

How Action Governs Intention

Nishi Shah

Amherst College

Draft of 05/07/07

On Monday you are offered a large sum of money to intend on Tuesday to drink toxin on Wednesday. The toxin will make you ill if you drink it, but you only have to *intend* to drink it in order to receive the money. Obviously, it is desirable for you to intend to drink the toxin, yet this fact seems incapable of moving you to so intend. Why is this so?¹ Does your inability to win the money reflect a flaw in your rationality?

Michael Bratman has pointed out that when one deliberates about the future, one's focus is on *what to do* in the future, not *what to intend now*, although a decision about what to do later does lead to an intention now about what to do later (1987, 103). Understanding why this is so is the key to solving the toxin puzzle.²

1. Practical deliberation

Before looking for an explanation of the phenomenon that Bratman describes, we need a more careful characterization of it. First of all, it is not true that any deliberation we engage in about the future is about what to do in the future; we also deliberate about what will be the case in the future. This latter kind of deliberation concludes in a future-directed belief rather than a future-directed intention. I will call deliberation that concludes in a belief **doxastic deliberation** and deliberation that concludes in an intention **practical deliberation**. Bratman's claim then is about practical deliberation: deliberation that concludes in an intention now for later is about what to do in the future.

¹ This, of course, is Gregory Kavka's famous toxin puzzle (See Kavka (1983)). One question that philosophers have discussed in relation to this example is whether it would be rational to take actions that would cause you to have the intention to take the toxin. This is not the question I am discussing. I am asking why you cannot intend to take the toxin directly on the ground that it would be desirable to so intend, without having to go through indirect means to induce the intention.

² Interestingly, Bratman doesn't seem to think that the explanation of this fact is this is the key to solving the toxin puzzle. For his own resolution of the puzzle, see Bratman (1999).

We can generalize this claim. Any deliberation that concludes in an intention to *A* – any instance of practical deliberation—is about whether to *A*, where *A* refers to an arbitrary action immediate or future.³ Aiming to arrive at an intention with respect to *A*-ing via deliberation thus requires that you determine whether to *A*. Initiating deliberation by asking yourself the question *whether to intend to A* makes this aim explicit. We thus can put Bratman’s point this way:

In order to answer the deliberative question *whether to intend to A* an agent must answer the question *whether to A*.⁴

Deliberating whether to intend to *A* just is aiming to arrive at an intention as to whether to *A* by way of reasoning.⁵ Whether a process of reasoning is an instance of deliberation does not depend on whether it is initiated by a deliberative question; it depends instead on the aim with which the reasoning proceeds. As long as your aim is to arrive at an intention, your reasoning constitutes practical deliberation, whether or not it is initiated by an articulation of the question *whether to intend to A*. Although you needn’t ask yourself this question in order to initiate practical deliberation, normally you conclude

³ In the case in which practical deliberation leads to an immediate action—where we talk about intention-in-action—it might be argued that there is no difference between intending to *A* and *A*-ing. At the very least, it seems that when one is deliberating about an immediate action, one cannot view one’s intention as preceding one’s action, and thus one cannot think that there is a step involved between one’s intention and one’s action that requires one to execute the intention. If this is so, then for the purposes of deliberation, the agent cannot think of his intending to *A* and his action of *A*-ing as distinct events. From his deliberative point of view it thus would make no sense to distinguish considerations favoring intending to *A* from considerations favoring *A*-ing.

⁴ What is the nature of this fact about our practical deliberations? Is it merely a contingent psychological fact about creatures like us that our deliberations *whether to intend to A* can only be settled by determining *whether to A*, or is it a fact that applies to practical deliberation as such? As we shall see, I favor the latter interpretation, but I don’t think that this interpretation is a deliverance of the phenomenology of practical deliberation itself. As far as our experience is concerned, all we know is that our deliberations are constrained in this way. Therefore, I will leave it open for now that, for all we know, the fact that the deliberative question *whether to intend to A* can only be settled by answering the question *whether to A* is merely a contingent fact about our own practical deliberations; that is, for all we know, there are creatures who can arrive at an intention to *A* just by answering the question *whether it would be desirable to so intend*.

⁵ Note that I did not say that deliberating *whether to intend to A* just is intending to arrive at an intention to *A* by way of reasoning. While one might successfully intentionally induce such a specific intention by indirect means, I do not believe that intending to intend to *A* can cause one to intend to *A* by way of practical deliberation. On the other hand, such a second-order intention may already involve the first-order intention that is its object, in which case there is no need to do anything to carry out one’s intention to intend to *A*.

deliberation by answering another one. Normally you can arrive at an intention to *A* just by answering the question *whether to A*, which itself normally can be answered by considering whether *A*-ing is desirable.⁶ You cannot, however, arrive at an intention to *A* just by answering the question *whether it is desirable to intend to A*.⁷ The correct account of intention should explain these facts.

Although an agent cannot arrive at an intention to *A* just by deliberating about whether it is desirable to so intend, such considerations nonetheless might causally induce such an intention. After all, wishful thinking sometimes does influence what people intend to do. An explanation of the facts about practical deliberation described by Bratman thus must be compatible with the causal influence of such processes.

Elsewhere I have argued that the hypothesis that the concept of belief includes a standard of correctness best explains the fact that deliberation *whether to believe that p* (where *p* stands for an arbitrary proposition) can only be concluded by answering the question *whether it is the case that p* (i.e. *whether p*).⁸ In this paper, I will argue that the hypothesis that the concept of intention includes a standard of correctness best explains the fact that in order to conclude deliberation *whether to intend to A* one must answer the question *whether to A*.

In the next section (section two) I discuss some doubts about whether my description of practical deliberation is accurate. In section three I argue that the deliberative question *whether to A* is the normative question *whether one ought to A*, not the descriptive question *whether one in fact will A*. In section four and five I examine two alternative explanations of why deliberation that concludes in intention is directed at this question,

⁶ Obviously, the desirability judgments that settle practical deliberation are comparisons relative to one's options. The judgment that settles practical deliberation thus must be that an action is more desirable, or no less desirable, than any of one's other options (including the option of not doing anything). For ease of exposition I will continue to speak merely of the desirability of an action when referring to this comparative judgment.

⁷ In cases of deliberation about immediate action, an agent may be able to arrive at the intention to *A* by determining that it is desirable to intend to *A*, but, as I said in footnote six, in such cases intending to *A* and *A*-ing are not distinct events from the agent's point of view, and thus, from his point of view, determining that it is desirable to intend to *A* would be equivalent to determining that it is desirable to *A*.

⁸ See Shah (2003) and Shah and Velleman (2005).

turning to my own hypothesis in section six. Finally (section seven) I apply my hypothesis to solving the toxin puzzle.

2. *Is it really like that?*

There are at least two reasons for doubting that practical deliberation answers the question *what to intend*. First of all, it might be argued that given a proper understanding of that question, answering it cannot issue in an intention. The question *what to intend* can either mean *what shall I intend* or *what ought I intend*. If it means the former, then the answer must be of the form ‘I shall (or shall not) intend to ...’ which is the expression of a second-order intention. If the question *what to intend* instead is interpreted as the question *what ought I intend*, the answer must be of the form ‘I ought (or ought not) to intend to ...’ which is the expression of a normative judgment, not an intention.⁹ Either way, answering the question *what to intend* fails to issue in an intention. Practical deliberation, though, can and often does issue in an intention, therefore it is incorrect to describe it as answering that question.

I agree (and will argue later) that the answer to the deliberative question *whether to intend to A* is a normative judgment and that a normative judgment is not identical with an intention. However, the normative judgment that I ought to *A* normally leads directly to the intention to *A*. Certainly no further question needs to be considered before deliberation can conclude in the intention. Deliberation that aims to conclude in an intention whether to *A* thus can proceed by answering the question *whether I ought to A*.

But how can the judgment that I ought to *A* be the answer to the question *whether I ought to intend to A*? One can answer one question by answering another if the answer to the latter question determines the answer to the former. For example, if act consequentialism were the correct moral theory then the question *whether A-ing is morally right* would be settled by answering the question *whether A-ing maximizes overall (expected) value*. Similarly, if the answer to the question *whether to A* determines the answer to the question *whether to intend to A*, one can answer the latter question by

⁹ Thanks to John Broome for raising this objection.

answering the former. I thus see no problem with claiming that the deliberative question *whether to intend to A* can be answered by a judgment that I ought to *A* and that deliberation that answers this latter question normally concludes in an intention.

A second reason for doubting that practical deliberation is directed at answering the question *what to intend* is the following line of thought. We can deliberate only about those things that are under our voluntary control. Intentions, unlike actions, are not under voluntary control. Practical deliberation thus is about what action to perform—what to do—not what to intend. It is only by mistakenly thinking of our intentions as actions that we are led to believe that we can deliberate about what to intend. Alternatively, it is only in those abnormal cases in which our intentions are under our voluntary control, and thus are action-like in the relevant way, that we can deliberate about what to intend.

I agree that intentions are not actions; rather, they are the vehicles by which we exercise voluntary control over our actions. Our voluntary control over our actions therefore cannot in turn depend on our exercising the same kind of control over these vehicles, on pain of setting off a vicious regress. I also agree that practical deliberation is directed at the question *what to do*, or equivalently, *what action to perform*. But I deny the premise that one can deliberate only about those things over which one has voluntary control. The fact that we do not exercise voluntary control over our intentions does not entail that we cannot deliberate about what to intend. We do not exercise voluntary control over our beliefs either, but we clearly can and do arrive at beliefs by deliberating about what to believe. And surely it is even more obvious that we can and do arrive at intentions via practical deliberation.¹⁰

Of course, if the question *what to intend* were in competition with the question *what to do*, then the fact that practical deliberation is aimed at answering the question *what to do* would entail that it does not answer the question *what to intend*. But the questions

¹⁰ Nor does the fact that we exercise voluntary control over our actions but not our intentions entail that we cannot exercise a different kind of control over the latter. When we arrive at an intention by way of practical deliberation, we exercise our rational powers and thus exert rational control over the formation of that intention, just as we do when we arrive at a belief by way of doxastic deliberation.

what to do and *what to intend* are clearly compatible, since the way to answer the latter is by answering the former. But if answering the question *what to do* settles *what to intend*, then why not say that practical deliberation is directed equally at answering both questions? After all, any practical deliberation that successfully answers the question *what to do* thereby has settled the question *what to intend*.

But aren't there cases in which settling the question *what to do* doesn't settle the question *what to intend*? For example, suppose I have determined that I ought to eat breakfast tomorrow and that I also know that I am in the habit of eating breakfast. I thus can just trust that when the time comes I will eat breakfast out of habit. There is no need to "plan ahead" in this case, and thus no need to adopt an intention now to eat breakfast tomorrow.¹¹ When one engages in practical deliberation one presupposes that one thereby exerts control over what one ends up doing. In this case, although it is true that I will eat breakfast tomorrow whether or not I adopt a prior intention to do it, my deliberative control over this action lies in the fact that I will not eat breakfast tomorrow if I adopt a prior intention to do something else. Or, at least, this is what I must assume if I am to engage in practical deliberation about the matter.

My description of practical deliberation accommodates the possibility of such cases. My claim is only that in order to settle the question *whether to intend to A*—in order to engage in reasoning with the aim of forming an intention either to *A* or not to *A*—one must answer the question *whether to A*. While there may be cases in which one can settle the question *whether to A* without settling the question *whether to intend to A*, it is not possible to settle the latter question without answering the former one. The question *whether to intend to A* is not in competition with the question *whether to A*. In those cases in which settling the question *whether to A* does settle the question *whether to intend to A*—which I take to be the majority of cases—practical deliberation settles both questions at once.

¹¹ While one thus is rationally permitted not to adopt an intention in this kind of case, we would not say that one is irrational if one does adopt the intention.

There is a sense in which intention is the focus of practical deliberation, which comes out most clearly in cases of future-directed practical deliberation. In deliberating whether to see a movie tomorrow, I do not think that I will arrive at an action at the conclusion of my deliberation. How could I? Going to the movie tomorrow is not itself something that is the product of my current deliberation. What I hope to achieve by my deliberation is a plan, or intention, that I will execute tomorrow by either going or not going to a movie. As I said earlier, what makes my reasoning an instance of practical deliberation is not that it is initiated by a deliberative question, but that it proceeds with the aim of concluding in an intention. This is the sense in which I deliberate about whether to intend to go to a movie tomorrow.

But even in present-directed practical deliberation, it is plausible that all that one can accomplish by way of deliberation is an intention to act. Whether one is able to execute one's intention in action is not itself a product of one's deliberation. As John Broome puts it, action requires both reasoning ability and physical ability.¹² Failure to do something that one intends to do is not a failure of one's deliberative capacities, although it may involve a failure of some other capacities. If I intend to raise my arm now to ask a question but my arm suddenly becomes paralyzed or I become paralyzed by fear, my deliberation has not failed, although I have.

There is, however, a sense in which the question *whether to A*, not the question *whether to intend to A*, is the focus of practical deliberation. One normally needn't explicitly ask oneself *whether to intend to A*, since that question can only be answered by deliberating *whether to A*. Explicitly asking oneself *whether to intend to A* thus is almost always deliberatively idle. If one wants to answer that question, one might as well just start one's deliberation with the question *whether to A*. Asking oneself *whether to A*, however, cannot be settled by asking the different question *whether to intend to A*, since the latter question can itself only be answered by deliberating *whether to A*. Asking oneself *what to do* thus is indispensable for practical deliberation, whereas asking oneself

¹² Broome (2002), 85. One may also need to exercise further non-reasoning mental capacities in order to carry out an intention on a given occasion. For example, one may need to overcome fear to carry out an intention to perform a dangerous action.

what to intend is not. But the fact that one does not explicitly need to ask oneself the latter question does not mean that practical deliberation does not answer it. Similarly, asking *whether to believe that p* is deliberatively idle. The only way to answer that question is by deliberating whether *p*, so one might as well start one's doxastic deliberation with that question. This does not show that one does not answer the question *what to believe* in doxastic deliberation, it just shows that one needn't explicitly ask that question in order to answer it. That deliberation about whether to *A* or whether *p* does aim to arrive at intention or belief becomes clear upon comparison with non-deliberative reasoning about the same questions. One can consider *whether to A* or *whether p* in the spirit of idly wondering, without aiming to make up one's mind—in which case, one isn't deliberating whether to intend to *A* or whether to believe that *p*. Idly wondering whether to *A* or whether *p* is different from trying to make up one's mind whether to *A* or whether *p* in that one doesn't feel required either to come to a conclusion or to accept whatever conclusion one happens to arrive at. The latter constitutes such deliberation despite skipping the deliberative question *whether to intend to A* or *whether to believe that p* because it is reasoning aimed at issuing or not issuing in an intention or belief. Adequate accounts of the concepts of belief and intention ought to allow us to understand why deliberation that aims to arrive at a belief whether *p* starts directly with the question *whether p* and deliberation that aims to arrive at an intention whether to *A* starts directly with the question *whether to A*.

I have claimed that practical deliberation is directed at the question *whether to A*, which is settled (in part) by considerations relating to the desirability of *A*-ing, rather than the question *whether it is desirable to intend to A*. But it might be objected that one focuses on the desirability of *A*-ing in an attempt to determine whether it is desirable to intend to *A*. After all, in order to determine whether it is desirable to intend to *A*, one must determine whether the consequences of intending to *A* are desirable. *A*-ing is by far the most significant consequence of intending to *A*, therefore determining whether *A*-ing is desirable is crucial to determining whether intending to *A* is desirable.¹³

¹³ Thanks to Niko Kolodny for raising this objection.

The preceding discussion points to an aspect of the phenomenology of practical deliberation that this objection overlooks. When we engage in practical deliberation with an aim to arriving at an intention with respect to *A*-ing, our attention *immediately* centers on the question *whether to A*. There is no inferential step between the question *whether to intend to A* and *whether to A*; the former question immediately gives way to the latter. This is why we can skip the question *whether to intend to A* and start right in with the question *whether to A* and yet be recognizably deliberating about what to intend. But if the question *whether to intend to A* were the question *whether intending to A is desirable*, there would be an inferential step between the question *whether to intend to A* and *whether to A*, bridged by the premise that the desirability of intending to *A* is determined by the desirability of *A*-ing. Even this premise wouldn't be sufficient to bridge the gap though, because in determining the expected value (and thus desirability) of intending to *A* one would need to discount the desirability of *A*-ing by the probability that one's intention would be executed. This is not an accurate description of what occurs in practical deliberation. We don't adjust our reasons for intending to do something by the probability that the intention will be carried out; we determine our reasons for intending directly by determining our reasons for action.¹⁴

This concludes my defense of the characterization of practical deliberation I gave in section one. I realize that I have only canvassed some of the reasons that one might have for resisting this characterization. I hope, though, that it will be granted that what I have said is at least true of paradigmatic cases of practical deliberation, and I do not believe that we possess an adequate explanation of even these cases. This is the situation that this paper aims to remedy. Only once we have an adequate explanation of paradigmatic cases of practical deliberation about which we have clear and stable intuitions will we be in a position to attempt to assess and explain cases about which we have unclear and fragile intuitions.

¹⁴ Even if the phenomenology of practical deliberation does not support the hypothesis that we determine whether an action is desirable in order to determine whether intending to do it is desirable, it may be thought that such a hypothesis is extensionally equivalent to the one that I opt for later, in that it delivers the same verdicts about which intentions we have most reason to adopt. In the Appendix I argue that this is not so—that there are cases in which the two hypotheses diverge in their verdicts about what we have most reason to intend.

3. 'What to do?' ≠ 'What will I do?'

Some philosophers claim that intending to *A* involves believing that one will *A*.¹⁵ If this were true, would it explain why practical deliberation starts with the question *whether to A*? No. Presumably, it is by interpreting the question *whether to intend to A* as the question *whether to believe that one will A* that a doxastic theory of intention attempts to shed light on the nature of practical deliberation. But the question *whether to believe that one will A* is a doxastic question that can only be settled by answering the question *whether one will A*, not *whether to A*. These are very different questions, and very different kinds of considerations are relevant to answering them. The answer to the practical question *whether to A* is determined by reasons that show *A*-ing to be desirable; the answer to the doxastic question *whether one will A* is determined by evidence for the proposition that one will *A*.¹⁶ The binge alcoholic may have good evidence on Wednesday that he will go on a bender over the weekend—his addiction has gotten the better of him the past several weekends after all—without having any reason to think that it is desirable to do so.¹⁷

Have I have begged the question against this doxastic view of intention by implicitly assuming that the question *whether to A* is identical to the normative question *whether one ought to A*? All I have shown, it might be claimed, is that this normative question cannot be answered by the descriptive question *whether one will A*. From this we only can conclude that *if* the question *whether to A* is identical to the question *whether one ought to A*, it cannot be answered by determining whether one will *A*. But those who claim that intending to *A* involves believing that one will *A* deny the antecedent of this

¹⁵ Here is just a small sample of philosophers who have held some version of this view: Hampshire and Hart (1958), Chisholm (1970), Grice (1971), Audi (1973), Harman (1976), Davis (1984), Velleman (1989), and Setiya (2004).

¹⁶ Anscombe (2000), 6, makes essentially this point, but in terms of the reasons by which a speaker justifies the description that he gives to express his intention:

We might attempt to make the distinction out by saying: an expression of intention is a description of something future in which the speaker is some sort of agent, which description he justifies (if he does justify it) by reasons for acting, sc. Reasons why it would be useful or attractive if the description came true, not by evidence that it is true.

¹⁷ See Bratman (1981), 255, for a similar example.

conditional; that is, they deny that the question *whether to A* is a normative question. I thus cannot assume such an equivalence, which I would need to do in order to draw the conclusion that the question *whether to A* is not answered by determining whether one will *A*, without begging the question against them.

While I do believe that the question *what to do* is best interpreted as a normative question, my objection does not depend on this claim. All I need is one of the points about practical deliberation with which we began: normally one can arrive at an intention to *A* by determining that it is desirable to *A*. If the question *whether to A* were equivalent to the doxastic question *whether one will A*, only evidence that one will *A* could settle it. The desirability of *A*-ing is not evidence that one will *A*. Thus, were the two questions identical, it would not be possible to arrive at an intention to *A* by determining that it would be desirable to *A*. But we know from our own experience of practical deliberation that it is possible to arrive at an intention to *A* by determining that it would be desirable to *A*. Therefore the question *whether to A* is not identical to the doxastic question *whether one will A*.

Some philosophers who hold a doxastic account of intention, though, think that intending to *A* involves the *self-fulfilling* belief that one will *A*.¹⁸ This version of the doxastic account of intention may be able to explain why one can arrive at an intention to *A* by determining that *A*-ing is desirable. What it cannot explain is why one *must* determine that *A*-ing is desirable in order to deliberately arrive at an intention to *A*—why, for example, one cannot arrive at an intention to *A* just by determining that it is desirable to so intend. Self-fulfilling beliefs, just like ordinary beliefs, are capable of truth or falsity, but unlike ordinary beliefs, they needn't be based on prior evidence, since they cause their own truth. If one conceives of oneself as deliberating whether to believe that one will *A*, where one understands this to be a self-fulfilling belief, one thus needn't be guided by evidence that one will *A*. In fact, it would seem that one is free to answer the question *whether to believe that one will A* on whatever basis one likes. After all, no matter how one forms the belief, one knows that it will be true. Therefore one would be

¹⁸ See Grice (1971), Harman (1976), Velleman (1989), and Setiya (2004).

free to believe that one will *A* on the ground that *A*-ing is desirable. But this freedom would also include the freedom to believe that one will *A* on the ground that it is desirable to believe that one will *A*. Were the deliberative question *whether to intend to A* identical to the doxastic question *whether to adopt the self-fulfilling belief that one will A*, one thus could arrive at an intention to *A* just by determining that it is desirable to so intend. But we know from our own experience of practical deliberation that it is not possible (for us) to arrive at an intention to *A* just by determining that it is desirable to so intend. Therefore the question *whether to intend to A* is not identical to the doxastic question *whether to adopt the self-fulfilling belief that one will A*.

Recall the facts about practical deliberation that we are trying to explain: normally we can arrive at an intention to *A* by determining that *A*-ing is desirable, but we cannot arrive at an intention to *A* just by determining that it is desirable to intend to *A*. I have not argued that intending to *A* does not entail the belief that one will *A*. Rather, I have argued that even if true, this claim cannot account for these facts. It can explain one fact or the other, depending upon how the ‘belief’ involved in intention is interpreted, but no single interpretation of it can explain both facts at once.

Interpreting the question *whether to A* instead as the normative question *whether one ought to A*, or equivalently, *whether A-ing is the thing to do*, would accommodate both these facts. Determining that *A*-ing is desirable clearly bears on *whether one ought to A*, even if other considerations also bear on this question, whereas determining that intending to *A* is desirable does nothing by itself to settle the question *whether A-ing is the thing to do* (after all, no evaluation of *A*-ing is mentioned in the claim that *intending to A* is desirable).

Don’t Buridan’s Ass cases, though, show that the question *what to do* is distinct from the question *what should I do?* In such cases one is faced with at least two equally desirable options; thus there is no fact about which action one ought to perform. Settling the question *what to do* therefore must be settled on some other basis.

Remember, the question *what to do* whose interpretation we are seeking is one that is answered by deliberation. Therefore, Buridan's Ass cases show that this question is not identical to the question *what should I do* only if choosing one equally desirable option over another is the conclusion of such deliberation. But this is not so. In these cases one just *picks* between the equally desirable options; one's choice between them is not based on any reasons that favor one action over the other. We can call picking one equally desirable option over another settling *what to do* if we like, but we must not be misled into thinking that this is the same question that we seek to answer in practical deliberation by the fact that we use the same linguistic formulation to express both questions.¹⁹

I thus suggest that the question *whether to A* is best interpreted not as the doxastic question *whether one will A*, but as the normative question *whether one ought to A*. Accordingly, from this point forward I will treat the locutions 'whether to A,' 'whether one ought to A,' and 'whether A-ing is the thing to do' synonymously. Also, for ease of exposition, from this point forward I will be making the simplifying assumption that the desirability of an action is the only normative factor that determines whether one ought to do it.²⁰

4. The aim of intention: the dispositional interpretation

Before turning to my own hypothesis about why settling the deliberative question *whether to intend to A* requires answering the question *whether to A* (interpreted now as the normative question *whether A-ing is the thing to do*), it will be helpful to see why two closely related hypotheses cannot adequately explain why this is so. Understanding where these hypotheses go wrong will put us in a position to construct a better explanation.

¹⁹ When faced with two equally desirable options, can't one arrive at an intention to take one of them by determining that it would be more desirable to adopt it? If so, wouldn't this show that question *what one ought to do* is different from the question *what to intend*? Yes, I believe it would. I claim only that one must determine whether one ought to A in order to determine whether to intend to A. In Buridan's Ass cases *A-ing* is permitted but not required, thus the answer to the question *whether one ought to A* is *no*.

²⁰ For those bothered that this assumption expresses a prejudice in favor of a particular normative theory of practical reasons, when I speak of the desirability of an action please interpret me as referring to those features of an action, whatever they are, that make it worth doing.

Just as we say that belief aims at truth, we might say that intention aims at to-be-doneness. One way of interpreting this metaphor in the case of belief is to say that belief is a state of mind that is regulated for truth, in the sense that the belief that *p* tends to be formed, revised, and extinguished in response to evidence of *p*'s truth. According to such a view, the dispositions that make belief responsive to evidence of truth distinguish belief from other cognitive states. In this sense, truth-directedness might be thought to be constitutive of belief.²¹

Furthermore, the truth-directed nature of belief appears to explain why the deliberative question *whether to believe that p* can only be settled by answering the question *whether p*. If being regulated for truth is a conceptual fact about belief, then anyone who deliberates about whether to believe that *p* recognizes that his deliberation can potentially produce the envisioned attitude—that is, a belief that *p*—only by regulating that attitude for truth. Thus the only way to deliberate about whether to believe that *p* is to deliberate about whether *p*.²²

Similarly, one might interpret the metaphor that intention aims at to-be-doneness by saying that intention is a state of mind that is regulated for to-be-doneness, in the sense that an intention to *A* tends to be formed, revised, and extinguished in response to reasons that *A*-ing would be desirable. According to such a view, these dispositions that make intentions responsive to the desirability of their objects distinguish intentions from other motivating propositional attitudes. In this sense, aiming at to-be-doneness might be thought to be constitutive of intention.

Furthermore, the fact that intention is regulated for to-be-doneness appears to explain why the deliberative question *whether to intend to A* requires answering the question *whether to A*. If being regulated for to-be-doneness is a conceptual fact about intention,

²¹ See Velleman (2000a) for a statement and defense of this view.

²² If this were the only way in which belief aims at the truth, then the aim of belief would have been naturalized. After all, the dispositions involved in regulating a state for truth presumably are naturalistic, causal features of our psychologies; and the properties that make beliefs true are themselves natural properties, at least if we stick to beliefs with empirical content.

then anyone who deliberates about whether to intend to *A* recognizes that his deliberation can potentially produce the envisioned attitude—that is, the intention to *A*— only by regulating it in accordance with the desirability of *A*-ing. But to determine that *A*-ing would be most desirable, and therefore that one ought to *A*, just is to settle the question *whether to A*. Thus the only way to deliberate about whether to intend to *A* is to deliberate about whether to *A*.²³

I will now argue that this account of the aim of intention cannot adequately explain the way that the desirability of an action governs practical deliberation. Later I will suggest an alternative way in which the aim of intention might be incorporated into an account of the concept of intention, one that does justice to the phenomenology of practical deliberation. My argument that the above account of intention's aim cannot adequately capture the way that an action's desirability governs practical deliberation parallels the argument I have given elsewhere that the analogous account of belief's aim cannot adequately capture the way that truth governs doxastic deliberation.²⁴ I thus begin by summarizing the latter argument.

The deliberative question *whether to believe that p* can only be settled by answering the question *whether p*. This, as I have said, is a fact about the phenomenology of doxastic deliberation. But it is also a widely recognized fact that non-evidential processes occasionally influence what we believe; accusations of wishful thinking are a common enough occurrence. When wishful thinking induces one to believe that *p*, the perceived desirability of believing that *p* causes one to hold that belief. The desirability of believing that *p*, however, cannot have an acknowledged influence in doxastic deliberation, because the desirability of a belief by itself does nothing to indicate whether the content of the

²³ Notice that, unlike the analogous way of explicating the aim of belief, this way of explicating the aim of intention would not give us a naturalistic account of that aim. According to this account, the dispositions constitutive of intention would be dispositions to form, revise, and extinguish intentions in light of a norm—whether the action intended is desirable (or to-be-done). Thus, on such an account, there is no way to specify the dispositions constitutive of intention without mentioning a normative concept, desirability. Unless there is some naturalistic account of this normative concept, the aim of intention will not have been naturalized.

²⁴ See the section titled 'The Teleologist's Dilemma' in Shah (2003).

belief is true. An adequate account of belief must respect these facts. This means that whatever explanation we give of the truth-governed nature of doxastic deliberation must not generalize in such a way as to rule out the impact of other, non-evidential processes on beliefs.

The dispositional account of belief's truth-directedness cannot accomplish this task. Being regulated for truth consists in responsiveness to evidence. The belief that p tends to be formed in response to evidence of p 's truth, to be reinforced by additional evidence of it, and to be extinguished by evidence against it. A crucial question is how strong these dispositions must be in order for the attitudes possessing them to qualify as beliefs. One might interpret the truth-regulating dispositions constitutive of belief to be so strong as to rule out the influence of any non-evidential considerations on the formation of beliefs. While this might account for the fact that only evidence can play an acknowledged role in doxastic deliberation, it would also imply the false claim that non-evidential considerations could never causally influence what we believe. One might instead interpret the truth-regulating dispositions constitutive of belief to be weak enough to allow for the influence of non-evidential considerations over our beliefs, but then one would not have accounted for the fact that non-evidential factors are incapable of playing an acknowledged role in doxastic deliberation. For when one deliberates whether to believe that p , this question not only gives way to the question *whether p* but does so to the exclusion of any other, competing question, such as *whether p would be in one's interest*. Yet if belief were required only to be weakly regulated for truth, then the potential outcome of deliberation could be envisioned as a belief that p so long as p 's truth were treated as relevant to that outcome, without necessarily being treated as having absolute priority over opposing considerations: the question *whether p* would not crowd out competing, non-epistemic questions

Thus, depending upon how it is interpreted, either the dispositional interpretation of the metaphor that belief aims at truth is false, or it fails to explain why doxastic deliberation focuses solely on truth. The dispositional account might capture one of the relations that belief has to truth, but it does not capture the relation that truth has to belief

in doxastic deliberation. For similar reasons, a dispositional account of the aim of intention cannot adequately capture the relation that the desirability of *A*-ing has to intending to *A* in practical deliberation.

In order to answer the deliberative question *whether to intend to A* one must determine whether *A*-ing is desirable. But just as evidentially insensitive processes can cause beliefs, so too processes that are insensitive to the desirability of actions can cause intentions to perform those actions. While one cannot arrive at an intention via deliberation if one judges that it would be undesirable to act on it, peer pressure might nonetheless induce one into adopting an intention to do something one thinks it is undesirable to do.²⁵ One might even be able to intentionally induce a desirable intention to perform an undesirable action by various means; if one were living in more pharmacologically advanced times, taking a pill might do the trick. The correct explanation of the fact that answering the question *whether to intend to A* requires determining whether *A*-ing is desirable thus must not generalize in such a way as to rule out the influence on intentions of processes that are insensitive to the desirability of what is intended.

The dispositional account of intention's aim cannot accomplish this task. The desirability-tracking dispositions that the account claims to be constitutive of intention operate in all contexts of intention-formation. The only way that the account can explain the way that the desirability of actions govern practical deliberation is thus by requiring that the desirability of actions govern all processes that induce the intention to perform them in this way. For example, one might interpret the desirability-tracking dispositions constitutive of intention to be so strong as to rule out the influence of any other considerations on the formation of intentions. While this might account for the fact that the desirability of an action plays a dominant role in practical deliberation, it would imply the false claim that other considerations could never causally influence what we intend to do. One might instead interpret the desirability-tracking dispositions constitutive of intention to be weak enough to allow for the influence of additional considerations over

²⁵ Thanks to David Velleman for suggesting this example.

our intentions, but then one would not have accounted for the fact that considerations relevant to determining whether to perform an action dominate other considerations in practical deliberation. For when one deliberates whether to intend to *A*, this question not only gives way to the question *whether to A* but takes precedence of any other, competing question, such as *whether intending to A would be in one's interest*. Yet if intention only was required to be weakly regulated for the desirability of its object, then the potential outcome of deliberation could be envisioned as an intention to *A* so long as the desirability of *A*-ing were treated as relevant to that outcome, without necessarily being treated as having priority over opposing considerations: the question *whether to A* would not crowd out competing questions.

Either the dispositional interpretation of the aim of intention is false or it fails to explain why the deliberative question *whether to intend to A* can only be settled by answering the question *whether to A*. The dispositional account might capture one of the relations that intention has to the desirability of its object, but it does not capture the relation that they bear to one another in practical deliberation.

5. Intentions as normative judgments

Obviously, there is a very close connection between answering the question *whether to A* and intending to *A*, and answering the question *whether p* and believing that *p*. Why not just say that the relation is one of identity?²⁶ This is the other hypothesis that I will examine before turning to my own.

If intending to *A* just is answering the question *whether to A* in the affirmative, by forming a judgment that one ought to *A*, then we can easily explain why only

²⁶ Allan Gibbard (2003), for example, claims that judging that *A* is the thing to do constitutes planning to *A*. I am unsure, however, whether his position is exactly the position that I describe in the text, because I am unsure whether his notion of planning is meant to capture our ordinary notion of intention.

considerations that bear on the desirability of *A*-ing can be reasons for intending to *A*.²⁷ Considerations bearing on the desirability of intending to *A* don't bear on whether *A*-ing is desirable and thus cannot constitute reasons for the judgment that is constitutive of intending to *A*.²⁸ Similarly, if believing that *p* just is answering the question *whether p* by forming a judgment that *p*, then we have an explanation for why the desirability of believing that *p* cannot be a reason for believing that *p*. The desirability of believing that *p* does not bear on the question whether *p*, and thus cannot constitute a reason for the judgment that constitutes believing that *p*.

However, believing that *p* and intending to *A* cannot be identical to answering these questions. First, at least in the case of intention (but I also think in the case of belief), one can intend to *A* even though one thinks that one ought not to *A*. This is what happens in cases of one type of *akrasia*.²⁹ Akratic intentions are genuine intentions, not merely powerful impulses that cause one's behavior without the participation of any of one's rational capacities. One determines how to execute one's akratic intentions, and in doing so one exercises one's instrumental rationality. This is reflected by the fact that, as Jay Wallace (2001) points out, we assess how intelligently people carry out even those intentions whose ends they think they ought not to pursue: ,

...it seems undeniable that agents can display a kind of instrumental rationality in the pursuit of ends that they do not themselves endorse, when for instance they are in the grip of *akrasia*. People sometimes exhibit great intelligence and skill in executing plans that they view as dubious or questionable—think, for instance, of the extraordinary talent many of us display at procrastinating when it comes to tasks that we regard as worthy but difficult. (1)

²⁷ Pamela Hieronymi, in her excellent paper "The Wrong Kind of Reasons" (2005), claims that answering the question *whether to A* in the affirmative amounts to intending to *A*. However, she does not think that this question is identical to the question *whether one ought to A*, nor does she suggest any other question that is identical to it. As far as I can see, she thus has no explanation of why the desirability of an action is even relevant to answering the question *whether to do it*. My ideas about how to extend my own hypothesis about belief to intention emerged out of thinking about Hieronymi's paper.

²⁸ I am ignoring Buridan's Ass cases in which it does seem that the desirability of an intention can be one's reason for intending one rather than another equally desirable action.

²⁹ The other type of *akrasia* occurs when one fails to do what one intends to do. For example, one intends to take the rollercoaster when one is next at the fair, but when the time comes one is too paralyzed by fear to follow through on this intention.

The fact that our judgments about what we ought to do do not always lead us to intend to do those things does not mean that any further reasoning is required to mediate the transition from judging that we ought to *A* to intending to *A*. Normally, making a judgment about what we ought to do directly leads us to intend to do it, but sometimes this is not so. We don't always intend to do what our reasoning dictates, but this doesn't mean that there are further considerations that are relevant to determining what to intend. Nor does it mean that there is some further psychological factor present that explains why we form an intention on the basis of our judgment about what to do in non-akratic cases; all that is required is that there be an absence of interfering factors that would prevent the normal transition from judgment to intention.

Second, one can believe that *p* or intend to *A* without having asked oneself any question at all. Asking and answering questions (i.e. deliberation) are activities that we engage in sometimes when we form beliefs and intentions, but certainly not always. For example, when forming a perceptual belief one does not normally ask or answer any question at all. Usually, it is only when we have reasons to doubt the accuracy of our senses that we raise questions about whether the way that things appear to us is really the way that they are. Similarly, normally when a desire for something and a belief about how to get it leads to the formation of an intention one doesn't ask oneself whether the action intended is desirable. This question usually only arises when we experience conflicting desires pulling us in different directions. Furthermore, we attribute beliefs and intentions to others in order to explain their behavior or other mental states of theirs, without implying that the subjects of our attributions normatively endorse these states (or their objects) or are even aware of them, and therefore, without implying that these states constitute answers subjects have arrived at to questions that they asked themselves.³⁰ As theorists we thus should not introduce such an endorsement or awareness into our account of the nature of these states.

6. The aim of intention: the normative interpretation

³⁰ See Velleman (2000b) for a nice example of a case of acting on an intention that one is unaware of having.

So how to explain the fact that in order to answer the deliberative question *whether to intend to A* one must determine whether *one ought to A* in a way that does not commit us to either the false claim that only the desirability of *A*-ing can causally determine whether one intends to *A*, or that intending to *A* just is settling the question *whether to A*? Unsurprisingly, my answer is similar to the answer I have given elsewhere to the corresponding question: How to explain the fact that the deliberative question *whether to believe that p* can only be settled by determining *whether p*?³¹ I will briefly rehearse that explanation, which will give us a blueprint for our explanation of the role of the desirability of an action in practical deliberation.

Recall that the problem with the dispositional view of belief's truth-directedness is that it fails to account for the unique way that truth governs one particular belief-forming process, doxastic deliberation. In order to explain this fact, I thus think we need to look to what is special about doxastic deliberation. Once we focus on this question, one thing that jumps out is that unlike other belief-forming processes, doxastic deliberation involves the subject's application of the concept of belief. After all, one cannot deliberate about what to *believe*, which entails either explicitly asking oneself *what to believe* or at least implicitly asking this question in virtue of aiming to form a *belief*, unless one employs the concept of belief. But one can certainly form beliefs by other means (e.g. perception) without employing the concept of belief. My hypothesis is that the concept of belief contains a standard of correctness, to the effect that a belief is correct if and only if it is true. Thus, when one deliberates *whether to believe that p*, one is deliberating about an attitude for which one already applies this standard. Applying this standard constrains one to reason with an eye solely to *whether p*.

This explanation of the role of truth in doxastic deliberation, unlike the one implied by the strong dispositional interpretation of belief's truth-directedness, avoids over-generalizing the role of truth in belief-formation. While other belief-forming processes might be regulated to a certain extent by truth-seeking dispositions, only in doxastic deliberation—the belief-forming process framed either explicitly or implicitly by the

³¹ See Shah (2003).

question *whether to believe that p*—is an agent required to apply belief’s standard of correctness—truth—to the exclusion of other considerations. This explanation thus accounts for why truth exclusively governs doxastic deliberation while leaving room for the possibility that beliefs arrived at in non-deliberative ways can be influenced by non-evidential considerations.

In order to explain the way that the desirability of an action governs practical deliberation, I think we should likewise focus on what is unique about practical deliberation. Once we focus on this question, one thing that jumps out is that unlike other processes of intention-formation, practical deliberation involves the subject’s application of the concept of intention. After all, one cannot deliberate about *what to intend*, which entails either explicitly asking oneself *what to intend* or at least implicitly asking this question in virtue of aiming to form an *intention*, without employing the concept of intention. But one can certainly arrive at an intention by other means without ever asking this question or even aiming to form an intention. My hypothesis is that the concept of intention includes a standard of correctness. Just as classifying an attitude as a belief entails applying the standard of being correct if and only if its content is true, likewise classifying an attitude as an intention entails applying to it the standard of being correct only if one ought to perform the action that is its object.³² When one deliberates whether to intend to *A*, one thus is deliberating about an attitude for which one already applies this standard. Applying this standard constrains one to reason with an eye to whether *one ought to A* before one considers any other question. If one concludes that one ought to *A*, then one’s acceptance of the norm for intention, combined with this judgment, will normally lead to an intention to *A*.³³ Things can go awry, and one can fail to intend to *A*

³² This isn’t quite right. We need to allow for Buridan’s Ass cases in which it can be correct to intend to *A* even though *A*-ing is no more desirable than at least one other option. Thus we should say that an intention to *A* is correct only if it is not the case that one ought not to *A*. I will leave out this complication for ease of exposition.

³³ This explanation relies on a very weak form of internalism about normative thought. The relevant form of internalism does not require a positive disposition to obey any norm that one applies; what it requires is the lack of a disposition to obey a different norm instead. One cannot genuinely apply the norm of action to an attitude while simultaneously trying only to make it conform to some other, unrelated norm. This form of internalism does not rule out obedience to additional norms compatible with the one applied. One can aim to arrive as quickly as possible at an intention with respect to *A*-ing—in which case, one will deliberate in accordance with a norm of speed as well as the norm of action. Also, as I said in the previous footnote,

even though one has settled the question *whether A-ing is the thing to do*, but it does not follow that there is another step that one must perform between judging that *A-ing* is the thing to do and intending to *A*.³⁴

Some philosophers have claimed that deliberating about what to do involves applying a norm implicit in the concept of action.³⁵ How is this view related to my own? I want to remain agnostic about whether there is a norm implicit in the concept of action that sets the standard for correct deliberation about what to do, or whether the norm for action is to be determined by substantive normative inquiry. My claim is that deliberating what to intend cannot involve applying a different norm from that which one accepts for action, however one's acceptance of the norm for action is determined.

It may appear, though, that my hypothesis rules out at least one basic and uncontroversial norm for action. Whenever you do something intentionally, there is something you are trying or aiming to do—something you are intending to do. If you do what you aimed to do, then there is a sense in which your action is a success. You kick the ball with the aim of scoring a goal. If the ball goes through the net, your action is a success, otherwise not. Doesn't the aim, or intention, thus set the standard of correctness for action? If so, the standard of correctness for action cannot itself set the standard for correct intention—as I have claimed, on pain of vicious circularity.³⁶

I accept the platitude that an action is successful whenever it fulfills the aim with which it is done. But an action's success condition cannot be equivalent to its correctness condition. To see this, notice that we do sometimes deliberate about what to do

there may be cases in which applying the norm of action doesn't yield a unique result—*A-ing* and not *A-ing* are equally desirable. In such cases one may use another norm to break the tie and arrive at an intention.

³⁴ Beliefs and intentions are mental attitudes, whereas judgments are mental acts. When we are rational, our attitudes conform to our judgments, but when we are not they may not. That our judgments conform to our attitudes may not be something that we have control over, but it is still a failure of our rationality if our attitudes fail to conform to our judgments. After all, in arriving at a judgment that *p* we are attempting to fix a belief *whether p* by way of reasoning. If we fail to believe that *p* upon arriving the judgment that *p*, then our reasoning has failed to determine what we believe. In these sorts of cases we are apt to say that some irrational condition (i.e. an addiction or phobia) had a dominant hand in determining what we believe.

³⁵ See e.g. Davidson (1980), Velleman (2000c), and Korsgaard (ms).

³⁶ Thanks to Tom Smith for raising this objection.

antecedently to adopting an intention to do anything: we don't know what to do—we have no plan, that's why we deliberate. But whether a movement constitutes a successful action can only be determined in light of an antecedent intention to do something. This success condition thus cannot set the standard for deliberation about what to do when one has no such antecedent intention. But it is only when one has no antecedent intention, or one is reconsidering an intention that one has adopted, that one deliberates about what to intend. The answer to the question *whether to A* that settles this deliberation thus cannot be that *A-ing* would be a success.

As I have said, deliberation whether to intend to *A* or whether to believe that *p* need not be initiated by an explicit articulation of the deliberative question: one can start right in with the question *whether one ought to A* or *whether p* and yet be recognizably deliberating whether to intend to *A* or whether to believe that *p*. Reasoning directly about whether one ought to *A* constitutes practical deliberation despite skipping the question *whether to intend to A* because it is reasoning that is aimed at issuing or not issuing in an intention in accordance with the norm of intention. Similarly, reasoning directly about *whether p* constitutes doxastic deliberation despite skipping the question *whether to believe that p* because it is reasoning aimed at issuing or not issuing a belief in accordance with the norm of belief.

Unlike the dispositional interpretation of intention's aim, according to which intending to *A* is an attitude that is regulated by dispositions that track the desirability of *A-ing*, my normative interpretation of intention's aim, according to which intending to *A* is an attitude that is *correct* only if one ought to *A*, explains the role of the desirability of *A-ing* in deliberating whether to intend to *A* without falsely attributing this role to other intention-forming processes. It is not the fact that an intention to *A* is regulated by dispositions that track the desirability of *A-ing* that explains why the desirability of *A-ing* is the dominant consideration relevant to determining the intention to *A* in practical deliberation. As I argued earlier, if intentions are regulated by these dispositions, they are everywhere so regulated. Therefore, such dispositions cannot explain the unique role that the desirability of an intention's object has in one particular intention-forming process,

practical deliberation. However, if it were a conceptual truth that an intention is an attitude that is correct only if one ought to perform the action that is its object, this would explain the unique role that the desirability of *A*-ing has in practical deliberation. While other intention-forming processes might be regulated to a certain extent by dispositions that track the object of the intention's desirability, only in practical deliberation—the intention-forming process framed either explicitly or implicitly by the question *whether to intend to A*—does an agent exercise the concept of intention and thus apply the norm that is contained in this concept. It is the application of this norm in practical deliberation that is responsible for the fact that the desirability of an action must be the focus of one's deliberation about *whether to intend to do it*. Since this norm is not applied in other intention-forming processes, intentions arrived at by means of non-deliberative processes are capable of being dominantly influenced by factors unrelated to the desirability of an action. My hypothesis thus explains why the desirability of an action governs practical deliberation while allowing that intentions arrived at in non-deliberative ways can be dominantly influenced by other considerations.

Unlike the hypothesis that intending to *A* is identical to judging that one ought to *A*, my hypothesis does not have the false implication that one cannot have intentions that one is unaware of or whose object one does not normatively endorse. Not only is intending to *A* not identical to judging that one ought to *A*, but also intending to *A* needn't even be based on the judgment that one ought to *A*. Whether an intention is based upon such a judgment depends upon whether practical deliberation has had a hand in determining the intention. The real connection between intending to *A* and the desirability of *A*-ing is a conceptual normative one. It is a conceptual truth that intending to *A* is an attitude that is correct only if *A*-ing is the thing to do. This holds true for all intentions, even those that are not themselves based upon judgments about the desirability of their objects. This is why intentions that we become aware of fall under the guidance of these judgments. Recognizing an attitude of ours as an intention to *A* rather than, for example, a wish to *A*, requires us to apply intention's norm of correctness to it.

If we find that we intend to do something that we don't endorse doing, this typically undermines the intention or at least alienates us from it.³⁷

7. *The toxin puzzle*

I am now in a position to address our questions about the toxin case. Because settling the deliberative question *whether to intend to A* requires answering the question *whether to A*, the fact that you would be given a desirable amount of money for intending to take the toxin cannot influence you in practical deliberation whether to intend to take the toxin.³⁸ The offer being made to you makes it desirable to intend to take the toxin, but it does not make taking the toxin desirable, since you would receive the money for intending to take the toxin, not for taking it. Therefore the fact that you would receive this money for intending to take the toxin cannot help answer the question *whether to take the toxin* which you must answer in order to settle your practical deliberation whether to intend to take the toxin. Until you have answered the question *whether you ought to take the toxin*, the desirability of intending to take the toxin cannot play any role in your practical deliberation.³⁹ But what can this tell us about whether you have reasons to intend to take the toxin?

³⁷ This might explain why intentions whose objects we don't endorse often need to exert their influence below the level of our conscious awareness.

³⁸ I realize that not everyone accepts that one cannot intend to take the toxin on the basis of the judgment that it would be desirable to so intend. In order to give a fully satisfying explanation of the toxin case, I thus would need to give a plausible explanation of why one might mistakenly believe that it is possible to take the toxin on such a basis. I hope to pursue this task on another occasion.

³⁹ In fact, even if you were offered the money merely to desire to take the toxin you couldn't win the money. It might be thought that, unlike with the intentions, this is due to the fact that desires are not under our deliberative control. But this is not true of all of our desires. With respect to desires devoid of a phenomenological aspect, we can deliberate about what to desire. For example, upon turning the television channel to a basketball game between the Suns and Pistons—two teams that I like—I may wonder which team to root for. This, though, is really just a way of asking myself which team I want to win. After all, rooting for a team is the manifestation of a desire. I figure out which team I want to win by determining which outcome is more desirable from my point of view. If I decide that it would be more desirable, from my point of view, if the Suns win, then I normally come to desire that the Suns win. Judging that it would be desirable to desire that the Suns win because everyone around me is a Suns' fan and I want to fit in, though, will not move me to desire that the Suns win (although I might fake it so that I can fit in). Of course, peer pressure often coercively induces such desires, just not by way of deliberation. These facts suggest that just as the concept of intention includes a standard of correctness, so too the concept of (non-phenomenological) desire includes a norm of correctness. Desires may be caused by their own desirability, but their desirability does not make them correct according to the standard built into the concept of desire that is applied in deliberation about what to desire. Thanks to David Owens for suggesting this parallel. In future work I plan to investigate to what extent the type of hypothesis I have argued for with respect to belief and intention applies to desire and other psychological concepts.

According to my hypothesis, the question *whether to intend to A* gives way to the question *whether to A* because the concept of intention contains the standard of correctness: intending to *A* is correct only if *A*-ing is the thing to do. Practical deliberation, in virtue of being framed by a question that contains the concept of intention—*whether to intend to A*—is governed by this norm of correctness. Similarly, the question *whether to believe that p* gives way to the question *whether p* because the concept of belief contains the standard of correctness that believing that *p* is correct if and only if *p*. Doxastic deliberation, in virtue of being framed by a question that contains the concept of belief—*whether to believe that p*—is governed by this norm of correctness.

The conception of deliberation that emerges is this: deliberation is reasoning aimed at issuing in some result in accordance with norms for results of that kind. Deliberating whether to ϕ is reasoning aimed at issuing or not issuing in ϕ -ing, in accordance with norms for ϕ -ing. Deliberating whether to ϕ is not merely one amongst other causal routes to securing results of kind ϕ ; it is the most explicit route by which we are guided by the application of norms in reaching results of kind ϕ . Deliberation whether to ϕ thus is precisely the place to look if we want to understand what can be a reason to ϕ , and not just a mere cause of ϕ . We uncover what sorts of considerations can be reasons to ϕ by seeing which norms are applied in deliberating whether to ϕ .

The constitutive norm for intention is this: intending to *A* is correct only if *A*-ing is the thing to do. Considerations that show *A*-ing to be desirable thus are reasons for intending to *A* and reasons that show *A*-ing to be undesirable are reasons against intending to *A*.⁴⁰ The money you are offered makes intending to take the toxin desirable,

⁴⁰ Because the norm of correctness that guides practical deliberation is contained in the concept of intention, it applies to any attitude that we (correctly) conceive of as an intention. This means that the norm applies even to those intentions of ours that are not currently the subject of our practical deliberation and those intentions that we attribute to others. One cannot conceive of an attitude as an intention to *A* without committing oneself to assessing it in accordance with the standard that it is correct only if *A*-ing is the thing to do. For example, if one were to negatively evaluate someone else's intention not to take the toxin on the ground of its undesirability, one would be applying a standard to this attitude that conflicted with the standard that one was already committed to applying to the attitude by conceiving of it as an intention. My point is not that one cannot judge whether an intention is desirable, but that one cannot use this judgment as

but it does not make intending to take the toxin correct according to the norm of correctness that governs intention, because it does not indicate any way in which taking the toxin is desirable. It may be possible for the offer to rationally influence you to intend to take the toxin if you have already determined that it is desirable to take it, but according to the original description of the case you lack any such reason to think that taking the toxin is desirable.⁴¹ You may come to desire to intend to take the toxin upon being offered the money, and this desire might cause you to so intend, either directly by means of wishful thinking or indirectly by means of some intention-inducing action. In neither case, however, would the desirability of this intention have caused your intention in the way that a reason can cause an intention, because in neither case would you have been moved to intend to take the toxin on the basis of your recognition that the norm(s) for such an intention had been met. Your inability to intend to take the toxin on the ground that it is desirable to so intend thus does not reflect a rational failure on your part.^{42 43}

the normative standard for assessing its rational credentials or for impugning the rationality of someone who is unable to adopt it.

⁴¹ Might you reason like this: Should I take the toxin? Intending to take the toxin is desirable. I cannot intend to take the toxin unless I take the toxin. Therefore I ought to take the toxin. One problem with this line of reasoning is that it is not true that you cannot intend to take the toxin unless you take the toxin, and you know this. The more fundamental problem, though, is that this form of reasoning is defective. It conceives the intention to take the toxin as a means to the end of taking the toxin, and then treats the desirability of this means—intending to take the toxin—as a reason for adopting the end—taking the toxin. But the fact that pursuing a certain end requires the use of certain desirable means is not itself a reason to adopt the end. The desirability of an end gives one a reason to take the means, but the desirability of a means does not give one a reason to adopt the end to which it is a means. If, however, we were to allow that such reasoning is cogent, while we would have to admit that the desirability of intending to *A* can contribute to the desirability of *A*-ing, we would not have to deny my thesis that considerations that speak to the desirability of an action are the dominant reasons for intending the action. Furthermore, if the toxin were lethal (and one knew this), this kind of reasoning would not allow one to conclude that taking the toxin is desirable, and so it would not yield the intention to take the toxin.

⁴² If the cognitivist account of intention is correct, there may be an additional reason why you cannot intend to take the toxin. If intending to take the toxin involves believing that you will take the toxin, then only judgments that can move you to believe that you will take the toxin can move you to so intend. And the judgment that it would be desirable to intend to take the toxin cannot move one to believe that one will take the toxin, although it may imply that it would be desirable to believe it and thus move one to take steps to induce the belief.

⁴³ Earlier versions of this paper were presented to a graduate seminar at the University of Massachusetts-Amherst, a workshop on Normativity and Reasons at the University of Fribourg in Switzerland, the 2006 OSU/Maribor/Rijeka Conference in Dubrovnik on Regulating Attitudes with Reasons, and the Philosophy department of Tufts University. For very helpful discussion I would like to thank Jennifer Acker, Stephanie Beardman, Jim Bell, Macalaster Bell, Davor Bodrozic, Michael Bratman, John Broome, Fabian Dorsch, Pascal Engel, Matt Evans, Dan Farrell, Allan Gibbard, Elizabeth Harman, James Harold, Pamela

Appendix

An extensionally equivalent hypothesis?

It may be instructive to compare my treatment of the toxin case with one derived from the hypothesis that I rejected earlier on phenomenological grounds, that the deliberative question *whether to intend to A* is answered by determining whether intending to *A* is desirable. According to this hypothesis, having judged that intending to take the toxin is more desirable than not intending to take it, one would be in a position to rationally adopt an intention to take the toxin. The problem, of course, is that while one may judge that it is desirable to intend to take the toxin, one is not able to adopt this intention just on the basis of this judgment. A defender of this hypothesis could bite the bullet and claim that a fully rational agent would intend to take the toxin, and thus that those of us who cannot arrive at such an intention are irrational. But there is another option available to the defender of this hypothesis. As discussed earlier, many philosophers have suggested that one cannot intend to do something unless one believes that one will do it if one intends to do it. If this is so—and I haven't given any reason to think it isn't, it might be argued that the reason that one cannot form the intention to take the toxin is that it would be *doxastically* irrational to believe that one will take the toxin if one so intends. The defender of the hypothesis that reasons for intending are considerations that show an intention to be desirable thus need not convict us of irrationality because we are unable to adopt the intention to take the toxin. Nor need he accept that the toxin case shows that practical reasoning is about something other than the desirability of an intention. In any circumstance in which it is *doxastically* rational to form the belief that one will do something if one intends to do so (and because one so intends), he would claim that one's practical reasoning can arrive at an intention just by determining that it would be desirable to so intend.⁴⁴

Hieronymi, Don Hubin, Nadeem Hussain, Dan Jacobson, Mark Kalderon, Niko Kolodny, Uri Leibowitz, David Owens, Casey Perin, Andrew Reisner, Geoffrey Sayre-McCord, Sigrun Svavarsdottir, Seana Shiffrin, Sharon Street, David Velleman, and two anonymous referees for the *Philosophers' Imprint*.

⁴⁴ Thanks to Niko Kolodny for suggesting this hypothesis.

This version of the hypothesis that I earlier rejected may seem to be extensionally equivalent to my own, but there are in fact cases in which their verdicts diverge. This is because one might possess reasons for believing that one will *A* if one so intends that do not bear on whether *A*-ing would be worth doing. One may know that one is resolute—that one tends to carry out one’s intentions and rarely reconsiders them. Or, more fancifully, one may know that if one intends to *A* one will be causally induced to believe that one will *A* or to think that *A*-ing is worth doing. In any of these cases, it is rational to believe that one will *A* if one intends to *A*, and thus according to the alternative hypothesis now being considered it is rational to intend to *A* as long as it is desirable to so intend. But even if all these conditions are met, *A*-ing itself may no be worth doing. For example, one may have reason to believe that if one adopts the intention to take the toxin one will be hypnotized into believing that one will take it, but these reasons don’t bear on whether it is actually worth doing. According to my hypothesis, as opposed to the alternative, it may be rational to believe that one will take the toxin if one intends to do so and to believe that it is desirable to intend to take the toxin, but nonetheless it is irrational to intend to take the toxin, because one is not in a position to judge that it is desirable to take the toxin, which is the standard of correctness for deliberation about whether to intend to take the toxin.

Finally, instead of (or in addition to) claiming that one cannot intend to *A* unless one believes that one will *A* if one so intends, it might be suggested that one cannot intend to *A* unless one believes that it would be rational to execute the intention—that one ought to execute the intention *A*. According to this hypothesis, deliberating *whether to intend to A* involves answering two different questions: 1) *whether intending to A is desirable*, and 2) *whether one ought to execute the intention to A if one adopts it*. Although knowing that if one adopts the intention to take the toxin one will be induced by hypnosis to believe that one will take it makes it rational to believe that one will take the toxin if one so intends, this knowledge does not make it rational to judge that one ought to execute such an intention. According to this hypothesis, one thus cannot rationally arrive at an intention to take the toxin even under these conditions.

The claim that one cannot intend to *A* unless one believes that one ought to execute the intention if one adopts it is too strong though, because it implausibly rules out cases of akrasia in which one intends to do something that one judges one ought not to do even if one adopts the intention to do it. A smoking addict may judge that he ought not to smoke, and judge that this would be true even if he adopts the intention to smoke. Nonetheless, he may end up adopting and executing an intention to smoke. A more plausible version of the claim is that one cannot *rationally* arrive at an intention unless one judges that one ought to execute the intention if one has adopted it.⁴⁵ It might even appear that this claim can be derived from my own hypothesis.⁴⁶ After all, executing an intention, as opposed to merely adopting it, is performing the action intended. Deliberating about whether one ought to execute an intention to *A* thus would appear just to be deliberating about whether one ought to *A*, and judging that one ought to execute an intention to *A* would appear just to be judging that one ought to *A*. According to my hypothesis, deliberating whether to intend to *A* requires applying the norm contained in the concept of intention—that intending to *A* is correct only if one ought to *A*—to one’s reasoning. One thus cannot arrive at an intention to *A* by way of practical reasoning unless one at least implicitly judges that one ought to *A*. And this might seem equivalent to saying that one cannot rationally arrive at an intention to *A* unless one at least implicitly judges that one ought to execute the intention to *A*.

The question *whether one ought to execute an intention to A—whether one ought to A if one adopts an intention to A*—however is not identical to the question that frames practical deliberation, which is *whether one ought to A tout court*. One’s *actual* intentions, not hypothetical ones, help to determine the answer to this question. If intentions are reasons for action, there will be cases in which the answer to the question *whether to A tout court* will differ from the answer to the question *whether to A if one adopts the intention to A*. It may be that one ought to perform an action (in part) because one has already adopted an intention to do it, but that one wouldn’t have sufficient reason to perform it if one had not already adopted the intention. In the toxin case, it might be

⁴⁵ Cf. Farrell (1989), 283.

⁴⁶ Michael Bratman suggested something like this to me.

argued that one ought to take the toxin if one has adopted the intention to take it, even though one has no reason to take the toxin antecedently to adopting the intention. If the rationality of intending to take the toxin were determined by both the desirability of intending to take it and the desirability of taking the toxin if one adopts the intention to take it, then it would be rational to adopt the intention to take the toxin. But according to my hypothesis, if it is not the case that one ought to take the toxin before one has adopted the intention, then, even if, were one to adopt the intention to take the toxin one would have sufficient reason to take it, it is not rational to adopt the intention. While admittedly it is only in very unusual cases that my hypothesis gives a different verdict than this last alternative, their possibility is sufficient to show that they are not extensionally equivalent.

Bibliography

- Anscombe, G. E. M. 2000. *Intention*. 2nd ed. Harvard: Harvard University Press.
- Audi, Robert. 1973. 'Intending.' *Journal of Philosophy* 70, 387-403.
- Bratman, Michael. 1981. 'Intention and Means-End Reasoning.' *Philosophical Review* (90): 252-65.
- 1987. *Intentions, Plans and Practical Reason*. Cambridge, MA: Harvard University Press.
- Broome, John. 2002. 'Practical Reasoning.' in *Reason and Nature: Essays in the Theory of Rationality*, edited by José Bermúdez and Alan Millar, Oxford University Press, 85–111.
- Chisholm, Roderick. 1970. 'The Structure of Intention.' *Journal of Philosophy* (67).
- Davidson, Donald. 1980. 'How is Weakness of the Will Possible?' *Essays on Action and Events*. Oxford: Oxford University Press, 21-42.
- Davis, Wayne. 1984. 'A Causal Theory of Intending.' *American Philosophical Quarterly* (21).
- Farrell, Daniel. 1989. 'Intention, Reason, and Action.' *American Philosophical Quarterly* (26), 283-95.
- Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Grice, H.P. 1971. 'Intention and Uncertainty.' *Proceedings of the British Academy* 57, 263-79.
- Hampshire, Stuart. and Hart, H. L. A. 1958. 'Decision, Intention and Certainty.' *Mind* 67, 1-12.
- Harman, Gilbert. 1976. 'Practical Reasoning.' in A. Mele, ed., *The Philosophy of Action*. Oxford: Oxford University Press, 1997: 149-177.
- Hieronymi, Pamela. 2005. 'The Wrong Kind of Reasons.' *Journal of Philosophy* (102), 37–57.
- Kavka, Gregory. 1983. 'The Toxin Puzzle.' *Analysis* (43): 33-36.
- Korsgaard, Christine. (MS). *Self-Constitution: Action, Identity, Integrity*.
- Setiya, Kieran. 2003. 'Explaining Action.' *Philosophical Review* (112), 339-93.
- Shah, Nishi. 2003. 'How Truth Governs Belief.' *Philosophical Review* (112), 447-82.
- Shah, Nishi and Velleman, J. David. 2005. 'Doxastic Deliberation.' *Philosophical*

Review (114), 497-534.

Velleman, J. David. 1989. *Practical Reflection*. Princeton: Princeton University Press.

———2000a. ‘The Aim of Belief.’ *The Possibility of Practical Reason*. Oxford: Oxford University Press, 244-81.

———2000b. ‘What Happens When Someone Acts?’ in *The Possibility of Practical Reason*, 123-43.

———2000c. ‘Deciding to Decide.’ Reprinted in *The Possibility of Practical Reason*, 221-43.

Wallace, R. Jay. 2001. ‘Normativity, Commitment, and Instrumental Reason.’ *Philosophers’ Imprint* 1.3:1–26.