if we decide to  $\phi$ , rather than by considering the practical reasons for and against  $\phi$ -ing (Marušić, 2015, p. 112).

What is the problem with epistemic evasion? According to Marušić, “The wrongness of epistemic evasion consists in the fact that we seek to predict what we will do when we should decide what to do. We thereby treat a matter that is up to us as if it were not up to us” (2015, p. 113).<sup>12</sup> In doing this, we fail to take appropriate responsibility for the way we exercise our agency (Marušić, 2015, p. 111).

---

<sup>11</sup> For a classic statement of this sort of position, see Hampshire and Hart (1958).

<sup>12</sup> Cf. Hampshire & Hart (1958): “If a man does claim to be able to predict with certainty his own future actions, basing his prediction on induction, then he is implying that the actions in question will be in some sense, or to some degree, involuntary, the effect of causes outside his own control” (p. 2).

If, on the basis of past experience, you predict you will be sarcastic and mean to your mother if you call her, you treat the question of how you will speak to your mother as though it were not a matter for you to decide on the basis of practical reasons. But this, Marušić claims, is to regard it as not up to you what you will say to your mother once you have her on the phone. And to regard it as not up to you is to treat it as something for which you are not responsible, at least not directly. Hence, in predicting your own eventual meanness, you regard yourself as not being responsible for what you would say to your mother, supposing you were to call her. But what you would say to your mother is a matter of the choices you make, choices for which you would, by hypothesis, be responsible. Therefore, in arriving at your view about how you'll behave by consulting evidence rather than engaging in practical reasoning, you fail to acknowledge your responsibility for the way you choose to respond to whatever provocations might emanate from the other end of the line.

I do not think this argument succeeds in showing that seeking to predict one's voluntary behavior necessarily amounts to evading responsibility for one's agency. When you predict that you're likely to be cruel to your mother if you call her, you do treat the question of how you will speak to your mother as one you are not in a position to settle by way of a decision you can make *now*, prior to placing the call. This is because you suspect that any decision or resolution to speak kindly to your mother is one you'll ignore or abandon once you are on the phone with her. But this is not equivalent to denying that what you say will ultimately be determined by decisions you yourself will make, and for which you will be responsible. Concluding that you are not now in a position to simply decide that you will speak kindly to your mother once you have her on the phone—in any sense that entitles you to regard the matter as settled—is consistent, therefore, with acknowledging that you are the one responsible for how you choose to speak to her if you call her. It does not in itself amount to a denial of responsibility for your agency.

This is not to say that Marušić isn't onto something important in worrying about the possibilities for evasion that arise in the kind of case we've been discussing. And he's right to think it's a problem—a failure to live up to an ideal of responsible agency—if you do not take yourself to be in a position to determine how you will behave on the basis of a conclusion about how you *should* behave. But I think we need to understand this as a problem with your agency—and thus as a problem you have to face—rather than as a problem with the suggestion that you ought to take account of evidence to the effect that you will probably fail to act as you should.

#### 4. Letting Oneself Off the Hook

Perhaps there's another way to understand the charge that relying on self-prediction in one's practical deliberations is evasive. Suppose we agree that relying on predictions about how one is likely to behave in some situation does not in itself constitute a denial that one would be directly responsible for that behavior. Still, there might seem to be a kind of bad faith involved insofar as one relies on such predictions to excuse oneself from obligations one would otherwise have. David Enoch has recently put the point like this: "If I say that you ought to save the drowning child, ...[and] you respond with 'but I am highly unlikely to' or 'but I'm not gonna,' your response—if it is supposed to be a response at all—is barely intelligible, and it certainly doesn't refute my ought-statement or the blame that is likely to follow if you violate it," (2018, p. 2). The choice not to pursue the optimal course of action seems in itself to constitute a failure to live up to the relevant standards of behavior that apply to your situation. To rationalize this choice by appeal to the fact that you would not have been willing to do the things required to meet those standards appears to be little more than a way to avoid taking responsibility for this failure.

One way of spelling this out would be in terms of the following claim:

(1) What a person has most reason to do does not depend on what she would or would not be willing to do.

That a person would not be willing to do what is required to carry out some course of action, cannot by itself establish that she should not pursue that course of action. (1) is *prima facie* plausible. But to assess it, we should first distinguish it from a closely related claim:

(2) What a person has most reason to do (as of a certain time) does not depend on what she is (at that time) willing or not willing to do (Cf. Goldman, 1978; Portmore, 2011)).

Here's the difference. (1) implies that how you would choose to behave in some situation *other* than the present one cannot make a difference to what you have most reason to do *in your present situation*, whereas (2) says only that what you have most reason to do at any given time cannot depend directly on what sorts of things you would *at that time* be willing to do. Insofar as it allows that what one might or might not choose to do at some point in the future may make a difference to what one now has most reason to do, (2) does not rule out the practical relevance of self-prediction. At any rate, this is so as long as your predictions about what you're liable to do

in this or that situation are not themselves based on your current intentions concerning how to behave in such situations.

Claim (2) seems to be supported by the idea that, if reasons are to serve the function of guiding and constraining a rational agent's decision-making, they must be suitably independent of the decisions that agent is anyway disposed to make (Cf. Kolodny, 2005; White, 2018). It also encapsulates an anti-bootstrapping principle, according to which one cannot simply create (or eliminate) reasons at will (Cf. Bratman, 1987; Broome, 1999; Brunero, 2007). But neither of these concerns provide a basis for (1). And once we distinguish (1) from (2), it might start to seem that insisting on (1) simply begs the question against those who think self-prediction can play some legitimate role in practical reasoning. Much of the intuitive force behind (1) seems to be captured by (2), and we are left to contend with the contrary intuition that, as imperfect agents, we have certain responsibilities of self-management, which require us to take a realistic view of when we're likely to behave in ways we shouldn't.

But there is more to be said. Consider another claim of Marušić's. He argues that insofar as we take it that what we have reason to do is constrained by the evidence concerning what we are likely to be willing to do, we fail to recognize the full weight or significance of our practical reasons. This, he says, is because we would then be treating the normative significance of considerations that favor a decision to  $\phi$  as limited by their capacity to support a prediction that we will indeed  $\phi$  if we decide to do so. But, this would surely be a mistake. As he put it, "The weight of our practical reasons, considered as practical reasons, does not correspond to their evidential weight, considered as predictors of what we will do if we intend to do something" (2015, p. 114).

Why should we think that, if we allow that self-prediction can bear on an agent's practical reasoning, we are committed to this conclusion—namely, that the weight or significance of a practical reason in favor of an action is determined by the extent to which it gives the agent reason to believe she will perform that action if she decides to? Marušić's argument for this begins with the idea that rationality requires one's intentions to cohere with one's beliefs about one's future behavior. That is, one is rationally required not to simultaneously intend to  $\phi$  while believing there is a significant chance one will not  $\phi$  (Marušić, 2015, pp. 57-58). This thought is supported by standard assumptions about the role that decisions and intentions for the future play in our mental economy. On Bratman's account, for instance,

intentions function to provide a basis for further planning. They do so in part by excluding from consideration alternatives that are either incompatible with what one intends, or redundant, and in part by allowing one to take for granted certain facts about one's future circumstances so that one can address downstream practical problems that are likely to arise (Bratman, 1987). But it would seem that an intention can rationally function in these ways to constrain and guide one's subsequent practical reasoning only if one assumes one will actually carry out that intention (Harman, 1976). So, it seems, it would not be rational for one to (flat out) intend something if one justifiably believed it was unlikely that one would carry out that intention.<sup>13</sup>

Suppose that's right—that it's irrational for one to intend to do something while believing there's a good chance one won't carry out that intention. And suppose one maintains that, in light of all the relevant information about oneself and one's situation, one should believe there's a good chance one would not  $\phi$  even if one were intend now to do so. In that case, if one is rational, one won't take oneself to be rationally permitted at this point to form an intention to  $\phi$ . Hence, whatever reasons there are that count in favor of  $\phi$ -ing under the circumstances, one won't regard them as sufficient to warrant a decision at this point to  $\phi$ . Rather, one will take some set of practical reasons to support a decision to  $\phi$  only insofar as one thinks that there's a sufficiently high probability that would follow through on an intention formed in light of those reasons. But this, Marušić claims, is to treat the weight or significance of the practical reasons counting in favor of the relevant course of action as determined by how, if at all, they affect the predictions it would be reasonable to make as to whether one will carry out that course of action. Again, this would be a distorted view of the normative significance of practical reasons in general.

This argument, however, is not sufficient to establish Marušić's conclusion. Suppose that, by considering what one knows about one's weaknesses, susceptibilities to temptation, etc., one came to believe that one would be unlikely to carry out any intention one might now form to act in a certain way in the future. Marušić is right that it would be objectionably evasive for one to conclude from this that because one is thus not in a rational position, at present, simply to decide or intend to do the thing in question, the considerations counting in favor of acting in that way

---

<sup>13</sup> The qualification "flat out" is meant to flag that we sometimes use "intention" to talk about less committal states, for example, what Bratman calls "endeavoring," which, he claims, is not subject to the same norms of consistency as intending (strictly speaking). See (Bratman, 1984). See also Richard Holton's account of partial intention in (Holton, 2008).

are therefore irrelevant and can safely be ignored. This would be to treat relevant practical considerations as if they did not matter. But a proponent of the idea that self-prediction can play a legitimate role in practical deliberation need not endorse that last inference. Even if one is not in a rational position (in light of one's predictions about how one is likely to behave) to flat out intend to do something in the future, there's no reason to think one should conclude from this that the considerations that would favor doing it are irrelevant. They may, for instance, be relevant to the decision as to whether one should take the initial steps in pursuit of a course of action *despite* the risk that one will fail to see it through.

Consider, for instance, the sort of view I sketched in §1, on which the justification of an agent's decision to do something depends on the expected value of doing it, where this takes into account how probable it is that she will go on to do various things in the future (Jackson, 2014). On this view, the fact that there's a good chance the agent would not carry out a certain plan of action can affect the justification for embarking on it; but it's not decisive on its own. The importance of the practical considerations favoring the relevant course of action also makes a difference. Even if the agent has reason to predict she will probably abandon the pursuit, it's perfectly possible that, given the significance of what's at stake, she should nevertheless take the first steps, in the hopes that she'll surprise herself and see the thing through. Imagine that Green has committed a serious crime and is now standing outside the police station, considering whether to turn himself in. Suppose, however, he knows it's likely that were he to enter the station to make his confession, he would simply turn around and walk out without saying a word. Even so, it could be that the moral importance of owning up to what he's done establishes that the expected value of entering the station is greater than not doing so, despite the high probability that it will come to nothing. Evidently, this view about how to understand the relevance of self-prediction does not end up conflating the significance of practical considerations as such with their significance as predictors of future behavior.

Still, one might feel that this response does not do justice to the intuition described at the beginning of this section. Suppose one chooses, on the basis of an (accurate) expected value calculation, not to pursue the course of action that would be best under the circumstances. Won't this nevertheless count as a failure to live up to the relevant standards of behavior that apply to one under the circumstances? And by appealing, in support of that choice, to the fact that one

would (probably) not do what's required to meet those standards in any case, isn't one just acquiescing in that failure?

In fact, I think there is something right about this last way of putting the objection. And, I will argue in the next section, it points to a serious problem with the straightforward expected value approach. But the problem is not with the general claim that an agent's predictions of her own behavior can be relevant to the choices she should make—can sometimes serve to justify opting for a second-best course of action. Rather, the problem is that it does not allow us to adequately distinguish between the case of an agent responsibly coping with her predictable failures, on the one hand, and the case of an agent merely acquiescing in them, on the other.

## 5. The Limits of Self-Prediction

Consider again the Jacksonian expected-value approach—the idea that the agent should assess the expected value of taking the initial steps involved in the optimal course of action, given the odds that she won't follow through, and compare this with expected value of the other options she's currently in a position to choose. If it is important enough to carry out a certain course of action, and if the opportunity costs are low, then it may be reasonable for the agent to embark on it even if the chances of follow-through are not good. But otherwise it will make sense to forgo the optimal course altogether.

This account can neatly explain why it might make sense for you to send your mother an email, rather than call her, on Mother's Day. If we assume that being cruel to your mother on Mother's Day is significantly worse than sending an email—which will be disappointing, but is better than nothing—then given a high enough probability that you'll end up saying something cruel if you call your mother, the expected value of calling will be lower than the expected value of sending the email.

Consider, however, a different kind of case. Pierre's friend has just revealed an embarrassing secret in the expectation that Pierre will keep it between them. Suppose that Pierre betrays his friend's confidence on the grounds that he will almost certainly do so eventually—he is bad at keeping secrets—and so he may as well do so now, given some opportunity for him to gain through the revelation. Perhaps telling what he knows in present company would improve Pierre's standing with the popular crowd. Or maybe it would just relieve him of the pressure of

keeping it to himself. Since he has little hope of remaining a loyal friend for long, there's not much to be gained by forgoing these benefits of immediate betrayal. Thus reasons Pierre.

Here's another one. Imagine that my doctor has told me that I need to keep a strict diet in order to bring my cholesterol down to a healthy level. But I've never been good at keeping diets and I think it quite unlikely I'll stick with this one long enough to do much good. So, I figure there's little point in passing up the delicious, extra-cheese pizza that's being served for lunch.

These do not seem to be responsible attempts by a person to manage the consequences of his vices. They seem rather to be attempts to rationalize and evade responsibility for succumbing to temptation. This suggests that the practical relevance of self-prediction is more limited than the expected-value approach implies. A prediction of less-than-ideal behavior on one's part does not always provide one with reason to do the next best thing.

But what exactly has gone wrong in these cases? I might have every reason to think that I will fail to stick to my diet for any length of time. And if that's so then forgoing the pizza in favor of salad today won't make any significant difference to my health—it will only deprive me of some pleasure. Similarly, though it would be ideal for Pierre to keep his friend's secret forever, if we assume he is likely at some point to betray her, it may in fact *be* better, or at least no worse, if he does so now. But what's the difference between these evaluations and the ones on which you base your judgment that you should just send your mother an email, cold as this may seem on Mother's Day, since it would at least be preferable to calling her up and being cruel to her, which is what you're likely to do if you call?

Perhaps the difference is that Pierre is acting on selfish motives, whereas your concern is still for your mother? This does not seem to me the best explanation. First, if the suggestion is that other-regarding considerations should trump self-regarding ones even where one predicts that one will fail to live up the relevant standards, this won't help explain what has gone wrong in my second, dieting example. In that case, all the relevant considerations are self-regarding. Second, we can imagine that Pierre is moved by purely altruistic considerations. He might, for example, want to tell his friend's secret in order to cheer up someone else who needs cheering up, and who would get a kick out of this particular story (perhaps it's a story that would make this other person feel better about some similarly embarrassing thing she has done). For Pierre to pursue even this altruistic aim still seems to me to constitute an objectionable acquiescence in his own failings as a friend.



Even where Pierre is not acting selfishly, he does not at all appear to be attempting to manage or mitigate the problems stemming from his tendency to betray those who confide in him. Rather, he seems to be accepting that as a given—a fixed feature of his motivational psychology.<sup>14</sup> Contrast our Mother’s Day case. In deciding not to pick up the phone, you are attempting to avoid a situation in which you don’t trust yourself to behave as you should. You can thus be seen as trying sincerely to avoid or at least mitigate the damage done by your tendency to react poorly when, for instance, your mother asks yet again about your love life.

Of course, there’s a sense in which Pierre’s revealing the secret now will partly mitigate the damage from his failure to be a trustworthy friend. After all, in doing so, he will avoid the worst overall outcome: namely the betrayal minus the benefits (to himself or others). But in seeking to avoid this outcome, the reasons that move Pierre to his betrayal are reasons that (we are assuming) have already been accounted for and overridden in the determination that he ought to keep his friend’s secret. When Pierre decides to reveal the secret, he is thus choosing to give in to the very considerations he rightly judges do not warrant betraying his friend’s confidence. The fact that he expects to betray her at some point does not introduce any new considerations. It merely leads him to bracket those that speak in favor of remaining loyal. In predicting that he will eventually give in to his desire to share the secret, he concludes he is entitled to ignore the value of ongoing fidelity. But this seems no more than a failure to acknowledge the force and seriousness of the reasons deriving from that value.

To elaborate: Consider what the main objections are to the worst of the courses of action at issue in the different cases. In the Mother’s Day example, the primary objection to calling up your mother and saying something cruel is presumably an objection to hurting your mother’s feelings, particularly on a day when she should be made to feel loved. In Pierre’s case, most would agree that the main objection to sharing his friend’s secret to no good end is not the *pointlessness* of the betrayal but the betrayal itself. And so, by sending the email to your mother (hence opting for the second-best course of action), you are acting so as to mitigate what is *primarily objectionable* about the worst-case scenario. We cannot say the same of Pierre’s choice to reveal his friend’s secret. In taking the second-best alternative (telling the secret when there is some advantage to be gained by doing so), Pierre does nothing to avoid what would be the main

---

<sup>14</sup> Ditto for my decision to break my diet.

objection to his conduct were he to end up taking the worse course of action—namely the betrayal of friendship involved in revealing the secret.

Why should this difference between the cases matter? The primary reason for preferring the ideal case over the worst case in the Mother’s Day scenario—namely, the reason to avoid causing your mother to feel hurt and unloved—is itself a reason to prefer sending an email over calling (given what is actually likely to happen if you call). The choice not to call is thus compatible with acknowledging the force of, and in some sense respecting, the original rationale for doing what, ideally, you should do. Insofar as you are responsive, at least in part, to this rationale, you have not entirely dissociated your present choice from the normative standards that would regulate your conduct if you were a more virtuous agent.

None of this is true of Pierre, though. For Pierre, the reason for choosing the second-best option (telling the secret now) is *not* itself a reason for preferring the ideal course of action over the worst course of action. Indeed, the reason motivating him to tell is in fact *opposed* to the considerations favoring the ideal course. The fact that he could, say, get a good laugh out of revealing the secret is by hypothesis overridden by the reasons that support ongoing loyalty to his friend. Thus, when Pierre lets the secret out for the sake of a laugh, he not only gives up on doing the best he could, under the circumstances; he also acts in pursuit of an *end* that, ideally, he would regard as inadmissible under the circumstances, given the other values at stake.

Thus far I have been criticizing Pierre’s motives. But on the assumption that the problem with them is that Pierre is motivated by considerations that do not constitute good reasons—reasons that would serve to justify his actions—my suggestion can be put as follows. Regardless of any predictions about how one is likely to behave, the fact that a course of action would serve some value or end can constitute sufficient reason for taking that course of action only if there would be no *conflict* between (a) acting for sake that value or end, given one’s circumstances and (b) performing the *optimal* course of action available under the circumstances—i.e., the course of action one would ideally perform, supposing one were virtuous and disciplined enough to see it through.

More formally, in cases of the kind I’ve been discussing, there are three alternative courses of action whose relative merits are at issue. There is an optimal course of action, *A*, which involves, say, doing  $\phi$  at  $t_1$  and  $\psi$  at  $t_2$ ; there is a second-best course of action, *B*, which precludes  $\phi$ -ing at  $t_1$ ; and there is a third, *C*, which involves  $\phi$ -ing at  $t_1$ , but not  $\psi$ -ing at  $t_2$ , and

which would be worse, all things considered, than either *A* or *B*. Independently of any consideration of what the agent is likely to do, we can distinguish between two classes of reasons that favor *B* over *C*. The first class consists of considerations that favor *B* over *A* as well. Call this the class of “ideally-defeated” reasons, since at least in the ideal case, where the agent has no reason to doubt that she will behave as she should, these reasons in favor of *B* would be overridden or cancelled by the reasons that favor *A*. In the second class, there are reasons for preferring *B* to *C* that do not represent any respect in which *B* would be preferable to *A*—either because they are also reasons to prefer *A* over *B*, or because they do not cut one way or another in favor of either. Call these the “ideally-undefeated” reasons, since they are in no way opposed to the agent’s performing the ideal course of action, supposing she could count on herself to follow through with it. My proposal is that, where the agent has grounds for predicting that she will end up doing *C* rather than *A* if she decides to  $\phi$  at  $t_1$ , she is permitted to opt for the second-best, *B*, if and only if the considerations in the class of ideally-undefeated reasons are themselves sufficient to justify this choice over  $\phi$ -ing at  $t_1$  (given the risk that the agent will end up doing *C* if she  $\phi$ ’s at  $t_1$ ). The considerations in the class of ideally-defeated reasons lend no justificatory support to the second-best option.

For example, the good of ensuring that your mother’s feelings aren’t unduly hurt is perfectly compatible with, and in fact furthered by, the best course of action open to you. It’s because of this that this aim remains a legitimate source of justification in your current situation. Hence, the fact that in your non-ideal condition this aim would be better served by pursuing a second-best course of action constitutes a valid (and in this case, probably decisive) reason for you to opt for that instead. By contrast the aim of improving his standing with the popular crowd, given that the only realistic way to do this is for Pierre to betray his friend’s secret, is itself opposed to (and overridden by) the considerations favoring the optimal course available to Pierre—the course he would take were he a better friend. And so, although we can concede it would be better, all things considered, for Pierre to betray his friend in such a way as to improve his standing with the popular crowd than it would be for him to forgo this benefit and betray his friend anyway, the respect in which this would be better does not constitute a good reason for him to choose the former option.<sup>15</sup>

---

<sup>15</sup> Note that, for this constraint to rule anything out, there must be reasons that can come into genuine conflict. Hence, if all reasons were grounded in a single value—e.g., the maximization of aggregate happiness—the anti-

Why should it matter that the reasons favoring Pierre's betrayal would be defeated if he were a more ideal moral agent, whereas the reasons favoring your choice not to call your mother are ideally-undefeated, given that the whole point is that you and Pierre are flawed, *non*-ideal agents? The intuition behind this proposal is that we should rule out forms of reasoning that would enable a person to take advantage of anticipated wrongdoing or imprudence to secure benefits she would have to forgo—or avoid costs she would have to bear—if she were more virtuous, rational, or strong-willed. Pierre should not betray his friend for the sake of amusement. He should not even betray his friend in order to cheer someone else up. His interest in getting a good laugh, or in brightening someone's day, is overridden by his duty not to betray his friend's trust. The fact that he expects to betray his friend sooner or later does not change this. He should not be entitled to use the fact that he's not going to remain loyal as itself a justification for pursuing what he would otherwise have to give up. To do that would be to treat his general lack of fidelity as a kind of boon from the standpoint of his desire to amuse others. And that does not seem like an acceptable attitude to take toward this flaw in his character. It's therefore illegitimate for him to use the fact that he's bound to betray his friend's trust eventually as a justification for revealing the secret when he calculates that doing so would be to good effect.<sup>16</sup>

Self-prediction can only contribute to the justification for pursuing a next-best course of action, then, subject to what we might think of as an anti-opportunism constraint. That is, it can help to justify a decision to pursue some less-than-ideal course of action only if the anticipation of failure cannot be viewed as an *advantage* relative to the motives for that decision. While one may seek to mitigate the foreseeable damage to oneself and others flowing from one's deficiencies as an agent, one may not take such deficiencies as license to act in pursuit of ends or values that, under the circumstances, one should give up or set aside.

---

opportunism constraint would be trivial. I take it, though, that this sort of monism is seriously implausible. It would imply not only that there was just one type of non-instrumental value (pleasure, say) but that one's relation to instances of that value (whether, for instance, it's one's own pleasure at stake, or someone else's) could introduce no conflict among one's reasons for various alternatives.

<sup>16</sup> Of course, we can imagine variations on Pierre's case in which he would be permitted to reveal the secret for reasons not ruled out by the anti-opportunism constraint. Imagine, for instance, Pierre predicts that, if he doesn't reveal the secret now (to get it off his chest), he's liable to do so when it would do significantly more damage to his friend's reputation—or even to some third party's reputation. Because the value of preserving the other's reputation is in no way opposed to his doing the optimal thing and keeping silent, it's the kind of value that could provide a valid justification for doing the next best thing and betraying his friend's secret in a context where the reputational damage would be minimal.

## 6. Agency-Supporting Acts

Of course, even if there are admissible reasons, not ruled out by the anti-opportunism constraint, for a person to opt for a less-than-ideal course of action in view of her anticipated failure to follow through on the optimal course, these may not ultimately be sufficient to justify doing so. I won't offer a full account of when such reasons would be sufficient to make it rationally and morally permissible to act in a way incompatible with the optimal course of action available. That would require a full normative theory. But some things can be said at this point about certain factors that are relevant to this question.

First, in determining whether, in a particular case, an agent ought to pursue a second-best course of action rather than attempting the best, we will need to consider just how unlikely it is that one will fail to follow through on the optimal course of action, supposing one were to attempt it. Other things equal, the less improbable it is that one will follow through, the more difficult it will be to justify not even trying. Second, it will matter whether the really important thing is to successfully carry out the optimal course of action or whether it is to avoid the worst-case scenario of beginning down the best path and then failing to follow through. In the former situation, the prospect of conducting oneself as one ideally should may be worth even a high risk of failure. In the latter case, where the important thing (relative to the values admitted by the anti-opportunism constraint) is primarily to avoid the disaster of a failed attempt, it may be that one should not take the risk, even where there is a decent chance one would manage to hold it together and succeed in doing the optimal thing.

I take it these points will be relatively uncontroversial. In the remainder of this section, I want to ask whether there might be other, more interesting constraints on the use of self-prediction in practical reasoning.

Consider another case: Patty is trying to quit smoking. She is aware that there is a good chance she will relapse at some point, and take it up again. (She's been through this before; she knows the statistics, etc.) Suppose, then, that she decides to hang on to the pack of cigarettes she bought a few days before resolving, yet again, to quit. She reasons as follows: "The odds that I will succeed in quitting for good this time are not great (let's be honest). And cigarettes these days are expensive. If I hang on to the pack I just bought, which is still almost full, then, if and

when I start smoking again, I will have thereby saved myself a good \$10. So, it would be foolish to throw it away.”

Intuitively, there is something wrong with Patty’s line of thought. It evinces a failure on her part to take seriously her current effort to quit, insofar as she seems to be choosing to act in a way likely to undermine that effort. But it’s not ruled out by the anti-opportunism constraint. Patty can’t be said to be treating her predictable failure to maintain her resolve as an opportunity to avoid costs she would—in the absence of any defect in her will—otherwise have to bear. Her frugality is, in this situation, perfectly consistent with the ideal of quitting once and for all.

This suggests we might need to posit a further constraint. We can begin by noting that, for the addict, tossing the leftover cigarettes is not just a matter of getting rid of clutter; it is part of a strategy for improving the chances of success. It is what I’ll call an agency-supporting act. It is agency-supporting, in this case, because it removes what is a predicable obstacle to successfully quitting—namely, the ready availability of a cigarette at moments of intense craving. Agency-supporting acts can take other forms—setting up external incentives, for example, or adopting techniques to concentrate one’s attention on certain considerations rather than others, or reaching out for various types of social support. Such actions often introduce new risks attendant on the failure to follow through with one’s endeavor, as the example of throwing out the leftover cigarettes illustrates: Tossing out her recently purchased pack means Patty runs the risk of bearing an additional financial cost if she takes up smoking again.

In general, where an act carries a risk of some particular negative consequence if and only if the agent subsequently fails to act as she should, let’s call that a “failure-associated risk.” The idea, then, would be that we should treat the failure-associated risks of agency-supporting acts differently than risks associated with actions that aren’t agency-supporting in the above sense. Perhaps, for instance, the failure-associated risk of tossing one’s leftover cigarettes should be ignored, or at least discounted as compared with, say, the failure-associated risk of throwing out one’s crystal ashtray.<sup>17</sup> It certainly seems less problematic from the point of view of a person’s commitment to quit smoking for her to store away a moderately expensive ashtray, so that she won’t have to shell out for a new one in the eventuality that she takes up smoking again, than it does for her to hang on to her cigarettes for the same reason.

---

<sup>17</sup> I’m assuming that, unlike cigarettes themselves, whether or not one keeps a fancy ashtray will have no meaningful effect on how readily one gives in to the temptation to light up.

I'm unsure whether we should accept an additional constraint along these lines. On the one hand, what seems problematic about focusing, in one's deliberation, on the failure-associated risks of an agency-supporting act like throwing out one's cigarettes is something that could arguably be analyzed in terms of the act's expected value. After all, such actions increase the probability of success, and this may be sufficient to accommodate our intuitions. Throwing out the cigarettes might only increase the probability that one will stop smoking by a small amount, but considering the great benefits associated with quitting, this could be enough to ensure the expected value of tossing the cigarettes is greater than keeping them, even considering the cost of buying a fresh pack. Because of this, putative counterexamples to the expected-value analysis will be difficult to assess.

On the other hand, I'm inclined to think that the rationale provided for the anti-opportunism constraint might be extended to deliberation about agency-supporting acts as well. The basic motivation for that constraint concerned the stance it is appropriate for a person to take toward her flaws as an agent. Extending this idea, we might hold that, in general, the choice to act in ways that are opposed to, rather than supportive of, one's well-functioning as a rational agent is objectionable at least in part because the attitude one thereby expresses is at odds with a proper conception of oneself as an agent who is answerable to independent rational and moral norms, and not only because of how it might affect the outcome. If this captures the proper stance one is to take toward oneself as a rational agent, albeit an imperfect one, then we will want to think differently about agency-supporting versus agency-obstructing choices, as compared with other kinds of choices informed by predictions of failure. In considering how best to pursue a difficult end, like giving up smoking, the goal of mitigating the costs of failure should be of secondary concern relative to settling on a plan that offers the best hope for success. And this is not (or not merely) explained by the promise of greater expected gains. Rather, the reason for this priority would be that it directly expresses a commitment to an ideal of rational agency.

## 7. Conclusion

I began with the intuition that there is an apparent tension in how we understand the deliberative point of view of a person who is cognizant of and attempting to cope with certain ways in which

she falls short of perfection as a rational and moral agent. On the one hand, the attempt to predict and respond to facts about how she is likely to behave under various conditions—even with regard to matters that are up to her—seems to be at least sometimes necessary as a way of managing such flaws. On the other hand, for an agent to treat her own predictable unwillingness to do the right or prudent thing as something that needs to be accepted as a parameter on her choices, and accommodated in her practical reasoning, can seem at odds with a full acknowledgment of her responsibility for how she chooses to behave in the situations being contemplated. I hope to have shed some light on the nature of this tension, and how far it extends.

We should not court moral and prudential disaster by ignoring what we know about our flaws as agents. There is thus a legitimate role for self-prediction in the context of responsible self-management. But this role is subject to constraint. Predictions that we are likely to behave badly in some respect or other may be relevant to the justification of choices we make in anticipation of such failings. But this justificatory role needs to be regulated by an ideal of rational agency, lest a realistic assessment of our vices is transformed into an excuse for their manifestation. In this respect, at least, expected failures of agency do not function in deliberation merely as more or less probable outcomes, relevant only insofar as they bear on the expected value of one's current choices.<sup>18</sup>

---

<sup>18</sup> For helpful comments and discussion, I am grateful to Mark Alznauer, Sara Aronowitz, Rima Basu, Agnes Callard, Johann Frick, Louis-Philippe Hodgson A.J. Julius, Joe Karbowski, Beri Marušić, Tori McGeer, Tom Parr, Philip Pettit, Doug Portmore, and Aarthy Vaidyanathan. Earlier versions of this paper were presented at the Einstein Ethics Group colloquium, the Ethics and Political Philosophy Colloquium at the University of Toronto, the Ethics and Values Symposium at Loyola University, Chicago, the Practical Philosophy Workshop at Northwestern University, the Practical Philosophy Workshop at University of Chicago, the Saint Louis Annual Conference on Reasons & Rationality, and the University Center for Human Values at Princeton. Thanks to everyone present on those occasions. Work on this paper was supported by a Laurance S. Rockefeller Faculty Fellowship at the Princeton University Center for Human Values.

## References

Bratman, M. (1984). Two faces of intention. *Philosophical Review*, 375-405. doi: 10.2307/2184542

Bratman, M. (1987). *Intentions, Plans, and Practical Reasoning*. Cambridge, MA: Harvard University Press.



- 
- Broome, J. (1999). Normative requirements. *Ratio*, 398-415. doi: [doi.org/10.1111/1467-9329.00101](https://doi.org/10.1111/1467-9329.00101)
- Broome, J. (2013). *Rationality Through Reasoning*. Oxford, UK: Wiley Blackwell.
- Brunero, J. (2007). Are intentions reasons? *Pacific Philosophical Quarterly*, 424-444. doi: [doi.org/10.1111/j.1468-0114.2007.00301.x](https://doi.org/10.1111/j.1468-0114.2007.00301.x)
- Enoch, D. (2018). Against utopianism: noncompliance and multiple agents. *Philosophers' Imprint*, 1-32. doi: [hdl.handle.net/2027/spo.3521354.0018.016](https://hdl.handle.net/2027/spo.3521354.0018.016)
- Feldman, F. (1986). *Doing the Best We Can*. D. Reidel.
- Goldman, H. (1976). Dated rightness and moral imperfection,. *Philosophical Review* , 449-487. doi: 10.2307/2184275
- Goldman, H. (1978). Doing the best one can. In K. J. Goldman A.I., *Values and Morals* (pp. 185-214). Springer Dordrecht.
- Hájek, A. (2016). Deliberation welcomes prediction. *Episteme*, 507-528. doi: [doi.org/10.1017/epi.2016.27](https://doi.org/10.1017/epi.2016.27)
- Hampshire, S., & Hart, H. (1958). Decision, intention and certainty. *Mind*, 1-12.
- Harman, G. (1976). Practical reasoning. *The Review of Metaphysics*, 431-463.
- Holton, R. (2008). Partial belief, partial intention. *Mind*, 27-58. doi: [doi.org/10.1093/mind/fzn002](https://doi.org/10.1093/mind/fzn002)
- Humberstone, L. (1983). The background of circumstances. *Pacific Philosophical Quarterly*, 19-34. doi: [doi.org/10.1111/j.1468-0114.1983.tb00181.x](https://doi.org/10.1111/j.1468-0114.1983.tb00181.x)
- Jackson, F. (2014). Procrastinate revisited. *Pacific Philosophical Quarterly*, 634-647. doi: [doi.org/10.1111/papq.12051](https://doi.org/10.1111/papq.12051)
- Jackson, F., & Pargetter, R. (1986). Oughts, options, and actualism. *Philosophical Review*, 95, 233-255. doi: 10.2307/2185591
- Kolodny, N. (2005). Why be rational? *Mind*, 509-563. doi: [doi.org/10.1093/mind/fzi509](https://doi.org/10.1093/mind/fzi509)
- Levi, I. (1997). *The Covenant of Reason*., Cambridge UK: Cambridge University Press.
- Liu, Y., & Price, H. (2018). Heart of DARCness. *Australasian Journal of Philosophy*, doi: [doi.org/10.1080/00048402.2018.1427119](https://doi.org/10.1080/00048402.2018.1427119).

- 
- Marušić, B. (2015). *Evidence and Agency*. Oxford, UK: Oxford University Press.
- Portmore, D. (2011). *Commonsense Consequentialism*. New York: Oxford University Press.
- Portmore, D. (2019). *Opting for the Best*. Oxford, UK: Oxford University Press.
- Rabinowicz, W. (2002). Does practical deliberation crowd out self-prediction? *Erkenntnis*, 91-122.
- Sartre, J.-P. (1956). *Being and Nothingness*. New York: Philosophical Library.
- Sobel, J. H. (1976). Utilitarianism and past and future mistakes. *Nous*, 195-219.
- Spohn, W. (1977). Where Luce and Krantz do really generalize Savage's decision model. *Erkenntnis*, 113-134. doi: doi.org/10.1023/A:1020106622032
- Thomason, R. (1981). Deontic logic and the role of freedom in deliberation. In R. Hilpinen, *New Studies in Deontic Logic* (pp. 177-186). Springer, Dordrecht.
- White, S. J. (2018). Intention and prediction in means-end reasoning. *American Philosophical Quarterly*, 251-266.
- Zimmerman, M. (1996). *The Concept of Moral Obligation*. Cambridge, UK: Cambridge University Press.
- Zimmerman, M. (2017). Prospective possibilism. *Journal of Ethics*, 117-150. doi: doi.org/10.1007/s10892-017-9245-1