Philosophers' Imprint

VOLUME 16, NO. 14
JULY, 2016

A PREFACE PARADOX FOR INTENTION

Simon Goldstein

Rutgers University

1. Introduction

Consider the following pair of norms for intention:

Noncontradiction S ought not: intend ϕ and intend not ϕ.
Agglomeration S ought not: intend ϕ, intend ψ, and not intend ϕ and ψ.

These norms are *prima facie* plausible. Many writers accept both of them in some form or other.[1]

These norms have direct analogues for belief: don't believe contradictory things; believe the conjunction of your other beliefs. However, there is a well-known counterexample to such norms: the preface paradox.[2] Imagine a historian who writes a long book on some topic, full of carefully researched claims. The historian seems perfectly rational to concede, in the preface of her book, that at least one claim is false. So either she does not believe the conjunction of every claim in the book, or her beliefs are inconsistent. In response, many have argued that Agglomeration is not a genuine norm on belief. In its place, they suggest a series of norms governing *partial belief*.[3] A rational agent's partial beliefs must satisfy the laws of probability. In doing so, her full beliefs may fail to agglomerate.

This raises a natural question: Does the preface paradox have an analogue for intention? I will argue that there is such an analogue. There is a preface paradox for intention that shows that there is a rational agent who does not satisfy both Noncontradiction and Agglomeration. In this section, I will present two instances of the paradox. Then I will give an argument that we should expect a preface paradox for intention, given some principles that connect belief and intention.

1. For discussion, see Bratman [1984] 380, Velleman [1989], Yaffe [2004], Bratman [1999] 194, Bratman [2009], Ross [2009], Broome [2013] 76.
2. Makinson [1965]. See Ryan [1991] and Foley [1993] 143 for formulations of the preface paradox for belief that uses principles analogous to Noncontradiction and Agglomeration.
3. For representative examples, see Foley [1993] and Christensen [2005].

1.1 *Two Preface Paradoxes*

Consider the following case:

(1) Sam is making her New Year's Resolutions, deliberating about what actions to take in the next year. She writes down each intention in her journal. She spends the day painstakingly considering what to do, and ends up with 100 well-researched new intentions, to take actions ϕ_1, ..., ϕ_{100}. Each one seems like the right thing to do in the next year. Each one is independent from the others, and each is compatible with the others.

However, Sam also knows that she makes mistakes. Every once in a while, she intends the wrong thing. More precisely, she knows that she sometimes intends an action that will be extremely bad for her. Knowing this, Sam desires that at least one of ϕ_1, ..., ϕ_{100} not occur.

But her wish is not just a desire. She also intends it. In particular, suppose that some omniscient demon assures her that at least one of ϕ_1, ..., ϕ_{100} would be terrible for Sam. At least one of these actions would be so bad that performing all 100 actions would be worse than not performing all 100 actions. The demon offers Sam to thwart the action, but won't tell Sam which it is. In addition, she gives Sam a button which will dissolve the offer if pressed. Sam accepts her offer, and adds to her journal the actions: don't press the button; thwart one of ϕ_1, ..., ϕ_{100}.

That's our first example. Here is a structurally different example:

(2) Susan is planning her trip to Europe. There are 20 cathedrals she would like to visit. Each one has a fee. She really would like to see each one. And so she makes quite specific plans for each cathedral about how to get there and when to go. She looks up the cost of admission for each one. Sadly, she discovers that the total cost of admission of all the tickets is just out of her budget; she can only afford 19 cathedrals.

Yet Susan also knows that not all of her plans will come about. She knows that sometimes cathedrals close for special events. Sometimes the transit workers are on strike. In fact, she is quite sure that on one of these days, she will not be able to visit the relevant cathedral. So she decides to simply plan out each trip to each cathedral, confident that she will only need to buy 19 tickets anyways.

Let the cathedrals number 1 through 20. And let ϕ_n be the action of visiting cathedral n. Susan intends ϕ_1, intends ϕ_2, ..., and intends ϕ_{20}. However, Susan knows that it is impossible for her to perform the conjunctive action ϕ_1, ..., and ϕ_{20}. She simply doesn't have the cash. And so Susan plans to skip at least one cathedral, intending the action: not ϕ_1 or ... or not ϕ_{20}.

In this example, Susan seems perfectly rational. Yet she violates the combination of Noncontradiction and Agglomeration. For by Agglomeration she is required to intend the conjunctive action ϕ_1 and ... and ϕ_{20}. But she also intends the action not ϕ_1 or ... or not ϕ_{20}. And by Noncontradiction she cannot rationally intend both this disjunctive action and the previous conjunctive action, as they are inconsistent.[4]

We have now seen two cases in which an agent has a rational set of intentions that are incompatible with Noncontradiction and Agglomeration. In the next section, we will see that each of these cases is tied

4. This example shares some features of Bratman [1984]'s video game case. In this case, an agent plays a game in which there are two incompatible means (hitting a target) to the end of winning the game. The player takes steps to achieve both means, while realizing they are not compossible. Of this case, Bratman [1984] suggests that the agent does not actually intend to hit each target. Rather, she simply intends to *try* to hit the target.

The crucial difference between the preface case and Bratman's is that Bratman's player is choosing between only two actions, while Susan's dilemma involves 20. It is plausible that an agent does not intend small numbers of inconsistent actions. But as the number of actions increases, this becomes decreasingly plausible.

In the limit case, imagine that Susan learns that *all of her intentions together* are not compossible. In this case, Susan can still be rational to intend each one. If we try to extend Bratman [1984]'s response to the video game case, however, we reach the result that Susan doesn't intend any of her original actions; she only intends to try. Moreover, what would she do if she learned that it is impossible to *try* to perform each action?

to a norm governing belief and intention.

1.2 *Belief and Intention*

In each preface paradox, we can give an argument that the agent is rational, by providing premises that link intentions and beliefs. First, consider the following principle:

Akrasia S ought not: intend ϕ and believe ϕ is worse than not ϕ.

Consider (1) again. For each single action ϕ, Sam believes that ϕ is better than not ϕ. So Sam satisfies Akrasia when she intends ϕ. However, Sam believes that the conjunctive action (ϕ_1 and ... and ϕ_{100}) is worse than the action not (ϕ_1 and ... and ϕ_{100}). For the demon tells her that there is one action that will spoil the rest (and Sam doesn't know which). So Sam would violate Akrasia if she intended the conjunctive action (ϕ_1 and ... and ϕ_{100}). This strongly suggests that Sam is rational in (1). If we strengthen Akrasia, and add one more assumption about the case, then we reach a principle that actually entails that Sam is rational. For consider:

Strong Akrasia S ought not: believe ϕ is better than not ϕ, and not intend ϕ.

We know that for each claim in Sam's journal, she has lots of evidence that it is a good action. After all, she spent a long time researching. Now suppose we add that Sam is required to believe what is supported by her evidence. Then it turns out that Sam is required to believe of each action in her journal that it is good, but also to believe that some action in the journal is bad. That is, Sam is in an epistemic preface paradox. Given this, Strong Akrasia requires Sam to intend each atomic action ϕ, since she believes it is better than not ϕ. Yet Akrasia forbids her to intend the conjunction. So, on pain of a rational dilemma, Sam is rational in (1).[5]

5. It may not be necessary for Sam to be in an epistemic preface paradox in order to apply Strong Akrasia. For it seems logically consistent that ϕ is better than $\neg\phi$, ψ is better than $\neg\psi$, and yet $\phi \wedge \psi$ is worse than $\neg(\phi \wedge \psi)$. This may happen with organic unities. If Sam knows this, then Strong Akrasia will again place her in a practical preface paradox.

We have now seen that our first preface paradox naturally arises if the intention to ϕ is connected with the belief that ϕ-ing is good. We will now see that our second preface paradox naturally arises if the intention to ϕ is connected with the belief that ϕ-ing is possible. Consider the following principle:

Possibility S ought not: intend ϕ and believe ϕ is impossible.

It is plausible that Susan is rationally permitted to intend each individual action $\phi_1 - \phi_{20}$. In addition, she is rational in believing that the actions are jointly unsatisfiable. Possibility entails that any such agent is rationally forbidden from intending the conjunctive action (ϕ_1 and ... and ϕ_{20}). So Possibility helps explain why Susan is in a preface paradox.

Donald Davidson famously claimed that an intention to ϕ has two parts: a desire to ψ, and a belief that ϕ is a means to ψ.[6] Each of our preface paradoxes has targeted a different one of Davidson's constituents. Our first paradox arose because the belief that ϕ-ing is better than not ϕ-ing (a proxy for an agent's desire to ψ) did not agglomerate. Our second paradox arose because the belief that ϕ-ing is possible (a proxy for an agent's belief that ϕ is a means to ψ) also does not agglomerate. The result is that, on pain of irrationality, intentions themselves cannot agglomerate.[7]

6. Davidson [1963].

7. Just like in the preface paradox for belief, our new preface cases are not only counterexamples to Agglomeration, but also to a principle of Consistency forbidding an agent to intend any set of actions that are jointly inconsistent. For each of our agents intends a series of actions ϕ_1–ϕ_n while also intending the negation of their conjunction. This set of actions is jointly inconsistent, and yet rationally permitted. Note that Consistency is a stronger norm than Noncontradiction, which merely requires that no *two* intentions of an agent be inconsistent. Since our agents do not intend the conjunction of their actions, they satisfy Noncontradiction without satisfying Consistency. Thanks to an anonymous referee.

1.3 Probabilism About Intention

I have argued that Agglomeration is not a norm on intentions. But what are the norms governing intentions? In the rest of this paper, I will pursue a hypothesis. Intentions come in degrees. There is a state of "partial" intention that stands to "full" intention as partial belief stands to full belief. One important set of norms for intention governs these partial intentions. These norms require that one's partial intentions satisfy the probability calculus.[8] [9] An agent intends to ϕ just in case her partial intention to ϕ is sufficiently strong. Call this theory "probabilism about intention".[10]

This hypothesis explains our preface paradoxes. In each case, our agent intends a series of actions to a certain degree. But since our agent's partial intentions are a probability function, the agent will never intend a conjunction of intentions to a higher degree than she intends the minimum conjunct; sometimes she will intend the conjunction less. As the number of conjuncts increase, her degree of intention in the conjunction will tend to lower. In both preface cases, the agent's degree of intention in the conjunction of each individual act is so low that she intends the negation of that conjunction to a very high degree. And so she intends each individual action, and also intends that one of the actions not occur.[11]

1.4 Outline

In the rest of this paper, I will explore and defend probabilism about intention. In the next section, I will present a positive theory of partial intention. Then I will use this theory to give an argument for probabilism about intentions. However, this argument relies on the particular theory of partial intention that I suggest. In the subsequent section, I offer a more general argument for probabilism about partial intention. I develop an analogue for intention of some decision-theoretic arguments that credence should obey probabilism. The result is a new type of decision theory governing an agent's degrees of intentions. These decision-

8. Many have endorsed a degreed solution to the preface paradox for belief. For example, Foley [1993] accepts a Lockean principle connecting full and partial belief, and rejects consistency and closure norms for full belief. Christensen [2005] defends full-blown probabilism about belief in response to the preface. And Stalnaker [1987] even suggests the stronger claim that probabilism provides the *only* norms for belief: "Once a subjective or epistemic probability value is assigned to a proposition, there is nothing more to be said about its epistemic status." (Stalnaker [1987] 81).

9. Holton [2008] also defends the view that there is a state of partial intention. However, Holton [2008]'s theory of partial intention does not resolve the preface paradox. For these partial intentions do not come in degrees. On this proposal, an agent's intention is partial just in case it is one of several incompatible plans for achieving an end. But neither of our examples has this structure. There is no common end that all of Sam and Susan's actions achieve. And even if there were, Sam's and Susan's actions are not incompatible. Finally, if the theory did not require that the relevant ends are incompatible, it still would not explain why Sam and Susan are under pressure to intend conjunctions of their intentions less than each conjunct.

10. What are the objects of intention? One option is that when S intends to ϕ, the object of her attitude is the proposition that S ϕs. Alternatively, the object of her attitude might be the property of ϕ-ing.

In this paper, I will not need to settle whether the objects of intention are propositions, properties, or something else. All I will assume is that the objects of intention form an algebra. That is, they are closed under operations of union, intersection, and complementation. This requirement is vindicated by either the propositional or property view.

11. Recently, Shpall [2016] has independently discovered a close cousin to the second preface paradox. In Shpall's case, an agent intends a series of actions while believing they are not jointly satisfiable. Shpall observes that this is a counterexample to the conjunction of two norms: (i) it is irrational to intend to ϕ while believing one will not ϕ; (ii) if an agent is rationally permitted to intend to ϕ and to intend to ψ, then she is rationally permitted to intend to $\phi \wedge \psi$.

While quite similar to the second preface paradox in this paper, Shpall's approach differs in a few important ways from this paper's. First, Shpall's agent does not actually intend that the conjunction of individual actions not occur. This is important, for one of Shpall's two responses to the problem is to give up (i). But this response does not help with our strengthened case, for the agent will still violate Noncontradiction if she satisfies either Agglomeration or (ii).

Shpall also offers an account of intention on which it comes in degrees. Shpall does not provide a metaphysical reduction of these "inclinations", and so the dispositional account of partial intentions developed here may be compatible with his own account. However, Shpall does not offer arguments that a rational agent conforms these degrees to the probability calculus. Such an argument is needed in order to use degrees of intention to explain the preface case. Such arguments will be provided later in this paper.

theoretic arguments require surprisingly few assumptions about the exact nature of partial intention.

2. A Dispositional Theory of Partial Intention

I've suggested that partial intentions can explain the behavior of Sam and Susan in the preface paradoxes. But what are partial intentions? In this section, I will propose a reductive theory of partial intention. Then, in the next section, I will show that this theory entails that an agent is irrational if her partial intentions are not probabilities.

I'll now pursue a dispositional account of partial intentions. The degree to which an agent intends to ϕ is simply the degree to which the agent possesses the dispositions characteristic of fully intending to ϕ. First, I'll sketch how the theory of dispositions in Manley and Wasserman [2008] allows them to come in degrees. With this sketch in place, I'll show how Bratman [1999]'s dispositional account of full intentions can be extended to partial intentions.

2.1 Partial Dispositions

Manley and Wasserman [2008] defend a theory of dispositions on which they come in degrees. The key motivation for their proposal is to explain the felicity of *comparative dispositional ascriptions*. Consider sentences like the following:

(3) Glass A is more fragile than Glass B.
(4) Glass A is more disposed to break than Glass B.[12]

These kinds of ascriptions suggest that dispositions come in degrees. Manley and Wasserman [2008] explain this by appeal to *proportions of cases*. In particular, they propose:

Prop N is disposed to M when C iff N would M in some suitable proportion of C-cases.

More N is more disposed than N′ to M when C iff N would M in more C-cases than N′.

What is a C-case? It is a set of worlds that specify a bunch of conditions relevant for a disposition. For example, a C-case for a glass being disposed to shatter when dropped is a set of worlds with the same height of the glass, gravitational constants, density of air, etc. Context will restrict the range of worlds included in the C-cases.

For our purposes, we will need to simplify this definition a bit. First, we will remove relativization to a circumstances parameter. The dispositions we are exploring simply involve an agent performing a behavior, not performing a behavior in a special circumstance. Second, we will need to evaluate all our dispositions relative to a *common* body of cases. So each disposition will occur at some proportion of a *common* body of cases.[13]

Next, we need to generalize More from comparisons on N to comparisons on M. For example, one could compare the degree a glass would shatter if dropped to the degree it would shatter into more than 100 pieces if dropped. The glass is disposed to shatter to a greater degree than it is disposed to shatter into more than 100 pieces. This is because more dropped-cases are shatter-cases than are shatter-into-100+-cases.

More gives us an ordering on various dispositions. This ordering can be mapped into the real numbers from 0 to 1 by assigning each disposition its proportion of C-cases. The degree to which N is disposed to M in C is exactly the proportion of C-cases in which N Ms.

We can use this ordering on dispositions to give a theory of partial intention. The degree of an intention is the strength of the dispositions that characterize intention. In the next section, I will draw on work by Michael Bratman to find these dispositions. Putting these together, we will have a theory of partial intention.

12. Manley and Wasserman [2008] 71.

13. Thanks to an anonymous referee here.

2.2 *Dispositions and Intention*

Michael Bratman has characterized intentions by a number of dispositions to act and reason in certain ways. Here is a summary from Bratman of the two kinds of dispositions involved in intentions:

> The descriptive aspect of the volitional dimension of commitment consists in the characteristic role of present-directed intention in controlling (and not merely potentially influencing) present conduct. **If I intend to A now, my intention will normally lead me at least to try to A.** ... The descriptive aspect of the reasoning-centered dimension of commitment, in contrast, consists in the characteristic roles of future-directed intentions in the interim between their acquisition and their execution. These roles include both their characteristic persistence and their part in guiding further practical reasoning, reasoning that issues in derivative intentions. Future-directed **intentions resist (to some extent) revision and reconsideration**. And future-directed **intentions involve dispositions to reason in appropriate ways: to reason about means**, preliminary steps, or just more specific courses of action; and to constrain one's intentions in the direction of consistency.[14]

Bratman suggests that an intention is a complicated dispositional state. This state involves both dispositions to act and dispositions to reason. Let's focus on the three emphasized parts. We have three dispositions:

Act If S intends to ϕ, then S is disposed to ϕ.
Don't Revise If S intends to ϕ, then S is disposed not to revise and reconsider whether to ϕ.
Search for Means If S intends to ϕ, then S is disposed to search for means to ϕ.

14. Bratman [1999] 108-109.

Before we go on, a disclaimer: these three dispositions may not be the *only* ones constituting intention. But they are certainly important components. This paper is about partial intention, not full intention. So I have to bracket the question of what *exactly* full intention is. For the rest of this section, I will focus on these three dispositions as a case study. I will show that if partial intention were constituted by just these three dispositions, then we could give an elegant theory of partial intention and also an argument for probabilism about intention. Since these three dispositions are not all there is to intention, the theory to come is somewhat incomplete. But what follows will highlight the general form that a theory of partial intention could take, and gives a recipe for arguments justifying probabilism about intention. By contrast, the decision-theoretic arguments in the second half of the paper will bypass questions about the exact dispositional nature of intention.

2.3 *A Theory of Partial Intention*

We can now use the previous two sections to give a theory of partial intention. The degree to which an agent intends to ϕ is simply a weighted sum of the dispositions above:

Partial Intention S intends to ϕ to degree n iff n = the weighted sum of the degree to which S is disposed to search for means to ϕ, to not revise or reconsider whether to ϕ, and to ϕ.

Partial Intention assigns each of three dispositions a certain weight. What weight is that? We won't need to answer that question for our purposes. But any answer is probably vague and context-sensitive.

Now we can connect partial intentions and full intentions, since we accept Prop, where an agent is disposed to M *simpliciter* iff she has a sufficiently high degree of disposition to M. In our context, this generates a Lockean theory of full intentions:

Lockean Intention S fully intends to ϕ iff S's degree of intention to ϕ is sufficiently high.

We now have a unified theory of intentions. An agent intends to ϕ just in case her degree of intention to ϕ is sufficiently high. An agent's degree of intention is the weighted sum of the degree to which she possesses the dispositions characteristic of intending. In the next section, we will use this theory of intention to give an argument that degrees of intentions must be probabilities.

3. An Argument for Probabilism About Intentions

I've given a theory of partial intentions. Now let's put it to work to get an argument for probabilism about intentions. Here's the structure of the argument: on the current theory, S's degree of intention to ϕ is simply a weighted sum of three dispositions. But the weighted sum of a series of probability functions is itself a probability function. So I will now give a series of arguments that an agent is irrational if any one of these three dispositions is not a probability function. This will entail that a rational agent's degree of intention is a weighted sum of three probability functions. And this means that her degrees of intention are themselves a probability function.[15]

A probability function is any function from an algebra of claims into the real numbers from 0 to 1 that satisfies the following three axioms:

Non-Negativity It is irrational to intend any action to a degree less than 0.

Normality It is irrational to intend any tautology to a degree less than 1.

Additivity If ϕ and ψ are mutually exclusive, then it is irrational to intend (ϕ or ψ) to a degree less than or greater than the sum of the degree to which one intends ϕ and the degree to which one intends ψ.[16]

In the rest of this section, I will argue that each of our three dispositions must, on pain of irrationality, satisfy each of the three axioms.

But our job isn't quite that complicated. It turns out that Non-Negativity comes for free. On our framework, the degree of a disposition is the proportion of worlds where the disposition manifests. And "proportions of worlds" obey Non-Negativity; no proportion is less than 0. And so we only have to check Normality and Additivity.

A word of warning: there will be two types of arguments to come. Some will show that it is metaphysically impossible for some disposition to violate some axioms. (Non-Negativity worked like this.) Others show that it is irrational for a disposition to violate an axiom. Together, these arguments support the claim that any agent whose intentions are not probabilistic is irrational.

3.1 Normality

Let's start with Normality. First, consider Act. S's degree of intention to ϕ is constituted in part by her degree of disposition to ϕ. Suppose ϕ is a logical truth.[17] Then ϕ is true in every world. So an agent manifests a disposition to ϕ at every world.

Now let's consider Search for Means. Any action is a sufficient

15. As an anonymous reviewer observed, this form of argument places a restriction on the degree of context-sensitivity involved in the weights assigned to each disposition. In any given context, the weights assigned to a disposition must be the same for each action. For suppose not, and imagine that the weight assigned to the disposition to act was 1 for ϕ and 0 for $\phi \vee \psi$. Then the degree an agent intends ϕ could be 1 while the degree she intends $\phi \vee \psi$ could be 0, violating Probabilism.

16. ϕ and ψ are mutually exclusive just in case $\{\phi, \psi\} \models \bot$, where $\models$ is a relation of logical entailment definable using the algebra on which the probability function is defined. For more on what exactly this algebra is, see the next footnote.

17. Here it might seem that I am assuming that the objects of intention are propositions. For suppose that when S intends to ϕ, the object of her intention is a property. What would it mean for a property to be a logical truth?

We can say that a property is a logical truth when every object is guaranteed to possess it. For example, the property of *going to the store or not going to the store* is possessed by every object. This conception of "logically true" properties will vindicate Normality. For if ϕ is a property possessed by every object, then every agent trivially will perform ϕ, and will perform the means to ϕ. And if ϕ is guaranteed to be instantiated, then it is a waste to reconsider whether to instantiate ϕ.

means for a logical truth; so an agent trivially performs the means to any logical truth at any world; so an agent searches for means to the logical truths to degree 1.

Finally, consider Don't Revise. This one is a bit trickier. Here, note that any reconsideration of whether to perform a logical truth is in some sense a waste of time, since the action is guaranteed to be performed regardless of what one decides. Resources are better spent deliberating on actions that are not guaranteed to come true.

One might be skeptical of these arguments. After all, in ordinary life we never describe agents as intending to (go to the store or not go to the store), or intending to be such that $2 + 2 = 4$. But there is a simple explanation of this fact. On our account, it is extremely easy to intend these actions. And so every ordinary agent that we encounter already intends the actions. And so it would be completely uninformative to describe an agent as intending such an action. And so Gricean reasoning predicts that it would be strange for a speaker to point out that an agent intends such an action.[18] After all, consider a related question: does any ordinary agent intend to any degree that a logical truth *not* be true?[19]

18. Grice [1975/1989a].

19. This raises some more general worries (thanks to an anonymous referee here). First, what are the objects of intention? Are they propositions or actions? Second, what kinds of events are intended? For example, can events in the past be intended?

For probabilism about intentions to hold, all we need to assume is that there is some algebra of events on which degrees of intention are defined. This algebra can be built on either propositions or actions. But this algebra does not need to include every possible event. For example, events in the past can be excluded from the algebra (whether represented as propositions or actions).

Nonetheless, to have an algebra we need the assumption that whenever ϕ and ψ are assigned a degree of intention, so are $\neg\phi$, $\phi \wedge \psi$, and $\phi \vee \psi$. So one potential concern for probabilism about intention would be a case where some ϕ and ψ are intuitively actions, but $\neg\phi$, $\phi \wedge \psi$, or $\phi \vee \psi$ are intuitively not actions. For example, GOING TO THE STORE is intuitively an action, while GOING TO THE STORE OR NOT GOING TO THE STORE is not an action.

Here are three ways to avoid this concern: First, one might resort to the Gricean strategy discussed in the main text. $\phi \vee \neg\phi$ is an action, but is a weird action to talk about someone intending. Second, $\phi \vee \neg\phi$ is not a *mysterious* action on the dispositional theory. For S to intend $\phi \vee \neg\phi$ involves S being such that $\phi \vee \neg\phi$ in sufficiently many cases. In addition, it involves S searching for means to being such that $\phi \vee \neg\phi$, and holding fixed being such that $\phi \vee \neg\phi$ in deliberation.

But if these responses fail, there is a more concessive alternative. Say that the dispositional theory and probabilism govern *proto-intentions* in *proto-actions*. Whenever ϕ is a proto-action, so is $\phi \vee \neg\phi$. Then we can use proto-intentions to give the norms on intention. Say that S ought to intend to ϕ to degree n iff S ought to proto-intend to ϕ to degree n and ϕ is an action. This allows us to retain our systematic explanation of the preface cases.

20. This premise is false if means are necessary rather than sufficient means. For more on this, see §6.5.

3.2 *Additivity*

In the case of Additivity, all of our arguments will have a similar structure. In each case, we have some dispositions associated with intending some incompatible actions ϕ and ψ. We will first suppose that these dispositions are manifested in some range of cases, for each ϕ and ψ. Then we will prove that, on pain of irrationality, the number of cases in which the disposition is manifested for the disjunctive action $\phi \vee \psi$ is equal to the sum of the two previous numbers of cases.

In the case of Act, the disposition to ϕ if one intends to ϕ, the argument goes as follows:

1. Suppose S performs ϕ in n cases and performs ψ in m cases.
2. Suppose ϕ and ψ are mutually exclusive.
   [Show: the number of cases in which S performs $\phi \vee \psi$ is n+m]
3. None of n and m overlap, from 2.
4. Any n case and any m case is a case of $\phi \vee \psi$.
5. Any case of $\phi \vee \psi$ is either an n case or an m case.
6. Therefore, the number of cases where S performs $\phi \vee \psi$ is n+m.

In the case of Search for Means, the argument is slightly different:

1. Suppose S searches for means to ϕ in n cases and searches for means to ψ in m cases.
2. Suppose ϕ and ψ are mutually exclusive.
   [Show: the number of cases of searching for means to $\phi \vee \psi$ is n+m]
3. Any means for ϕ and any means for ψ is a means for $\phi \vee \psi$.[20]

4. Any means for $\phi \vee \psi$ is either a means for ϕ or a means for ψ.
5. From 3 and 4, if no n case is an m case, then the number of cases of searching for means to $\phi \vee \psi$ = n+m.
6. If S is rational, then no n case is an m case.
7. Therefore, from 5 and 6, if S is rational then the number of cases of searching for means to $\phi \vee \psi$ = n+m.

To finish the argument, we just have to prove (6). Whether (6) is plausible depends on what "searching for means" means. Here's one gloss on searching for means:

Search For Means' If S intends ϕ, then S is disposed to be such that there is some ψ such that S believes that ψ is sufficient for ϕ, and S intends ψ.

The idea behind SM' is that searching for means to an action is just intending a (believed) sufficient means for it. With this reading, we can justify (6). If there is an n and m case in common, this would then be a case where an agent intends a sufficient means to ϕ and also intends a sufficient means to ψ. But this is irrational because it violates Noncontradiction.

Finally, let's turn to the argument that Don't Revise satisfies Additivity:

1. Suppose S does not reconsider ϕ in n cases and does not reconsider ψ in m.
2. Suppose ϕ and ψ are mutually exclusive.
   [Show: the number of cases of not reconsidering $\phi \vee \psi$ is n+m]
3. Suppose there is an n (/m) case where S reconsiders $\phi \vee \psi$.
4. By 1, S would be reconsidering an action that his other commitments entail.
5. It is irrational to reconsider an action entailed by what you do not reconsider.
6. From 3–5, if no n case is an m case, then the number of cases of not reconsidering $\phi \vee \psi$ = n+m.
7. Suppose that some n case is an m case.
8. By 2, S will not reconsider two intentions that cannot both be satisfied.
9. By Noncontradiction, it is irrational to be committed to two actions that can't both be performed.
10. By 7–9, if S is rational, no n case is an m case.
11. Therefore, from 6 and 10, if S is rational then the number of cases of not reconsidering $\phi \vee \psi$ = n+m.

This argument is similar to the argument about searching for means. Both invoke Noncontradiction. One might worry that this is dialectically inappropriate. We wanted coherence constraints on partial intention to explain the rational requirements on intentions. But I've had to appeal to the Noncontradiction norm I started with.

I have two responses. First, I have made no appeal to Agglomeration, and have argued that it does not hold. So we still have an overall theory that explains the preface paradox. Second, one might at this point distinguish (dispositional) intentions from (occurrent) premises in practical reasoning. A premise in practical reasoning is an occurrent token of a step of a mental process. Perhaps it is a tokened sentence in mentalese. This is picked out by the term 'commitment' in premise 4. By contrast, an intention is a disposition to act in certain ways, and to token certain premises in practical reasoning. On the current account, facts about our actual intentions are explained by facts at various nearby possible worlds about the distribution of occurrences of tokens of premises about means, and the lack of revision of these tokens. I appeal to Noncontradiction only when it comes to premises in practical reasoning. But the norms directly governing intentions are the axioms of probability.

The first response is still important because it prevents a preface paradox for tokenings of premises. After all, Sam can consider all 101 actions in a single bout of practical reasoning. But, crucially, no two premises in Sam's reasoning are inconsistent. So we have used a weak consistency norm on tokenings of premises in practical reasoning, plus facts about the nature of dispositions, to argue for some stronger co-

herence norms on partial intention.

3.3 Limits

This completes our first argument for probabilism about intention. We've seen that each disposition constitutive of partial intention must be a probability function. Partial intention is a weighted sum of such probabilistic dispositions. And the weighted sum of several probability functions is itself a probability function.

This argument has limits. For example, we already saw that the argument took Noncontradiction as primitive. One might wonder whether there is a way to derive this norm instead. But more importantly, this argument depends on the particular theory of partial intention that I defended. It may be plausible that partial intentions are in some way a matter of the dispositions characteristic of full intention. But it is unclear whether, at the end of the day, the exact dispositions involved are Act, Don't Revise, and Search for Means.

In fact, there seem to be cases where these dispositions are not exactly what we want. For example, it seems like an agent can have an extremely high degree of intention to perform an act ϕ that is extremely difficult. They may only ϕ in a small proportion of worlds. And so if Act is an important component of partial intention, such an agent will not have an extremely strong partial intention to perform the act.[21]

The attraction of the framework above is that it gives us a recipe for arguments in favor of probabilism. But the argument is not conclusive, since it awaits a complete theory of the dispositions that characterize full intentions. This raises a question: Is there an alternative way to defend probabilism about intention that does not rely on a particular theory of partial intention? In the rest of the paper, I will show that this can be done using some tools from decision theory.

21. In response to this worry, one might revise Act so that it involves trying to ϕ, rather than ϕing. But this will complicate premise 3 in the argument for Additivity.

4. A Decision-Theoretic Argument for Probabilism

In this section, I will give an argument that an agent's degrees of intention should obey the probability calculus. I will argue for this claim by extending Jim Joyce's arguments that degrees of belief should be probabilities.[22]

Joyce gives an "epistemic utility" argument for probabilism. He treats the question of what degrees of belief to have as a decision problem. Having a credence function provides the agent with a different level of value in different possible worlds. The rational credence function for an agent is the one that best balances the value of that credence function in each possible world.

Joyce's argument proceeds in three steps.[23] First, one needs a way of calculating how valuable a credence function is at a possible world. Second, one needs a decision rule that says what credence functions to have, given how valuable the credence functions are at each world. Finally, one shows that credence functions that satisfy the probability calculus are rationally required, given the decision rule and the theory of value.

I will provide an analogous argument that an agent's degrees of intentions must satisfy the probability calculus. To do so, I will construct analogues of Joyce's theory of value and of decision rules. Here is the result I will establish:

Dominance If an agent's degrees of intention are not a probability function, then there are some other potential degrees of intention that more closely match any candidate for the best of all possible worlds.
If an agent's degrees of intention are a probability function, then there are no other potential degrees of intention that more closely match any candidate for the best of all possible worlds.

22. See Joyce [1998]; Joyce [2009]; Pettigrew [2013]; De Finetti [1974].
23. Pettigrew [2013] 899.

4.1 Assessing the Value of Intention

Joyce's epistemic argument begins with the claim that the value of a credence function is its accuracy. On this picture, we can assess the value of a credence function relative to different possible worlds. Relative to a possible world, the value of the credence function is its accuracy. That is, the value of the credence function is simply the degree to which that credence function matches the world.

More precisely, Joyce defines the value of a credence function at a world in two steps. First, he says which credence function c is *most valuable* at world w. This is the credence function that perfectly matches the world. In this case, we say that c is *vindicated* by w. Second, he defines a distance measure between the most valuable credence function and every other credence function. The value of a credence function at c is a function of its distance from the most valuable (accurate) credence function.

To extend Joyce's argument from belief to intentions, we must first determine what makes a partial intention function valuable. Here is my proposal: Just as credence aims at the truth, intention aims at the good. Just as the best credence function is the one that perfectly matches the actual world, the best intention function is the one that perfectly matches the best world.

Joyce provides a decision-theoretic model for the intuitive claim that belief aims at truth. We need an analogous model of how intention aims at the good. I suggest the following: Any theory of the good induces an ordering on possible worlds. This ordering will generate a set of worlds that are best — ranked highest in the ordering.[24] We can then assess the value of an intention against any of these "best of all possible worlds".

I propose that we assess intentions for value relative to some candidate for one of the best of all possible worlds. Let I(·) be a function that represents the agent's degrees of intention. Let g be a candidate for the best of all possible worlds. Call g a "goal". We can define the value of I relative to g in two steps. First, we find the intention function that is best relative to g. Second, we measure the distance between I and this "vindicated" intention function.

But first, a disclaimer. Here I assume that we assess intentions against candidates for the best of all possible worlds. But my argument for probabilism will not need to assume any particular conception of the good. The argument will be that whichever possible worlds are best, a probabilistic intention function does better by that standard than a probabilistically incoherent intention function. In this sense, our argument is "procedural" rather than "substantive". The good may simply consist in the satisfaction of an agent's desires. Or the good might be something more objective. Our argument will show that whatever the good is, probabilistic intentions do better by it than incoherent intentions.

4.2 Vindication

In this section, we will explore the two steps required to determine the value of an intention function I relative to some best possible goal g. First, we specify the vindicated intention function for each goal g. Second, we measure the distance between any intention function and the vindicated function.

The vindicated intention function relative to g assigns the degrees of intentions that are best, assuming that g is the best of all possible worlds. Here is a natural proposal: the best intentions to have, given goal g, assign a degree of 1 to g and a degree of 0 to any other goal. This way, the intentions perfectly match the goal. More precisely, let v_g be the degree of intention function vindicated by g. And let a goal g be a maximal, consistent set of actions ϕ. Then we say:

Definition 4.1 (Vindication). $v_g(\phi) = \begin{cases} 1 \text{ if } \phi \in g \\ 0 \text{ if } \phi \notin g \end{cases}$

24. See chapter 5 of Lewis [1973] for how to extend these orderings to cases where there is no best world.

This definition is exactly analogous to Joyce's definition of the vindicated credence function at a world.[25]

One might challenge this "matching" conception of vindication. For example, one might think that the best intention function relative to g is the one with the highest chance of actually bringing g about. However, I think there is a good reason to prefer a matching conception of vindication to a causal conception.

Consider Kavka's toxin case:

> An eccentric billionaire places before you a vial of toxin that, if you drink it, will make you painfully ill for a day, but will not threaten your life or have any lasting effects. The billionaire will pay you one million dollars tomorrow morning if, at midnight tonight, you intend to drink the toxin tomorrow afternoon. He emphasizes that you need not drink the toxin to receive the money; in fact, the money will already be in your bank account hours before the time for drinking it arrives, if you succeed. All you have to do is . . . intend at midnight tonight to drink the stuff tomorrow afternoon. You are perfectly free to change your mind after receiving the money and not drink the toxin.[26]

Intuitively, it is irrational or incoherent for you to intend to drink the toxin. What explains this?

In this case, the matching and causal criteria for vindication make different predictions. Here, your goal is to avoid drinking the toxin, but get the million dollars if possible. On the matching conception, the best intention function given that goal thus intends not to drink the toxin, and intends to get the million dollars if possible. By contrast, on the causal conception, the best intention function is the one that intends to drink the toxin, since this will cause you to get the million dollars. The matching conception, but not the causal conception, can therefore explain why you would be irrational to intend to drink the toxin.

With our vindicated intention function in place, we can now define the distance in value between any two intention functions. The most popular distance function in the case of credence is the *Brier score*. The Brier score sums the squares of the differences in degree of intention between the two functions, for each action ϕ. More precisely (letting Φ be the set of acts):

Definition 4.2 (Distance). $d(v_g, I) = \sum_{\phi \in \Phi} |v_g(\phi) - I(\phi)|^2$

This definition of distance is exactly analogous to Joyce's definition of distance for credence functions.[27] In fact, the results that follow will hold for a larger class of distance measures — those that are *strictly proper*.[28] For concreteness, I focus on the Brier score. We now know which intention function is best for a given goal. And we know how

25. $v_w(p) = \begin{cases} 1 \text{ if } p \in w \\ 0 \text{ if } p \notin w. \end{cases}$

26. Kavka [1983], 33–34.

27. Pettigrew [2013] 899.

28. One of the main arguments for strict propriety generalizes smoothly to intentions. A scoring rule is strictly proper whenever it generates the verdict that any probability function is the unique function that minimizes expected score, relative to itself. Joyce [2009] observes that this property follows from two more properties–*immodesty* and *minimal coherence*. A scoring rule is immodest whenever it generates the verdict that any function uniquely minimizes expected score, relative to itself, if that function could be rational. A scoring rule is minimally coherent when it says that each probabilistic intention function is uniquely rational in some situation. Together, immodesty and minimal coherence entail strict propriety. Thus, we must provide an argument that degrees of intention satisfy immodesty and minimal coherence.

The requirement of *immodesty* looks plausible for intentions. Here, we can require a distance measure on which every partial intention function maximizes the expectation of value-at-goal, weighted by the degree to which each goal is intended. Now here's an argument for *minimal coherence*. Suppose the following akrasia norm holds: if an agent is certain that she ought to intend ϕ to degree n, then she is required to intend ϕ to degree n. Suppose further that we follow Joyce [2009] in allowing that credences satisfy minimal coherence. Then we can construct an evidential situation in which an agent is required to be certain that she ought to adopt intention function I. Given our akrasia norm, this entails that there is an evidential situation in which she ought to adopt intention function I. Thanks to an anonymous referee for help here.

to calculate the distance between intention functions. We can put these two notions together to calculate the value of every intention function at every goal. The utility of an intention function at a goal is simply the function's distance from that goal's vindicated intention function. More precisely, let B measure the utility of an intention function relative to a goal. It is a function from intention functions and goals to the real numbers:

Definition 4.3 (Utility)**.**
$B(I, g) = 1 - d(v_g, I) = 1 - \sum_{\phi \in \Phi} |v_g(\phi) - I(\phi)|^2$.

This definition of utility is exactly analogous to Joyce's definition of utility for credence functions.[29] The definition of utility calculates how valuable any intention function is relative to any goal.

4.3 Dominance

We have calculated the utility of the degrees of intention of an agent relative to a particular goal. We will now give a decision rule that deems certain degrees of intention irrational as a function of their utility relative to goals. Our decision rule uses the concept of dominance. One intention function I* dominates another, I, iff I* does better than I for every goal. More precisely, this holds iff I* has a greater utility to I relative to every goal:

Definition 4.4 (Dominance)**.** I* dominates I iff for every goal g, B(I, g) $<$ B(I*, g).

With this definition in place, we can give a decision rule: It is irrational to have a set of degrees of intention that are dominated. If your degrees of intention do worse than some other degrees relative to every candidate for the best possible world, then your degrees of intention are irrational. (This rule only holds when the dominating degrees of intention are themselves not dominated. If every set of degrees of intentions is dominated, then all bets are off.) More precisely:

Dominance Rule If (a) there is some I* such that I* dominates I; and (b) there is no I** such that I** dominates I*, then an agent is irrational if her degrees of intention match I.

4.4 Result

We have now provided a theory of the utility of I relative to a goal. And we've provided a sufficient condition for I being irrational. We can use these points to show that an agent is irrational if her degrees of intention are not a probability calculus:

Theorem 4.1 (Modified De Finetti)**.** (a) If I is not a probability function, then there is some I* that is a probability function such that I* dominates I. (b) If I is a probability function, then there is no I* such that I* dominates I.[30]

Together, Theorem 4.1 and Dominance Rule entail that an agent is irrational if her partial intentions are not a probability function.

We've now finished our argument that an agent is irrational if her degrees of intention are not a probability calculus. If this were to happen, then her degrees of intention would do a worse job relative to any candidate for the best possible world. This can't be rational, for it violates the commonplace thought that intentions aim at the good. In the next section, we will give a similar decision-theoretic argument for a Lockean thesis for intention.

5. A Decision-Theoretic Argument for the Lockean Thesis

We have now seen that an agent is irrational if her degrees of intentions are not a probability function. We can now explore the relationship between degrees of intention and full intentions. Many Bayesians endorse a Lockean theory of the relation between credence and belief.

29. Compare Pettigrew [2013] 900.

30. For proofs of the analogous theorem for credences, see De Finetti [1974] 87–91; Joyce [1998]; Pettigrew [2013] 907.

On this proposal, an agent believes p iff her credence in p is above some threshold.

In this section, I will extend recent work by Kenny Easwaran to provide a decision-theoretic argument for the Lockean thesis for intention.[31] That principle says:

Lockean Intention S fully intends to ϕ iff S's degree of intention to ϕ is sufficiently high.

Easwaran gives an epistemic utility argument that agents are irrational if their doxastic states do not satisfy the Lockean thesis. Easwaran's argument can be extended to intentions, to show that an agent is irrational if she violates Lockean Intention. The basic idea will be this: If an agent violates Lockean Intention, then her full intentions will violate a rational norm. The agent's full intentions will not be her best attempt at satisfying her goals, weighted by how strongly she intends each goal.

I will proceed in 3 steps. In step 1, I define the value of a set of full intentions *relative to a goal*. This is a function of two factors: first, how many intentions *agree with the goal*; second, how many intentions *disagree with the goal*. The value of a set of full intentions is a weighted sum of these two factors.

In step 2, I offer a decision-theoretic norm on sets of full intentions. An agent should possess a set of intentions only if it maximizes the expectation of goal-relative value, where each goal is weighted by her degree of intention for that goal. Here's the idea: we can assess an action for *how well it satisfies an agent's goals*. But an agent has different goals to different strengths. We want to weight the value of an action by how well it satisfies each goal, and by how strongly the agent is committed to that goal.

Finally, in step 3 I will use this decision-theoretic norm to defend Lockean Intention. If an agent violates Lockean Intention, she also does not maximize the expectation of goal-relative value. Extending a result from Easwaran [forthcoming], we will see:

Theorem 5.1 (Lockean)**.** An agent's full intentions maximize the expectation of her goals, weighted by her degrees of intention, iff the agent satisfies Lockean Intention for a particular threshold.

5.1 The Goal-Relative Value of a Full Intention Set

We are now interested in what set of full intentions an agent should have. Let **I** denote the agent's set of full intentions. We are assessing the value of **I** relative to some goal, g. Again, the goal is a maximal, consistent set of claims, specifying exactly how the agent would prefer the world to be.

Following Easwaran, we can measure the value of **I** at g (v(**I**, g)) as a function of the propositions on which **I** and g disagree.[32] When **I** and g agree over p, **I** receives some positive value, R. When **I** and g disagree over p, **I** receives some negative value, W. We assume that the value of **I** and g supervenes on R, W, and the number of claims on which **I** and g agree or disagree.

Here's the rough idea: relative to a certain goal, an agent does best in her intentions if she intends exactly the actions that are part of the goal. Each time she intends an action that is part of the goal, her intentions become better. Each time she intends something inconsistent with the goal, her intentions become worse. An agent has two basic values. R is the value associated with matching the goal. W is the (dis)value associated with violating the goal.

Summing up, this is our current theory of value:

Definition 5.1 (Value)**.** $v(\mathbf{I}, g) = R \times |\{\phi : \phi \in \mathbf{I} \ \& \ \phi \in g\}| - W \times |\{\phi : \phi \in \mathbf{I} \ \& \ \phi \notin g\}|$.

5.2 The Expected Value of a Full Intention Set

I have now defined the value of a set of full intentions relative to a particular goal. We can use this to define the value of a set of full in-

31. Easwaran [forthcoming].

32. Easwaran [forthcoming] 5.

tentions *simpliciter*. This quantity measures the value of a set of full intentions relative to all the goals an agent has. To measure this quantity, we can calculate a type of *expected value* for full intentions. We can weight the value of the full intention at each goal by the *degree to which the agent intends that goal*. To do this, we can introduce the agent's degrees of intention function, I. This function assigns a certain degree of intention to each goal g. And this allows us to calculate the weighted value of **I** at each goal, weighted by the degree of intention in that goal (letting G be the set of goals):

Definition 5.2 (Expected Value)**.** $EU(\mathbf{I}, I) = \sum_{g \in G} I(g) \times v(\mathbf{I}, g)$.

What is the interpretation of this value? Simply this: it is how good an action is at satisfying various goals, *weighted by how strongly an agent is committed to each goal*.

This quantity suggests a norm. An agent should not possess a set of full intentions **I** and a partial intention function I when **I** does not maximize expected value given I:

Maximization Rule S ought not: possess **I** and possess I, if **I** does not maximize $\sum_{g \in G} I(g) \times v(\mathbf{I}, g)$.

This norm is a bit stronger than the dominance norm defended earlier. Let's look at an example that explains why the norm makes sense. Suppose that Susan is deciding what to do tomorrow. She can either go to the office or go to the movie theater. Suppose further that Susan is considering a few different goals. First, she might want to finish a draft of a paper she has been working on. Second, she might want to see the new Kenneth Branagh movie. What should Susan do? The following points seem true: (a) if she wants to finish a draft of the paper, then she ought to go to the office; (b) if she wants to see the new Kenneth Branagh movie, she ought to go to the movie theater. But what if Susan is torn? In that case, what she ought to do is a function of *how strongly she is committed to each goal*. If her commitment to finishing the paper is greater than her commitment to seeing the movie, then she ought to go to the office. If her commitment to seeing the movie is greater, then she ought to go to the movie theater. Maximization Rule vindicates this common-sense thought. Susan's commitment to her goals can be modeled using a function, I, that assigns a weight from 0 to 1 to each of her two goals. The action she should perform is the one that does best, given how strongly she is committed to each goal.

5.3 Result

Here is a question: For a given degree of intention function, I, what set of full intentions **I** maximizes expected utility? Given Maximization Rule, an agent is required to possess this set of full intentions, or else revise her degrees of intention.

We can extend Easwaran's work to give a surprising answer to our question.[33] Easwaran's results show that the maximal set of full intentions for I is the one that obeys a particular Lockean threshold with respect to I. The Lockean threshold is $\frac{W}{W+R}$, a function of the agent's dispreference for violating a goal and her preference for satisfying a goal. More precisely:

Theorem 5.2 (Lockean)**.** **I** maximizes EU with respect to I iff for every action ϕ, $\phi \in \mathbf{I}$ iff $I(\phi) > \frac{W}{W+R}$.

Theorem 5.2 and Maximization Rule entail that an agent must satisfy Lockean Intention on pain of irrationality. So if an agent wants to maximize the value of her full intentions relative to goals, weighted by how strongly she intends each goal, she must satisfy a particular Lockean threshold for her degrees of intention: $\frac{W}{W+R}$.

6. A Decision-Theoretic Argument for Conditionalization

Our decision-theoretic framework is quite rich. So far, we've used it to give arguments for probabilism about intention, and a Lockean the-

33. Easwaran [forthcoming] 13.

sis. How far can we go? In this section, I'll show that our decision-theoretic framework can also provide norms for how to change partial intentions over time. These norms ultimately vindicate the traditional requirement that agents take the means to their ends.

How should an agent change her intentions over time? One answer involves means and ends. When an agent adopts a new end, she should search for the means to that end. We might express this requirement with a diachronic norm on intention. On this view, an agent is irrational if she forms an intention to ϕ without also forming an intention to take the means to ϕ:

Mean-End Coherence S ought not: form an intention to ϕ between t and t', and not intend the means to ϕ at t'.[34]

Can we give arguments for diachronic norms on intention of this type? In this section, I will give an argument for a related diachronic norm on intention. I will argue that an agent should update her degrees of intentions over time by conditionalization. More precisely, suppose that S possesses partial intention function I at t. Consider the following norm:

Conditionalization If S forms a maximally strong intention to ϕ between t and t', then S's partial intention function at t' ought to be $I(\cdot|\phi)$.

Greaves and Wallace provide a decision-theoretic argument that an agent should change her credence over time by conditionalization.[35] In this section, I will show how to extend Greaves and Wallace's argument to provide a decision-theoretic argument that an agent should change her partial intentions over time by conditionalization. Then we will see how Conditionalization relates to Means-End Coherence.

Just like Greaves and Wallace's, the argument proceeds in three steps. First, I define the value of an *update procedure*, a method of changing degrees of intention over time. Second, I provide a decision-theoretic norm that connects this value with which update procedures an agent ought to use. Third, I show that conditionalization is the update procedure recommended by this decision-theoretic norm.

6.1 Update Procedures

Let's start by defining an update procedure. An update procedure is a function from an action to a partial intention function. Intuitively, this procedure says what partial intentions an agent would have if she formed a maximally strong intention to perform some action.[36]

To model update procedures, let's introduce a partition of actions **A**. Call this partition a "decision". We can think of this set of actions as a number of alternative actions. The agent will resolve to perform exactly one of them. Once she resolves to perform this action, her degree of intention to perform the action will be 1. And let $\mathbb{I}$ be the set of all partial intention functions. We then let an update procedure be a function u from **A** to $\mathbb{I}$.

We now know what an update procedure is. We are interested in showing that one update procedure is better than every other. This update procedure is conditionalization on the agent's prior intention function. So consider some agent with intention function I facing the decision **A**. Let $\mathbf{Cond}^I$ be the update procedure u where for every A $\in$ **A**, $u(A) = I(\cdot|\ A)$.[37] I will argue that any rational agent with partial intention function I uses $\mathbf{Cond}^I$ as her update procedure.

6.2 The Value of an Update Procedure

Now that we have defined an update procedure, let's calculate its value. Here, we can begin with our earlier concept of the *goal-relative value of a set of partial intentions*. Again, let a goal g be a maximal, consistent set of claims. A goal is a maximally precise, coherent way that an agent could wish the world to be. And for each goal g, define the best possible

34. For discussion, see the sources in footnote 1, as well as Wallace [2001].
35. Greaves and Wallace [2006].
36. Greaves and Wallace [2006] 612.
37. Greaves and Wallace [2006] 613.

partial intention function. We call this the vindicated function, v_g. And let the value of a partial intention function I at g, B(I, g), be the distance between I and v_g. As before, we require that this distance be a *proper scoring rule*, such as the Brier score.[38]

With this framework in place, we can define the *goal-relative value of an update procedure*. First, an update procedure provides an agent with a unique partial intention function for each act in **A** that she might maximally intend. But, second, we can extend update procedures so that they provide an agent with a unique partial intention function for each *goal* that she might have. **A** is a partition on the set of goals. And so each goal can be associated with a particular member of **A**, the action in which the goal is satisfied. For any particular goal g, we can think of our update procedure as providing an agent with the partial intention function that is the output of the cell of **A** where g is satisfied.[39] To simplify, let's extend our notation for update procedures u so that u(g) = u(A) iff g ∈ A.

We now know what partial intention function is recommended by an update procedure relative to each goal. We can use this information to calculate the *goal-relative value of an update procedure*. The goal-relative value of an update procedure is simply the goal-relative value of the output of the update procedure when given that goal as an input. More precisely, suppose that B(I,g) is the goal-relative value of I relative to g. And let $B^+(u, g)$ be the goal-relative value of update procedure u relative to g. Then we say that:

Definition 6.1 (Goal-Relative Value of Update Procedure). $B^+(u, g) = B(u(g), g)$.

Let's summarize. How valuable should an agent consider an update procedure in light of a particular goal she might have? For an agent with a goal, the value of the update procedure is simply the value of the partial intention function that the update procedure provides, given that goal.

We have now defined how valuable an update procedure is relative to a goal. We can use this concept to define the *expected value of an update procedure*. Here, we weight the value of an update procedure at each goal, by how strongly the agent intends each goal. More precisely, we say:

Definition 6.2 (Expected Value of Update Procedure). $EU(u, I) = \sum_{g \in G} I(g) \times B^+(u, g)$.[40]

We want to know the expected value of u, some update procedure. We begin by considering the goal-relative value of u, for each goal g. This is simply the goal-relative value of I, where I is the partial intention function recommended by u when an agent maximally intends A, where A is the act in which g is true. Once we have calculated the goal-relative value of u, for each goal g, we weight this value by the degree to which the agent intends each goal. Again, this tells us how valuable her plan for updating is, relative to any goal she has, weighted by how strongly she has the goal. This allows us to calculate how valuable an update procedure is for an agent, *given her current partial intention function*.

6.3 *Maximization Norm*

To reach our conclusion, we must add a norm connecting expected value and norms for changing intentions. Here is the norm I propose: an agent should change her intentions in the way that maximizes expected value. More precisely, an agent should always conform to the update procedure that maximizes expected value, given her current degrees of intention. Summarizing:

Maximization Rule If S possesses partial intention function I, then S ought to adopt the update procedure that maximizes expected value

38. Greaves and Wallace [2006] 627.
39. Here we follow Greaves and Wallace, who say that a doxastic update procedure associates each world w with the output of the update procedure for the piece of evidence $E_j \in$ E where w is true. See Greaves and Wallace [2006] 612.
40. See Greaves and Wallace [2006] 615.

relative to I.

6.4 Result

We can now adopt Greaves and Wallace's result to justify conditionalization. For we can extend Greaves and Wallace to show that conditionalization on I is the unique update procedure that maximizes expected value relative to I. More precisely:

Theorem 6.1 (GW)**.** Where B is determined by a proper scoring rule, $\mathbf{Cond}^I$ is the update procedure u that maximizes $\sum_{g \in G} I(g) \times B^+(u, g)$.[41]

Theorem 6.1, Definitions 6.1–6.2, and Maximization Norm entail that an agent is rational only if she satisfies Conditionalization. This completes our decision-theoretic argument that rational agents update their partial intentions by conditionalization.

6.5 Means-End Reasoning and Conditionalization

What is the philosophical significance of Conditionalization? Conditionalization gives a precise vindication of the folk-psychological norm that an agent ought to intend the means to her ends.

Suppose an agent forms a maximal intention towards some end. Conditionalization requires the agent to conditionalize her partial intention function on this end. This means that the agent must now become maximally committed to any logical consequence of the end. And this vindicates the idea that an agent should maximally intend any necesssary means to her end. For a necessary means to an end is simply a consequence of that end.

What about sufficient means? A sufficient means to an end entails the end. And so the probability of the end, conditional on the means, is 1. By Bayes' Theorem, we know that the probability of the means, conditional on the end, is positively correlated with the probability of the end, conditional on the means. And Conditionalization requires that the agent's posterior commitment to the means equals her prior commitment to the means, conditional on the end. Thus becoming committed to an end will tend to increase the agent's commitment to sufficient means.

Finally, it's worth observing that Conditionalization allows us to explicate the distinction between intended means and merely foreseen side effects. For an agent's partial intention function is distinct from her partial belief function. Consider a case where an agent foresees that if she ϕs, she will ψ, but where ψ is not intended. Since ψ is foreseen, we may suppose that the agent's credence in ψ, conditional on ϕ, is quite high. But since ψ is not intended conditional on ϕ, we may suppose that her degree of intention to ψ conditional on ϕ is low. How is this possible? We may suppose that although the agent thinks ψ is quite likely to occur, given ϕ, she is also quite committed to it not occurring. And so she actively seeks means to thwart ψs occurrence, conditional on ϕ. Now suppose that the agent forms a maximal intention to ϕ, and becomes certain that she will ϕ. In this case, her credence that she will ψ will become quite high, while her degree of intention to ψ will become quite low. So we predict that an agent can foresee an effect of an action without intending that act as a means.

Conditionalization shows that rational agents become fully committed to necessary means of the ends they are fully committed to. And rational agents tend to become more committed to the sufficient means of an end, once they become fully committed to that end. We have now provided a decision-theoretic argument for the means end principle of reasoning that we started with.

7. Conclusion

In this paper, I have explored the preface paradox for intention. I have offered a solution to the paradox: intentions come in degrees.

I offered several arguments for this solution. First, I provided a model of partial intention on which it is metaphysically respectable. We need only commit to dispositions coming in degrees — and we

41. Greaves and Wallace [2006] 623.

needed that anyways. Then I showed that on this model, we would expect rational agents to conform their partial intentions to the probability calculus.

Yet one might not want to hold the norms of partial intention hostage to a particular metaphysical interpretation. So in the last sections of the paper, I developed a series of decision-theoretic arguments that rational agents conform their partial intentions to the probability calculus. These arguments also showed that rational agents satisfy a Lockean thesis for intentions, and that rational agents update their intentions over time through conditionalization.

What results is a unified theory of intention and belief. While these states are different, they have a common structure. Each state is ordered by degrees, and each state is subject to a common set of norms.[42]

References

Michael Bratman. Two faces of intention. *The Philosophical Review*, 93 (3):375–405, July 1984.

Michael Bratman. *Intention, Plans, and Practical Reason*. Center for the Study of Language and Information, 1999.

Michael Bratman. Intention, belief, practical, theoretical. In Simon Robertson, editor, *Spheres of Reason*, chapter 2. Oxford University Press, 2009.

John Broome. *Rationality Through Reasoning*. Wiley-Blackwell, September 2013.

David Christensen. *Putting Logic in its Place*. Oxford University Press, 2005.

Donald Davidson. Actions, reasons, and causes. *The Journal of Philosophy*, 60(23):685–700, November 1963.

Bruno De Finetti. *Theory of Probability*. Wiley, 1974.

42. I am grateful to Bob Beddor, David Black, Michael Bratman, Ben Bronner, Megan Feeney, Branden Fitelson, Kathryn Goldstein, Charles Hermes, Daniel Rubio, Ernie Sosa, Sergio Tenenbaum, Tobias Wilsch, audiences at the 2014 Joint Session and 2015 Pacific APA, and two referees for helpful discussion.

Kenny Easwaran. Dr. Truthlove or: How I learned to stop worrying and love Bayesian probabilities. *Noûs*, forthcoming.

Richard Foley. *Working Without a Net*. Oxford University Press, 1993.

Hilary Greaves and David Wallace. Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459): 607–632, July 2006.

Paul Grice. Logic and conversation. In *Studies in the Way of Words*, chapter 2. Harvard University Press, 1975/1989a.

Richard Holton. Partial belief, partial intention. *Mind*, 117(465):27–58, January 2008.

James M. Joyce. A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4):575–603, December 1998.

James M. Joyce. Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In Franz Huber and Christoph Shmidt-Petri, editors, *Degrees of Belief*, pages 263–300. Springer, 2009.

Gregory S. Kavka. The toxin puzzle. *Analysis*, 43(1):33–36, January 1983.

David K. Lewis. *Counterfactuals*. Blackwell, 1973.

D.C. Makinson. The paradox of the preface. *Analysis*, 25(6):205–207, June 1965.

David Manley and Ryan Wasserman. On linking dispositions and conditionals. *Mind*, 117(465):59–84, January 2008.

Richard Pettigrew. Epistemic utility and norms for credence. *Philosophy Compass*, 8(10):897–908, October 2013.

Jacob Ross. How to be a cognitivist about practical reason. *Oxford Studies in Metaethics*, 4:243–281, 2009.

Sharon Ryan. The preface paradox. *Philosophical Studies*, 64(3):293–307, December 1991.

Sam Shpall. The calendar paradox. *Philosophical Studies*, 173(3):801–825, March 2016.

Robert C. Stalnaker. *Inquiry*. MIT Press, 1987.

James David Velleman. *Practical Reflection*. Princeton University Press, 1989.

R. Jay Wallace. Normativity, commitment, and instrumental reason.

Philosophers' Imprint, 1(3), December 2001.

Gideon Yaffe. Trying, intending and attempted crimes. *Faculty Scholarship Series*, (Paper 3712), 2004.