The Desire-Belief Account of Intention Explains Everything

NEIL SINHABABU

National University of Singapore

Views on which intentions are desire-belief pairs have been widely criticized for failing to explain basic facts about intention. Kieran Setiya writes that "If A is doing φ intentionally, A believes that he is doing it or is more confident of this than he would otherwise be, or else he is doing φ by doing other things for which that condition holds" (391).[1] I'll call this phenomenon "belief about doing." Setiya also writes that "we *choose* the reasons on which we act" (39).[2] I'll call this phenomenon "reason-choosing." Michael Bratman notes "three central facts about intentions: they are conduct-controlling pro-attitudes, they have inertia, and they serve as inputs into further practical reasoning" (27).[3] Bratman treats these facts as "a source for the challenge to the descriptive aspect of the desire-belief model" (27). Setiya regards belief about doing and reason-choosing the same way.

I'll argue that a desire-belief account can explain these phenomena whenever they obtain. First I'll present a desire-belief account. Then I'll show how it explains belief about doing, reason-choosing, and the three facts Bratman describes. Not all of these phenomena are present in all cases of intention. We should accept the desire-belief account because it explains their presence and absence simply and accurately.

1. The Desire-Belief Account

A intends that φ if A has a desire and belief such that for some behavior B and situation S:

(1) A desires that φ.
(2) A believes that S will obtain, and that A's B-ing in S would make φ more likely.
(3) If A were to believe that S obtained, the desire and the belief would, without further practical reasoning, produce motivational force causing A to initiate B-ing.

This account is similar to those of Donald Davidson and Robert Audi, and some objections I'll address were originally intended for them.[4]

I take A's intention to be the desire from (1), while (2) and (3) are background conditions necessary for a desire to be an intention. An intention that φ thus is an appropriately situated desire that φ. If you think intentions are composed of all the

psychological states they involve, you might instead take A's intention to consist of the desire from (1) and the belief from (2), positioned in the agent's psychology as described in (3). My arguments will support both views alike. I won't give an account of φing intentionally.[5]

The desire in (1) can be instrumental or final. In typical cases of intention, the final desire for some goal will generate motivational force favoring not only the goal, but a chain of events leading from the behavior to the goal, each event raising the probability of the next. The agent instrumentally desires and thus intends each event in the chain leading to the goal. For the purposes of this formulation, these instrumental desires have motivational force, though all the things someone instrumentally desires as a means to satisfying one final desire obviously shouldn't be separately added into the motivational calculus. Perhaps we should say that the motivational force of an instrumental desire is just the motivational force of the final desire or desires from which it is derived. φ can be the action, the goal, or any intermediate step in the chain, as an intender will instrumentally or finally intend all these things. The agent's goal may just be performing behavior B, giving the chain of events only one link. Performing a behavior is a straightforward way of making its performance more likely.

In (2), situation S may either be a present situation in which the agent is acting or a possible future situation. This provides a unified treatment of intention in action and prior intention.[6] The agent need not characterize B or S precisely. An earnest college student may fervently intend to do something, someday, to reduce global poverty, but lack knowledge of his future so that he can't fill out the intention any further. All one must believe is that S will obtain, and that B-ing then will make φ more likely. If one's credence that S will obtain falls short of belief, but all the other conditions in the account are met, one has a conditional intention.

(2) allows agents to intend things they believe are highly improbable. The basketball player who shoots from behind halfcourt just before time expires intends that the ball go into the basket, despite knowing that long-distance shots like this rarely go in. This is the intuitive way to describe his intention. Think how upset his teammates would be if he said afterwards that he didn't intend that the ball go into the basket. By requiring only that the agent believe that his action would make φ more likely, not that φ will occur if he acts, (2) accounts for desperation shots and other cases where we rationally intend the improbable.[7] The use of subjective probability in (2) is part of why I discuss "intending that" rather than "intending to"—subjective probabilities are more straightforwardly assigned to φ if it's a state of affairs than if it's an action. I hope this account of "intending that" will smoothly extend to "intending to," but I won't explore this here.

"Motivational force" in (3) is the product of the desire's strength from (1) and the rise in its expected probability of satisfaction from (2). This multiplication of desire strengths by subjective probabilities is familiar from desire-belief accounts of motivation and causal decision theory. If multiple desire-belief pairs jointly cause the action, as when one drinks whiskey both for its taste and for intoxication, one has multiple intentions and intends every value of φ in each of them. If one desire-belief pair produces motivational force favoring action and another produces

motivational force against action, the force favoring action will have to be greater to cause action. If the desire-belief pair favoring action is not strong enough to cause action, this prevents it from being an intention, as (3) implies.[8]

(3) rules out cases where the agent desires something and has a belief about how to attain it, but hasn't yet combined these mental states and formed an intention. In these cases, intention-formation requires a further step of practical reasoning in which the desire and belief are combined so they can motivate action. I talk of "initiating" rather than "performing" an action to accommodate cases of paralysis. "Initiating" includes both the beginnings of ordinary action by non-paralyzed people and whatever fragment of action occurs when a paralyzed person tries to move, but finds that he can't. Even if paralyzed people can't act, this lets them have intentions.[9]

This desire-belief view is only a set of sufficient conditions for intending and an account of how intention is psychologically realized in human beings.[10] It isn't a set of metaphysically necessary conditions—there might be many metaphysically possible ways to combine mental states (including ones that humans don't have) into psychological structures with the right functional properties to be intentions. Neither is it a conceptual analysis, an account of what makes action rational, or an account of what makes something an action. I happily concede to opponents of the desire-belief view that there are other ways in which intentions could be psychologically realized. Following Bratman, we can make artificial intelligences whose intentions aren't composed of beliefs and desires. Maybe there are aliens whose Velleman-style intentions always involve desires for self-knowledge.[11] I can imagine angels whose intentions are constituted by a kind of Anscombean practical knowledge that is "the cause of what it understands" (87).[12]

We need not be like these creatures to have intentions. In fact, we aren't like them. An appropriately related desire-belief pair is sufficient for intention, and that's how human intentions are composed. My opponents deny this, presenting familiar phenomena from actual human deliberation and action which they claim the desire-belief view can't explain. In explaining these phenomena, I'll appeal to many things desire does, like motivating action, directing attention, and increasing in intensity when we have a vivid image of what is desired. Often I'll argue that the phenomena my opponents cite aren't as ubiquitous as they claim, and that the desire-belief view successfully explains both their presence in some cases and their absence in others. I hope to leave you with simple, illuminating explanations of several phenomena surrounding intention. That the desire-belief view provides such explanations is the best reason to accept it.

2. Belief about Doing

First, I'll show how the desire-belief account explains belief about doing. I'll focus on Setiya's recent description of the phenomenon and argue that the ability of desire to drive an agent's reasoning plays a key role in explaining it. The desire-belief account won't make belief about doing a necessary feature of intentional action. I'll present an example of intentional action without belief about doing to show that this is an advantage.

In "Practical Knowledge," Setiya presents belief about doing as follows: "If A is doing φ intentionally, A believes that he is doing it or is more confident of this than he would otherwise be, or else he is doing φ by doing other things for which that condition holds" (391). In his earlier book, *Reasons Without Rationalism*, he uses a stronger version of this principle, which he calls *Belief*, against desire-belief models of intention. The stronger version is not put in terms of greater confidence, but just in terms of belief. It has problems with cases of the sort where someone is recovering from paralysis in his hands and doesn't yet know whether his attempts to clench his fists will be successful. While intending to clench one's fists at some moment should make one more confident that one is clenching one's fists, this increase in confidence might fall short of generating full-blown belief that one is clenching one's fists or doing anything at all. Setiya's new formulation avoids this problem.

In his book, Setiya calls the desire-belief picture the "standard model" of intention, following Bratman. He claims that it can't explain *Belief*:

> [A]n agent can be motivated by the desire for an end, and the belief that doing x is a more or less effective means to that end, without having the belief that he is doing x. The presence of that belief does not follow from belief-desire motivation; nor does it follow that there is anything he does in the belief that he is doing it. If the standard model were correct, *Belief* would not be a necessary truth. But it is, and so the standard model is false. (33)

Setiya is right that the desire-belief account doesn't imply that agents necessarily believe they're doing anything intentionally when they act. The belief from (2) can't do it alone, because it only implies that acting in some situation might have some effect, not that one will act. The belief that they're in the appropriate situation, from (3), doesn't do it either—as a precondition of action, it doesn't include the belief that they're acting.[13]

Setiya rejects Paul Grice's inferential view on which "knowledge of what I am doing in acting intentionally is inferred from prior knowledge of my will" (394).[14] Setiya writes:

> If, when I am clenching my fist intentionally, the belief that I am doing so had to be inferred from the premise that I intend to be clenching my fist and from empirical knowledge of the conditional that if I intend to do so, I will, it would be possible for the inference not to take place. I might simply fail to put two and two together.

Setiya objects that "The problem with such an account is that it cannot explain why it should be a necessary truth that doing something intentionally is doing it knowingly." This objection is significant. Desire-belief theorists must explain why we don't fail to put two and two together.

The explanation comes from desire's ability to direct our attention on things relevant to its satisfaction and drive our reasoning about these things. David Hume describes how a desire, "making us cast our view on every side, comprehends whatever objects are connected with its original one by the relation of cause and

effect. Here then reasoning takes place to discover this relation" (2.2.3).[15] Desire's ability to make us reason about how its object can come about is familiar. Someone may believe all the premises from which he could discover how to enter his house without a key, but not attend to this matter and draw the conclusion until he's locked out and desires to get in. Desire directs even reasoning that isn't oriented towards action. On election night after voting ends, partisans consider the vote totals from areas that report results early and draw conclusions about how many votes they'll need from other areas. Even though their actions can't affect the election anymore, the desire for victory directs their attention at the results and drives reasoning about how victory might be attained.

Desire's ability to drive reasoning explains why agents infer that they're acting. When someone is about to act on an intention, the desires motivating the action are occurrent. They drive reasoning about how their object might be attained. The action at hand is the means to attain their object, so it's especially salient. The agent knows that he intends to act, and he knows that if he intends to act, he'll act. His desires drive him to reason from these premises to the conclusion that he'll act.[16]

The desire-belief view explains a difference between the way we know that we're acting and the way we know other things. Suppose I'm writing a paper while kicking the wall in rhythm to the music from my headphones. Writing the paper is an intentional action, while kicking the wall isn't. I know I'm writing the paper, but I may not know I'm kicking the wall until I hear a thump from the other side of the wall, where my annoyed colleague is trying to make me stop. I didn't know I was kicking the wall because I had no desire that would focus my attention on the wall-kicking. The desires that motivate philosophical writing focused my attention and my reasoning on things associated with their objects, and I didn't even notice my other movements.

The desire-belief view doesn't explain the supposed necessity of belief about doing. As it's just a contingent psychological fact about human desire that it directs attention, desire-belief theorists are committed to the metaphysical possibility of creatures who lack belief about doing because their desires don't direct their attention. One might even argue that desire-belief views can't explain why belief about doing is ubiquitous in human action. While the psychological connection between attending to something and forming beliefs about it is strong, it can be overcome. If the forces producing belief about doing are mere psychological tendencies, why is this phenomenon so robust?

Sarah Paul suspects that the phenomenon isn't so robust.[17] She discusses "the classic example of driving home on autopilot," which we'd describe as something done intentionally, but during which the driver "may have no belief that he is doing these things, and might only be able to discover that he is doing them by observation" (5). She takes cases like this to suggest that the adverb "intentionally" may not track belief about doing. However, she allows her opponent the response that "action for a reason" may require belief about doing. I'll argue against even this response by offering a strange case in which someone acts for a reason, but is *less* confident that she's acting than she would be if she hadn't formed the intention.[18]

Fidgety young Bridget, who has a tendency to unintentionally bounce up and down in her chair, has been kidnapped by the evil wizard Zur. Zur lies that he has enchanted her so that if she ever intends to bounce, she'll immediately stop moving while being under the illusion that she's bouncing. Bridget believes Zur's lie and is 99% sure that he has cast the enchantment, though she doesn't really care. Later, when Zur sees Bridget unintentionally bouncing in her chair, he summons Yawgmoth, a demon who devours humans, but never those who are bouncing. Bridget desires to survive, and knows that in this situation she can only survive by bouncing. She can't keep her bouncing unintentional, and she starts to bounce intentionally.[19] As this happens, she infers from knowledge of her new intention to bounce and her belief in Zur's enchantment that there's only a 1% chance that she's bouncing. She wishes she hadn't formed the intention, believing that it has made her sit still when she'd otherwise be bouncing unintentionally, until Yawgmoth sees that Zur isn't bouncing and eats him instead.

When she wishes she hadn't formed the intention, Bridget is bouncing intentionally, while not believing that she's bouncing and having less confidence that she's bouncing than she would if she hadn't formed the intention. She isn't bouncing by doing anything else, once we exclude her "inner act of volition" (390) or "the firing of nerves," (391) to which Setiya rightly says he can't appeal. (Nothing supernatural happens to Bridget—Zur doesn't actually enchant her.) Therefore, Bridget's case is a counterexample to Setiya's principle.

Setiya's pre-emptive response addresses how "more confident of this than he otherwise would be" should be understood:[20]

> The counterfactual in this principle must be handled carefully. Couldn't there be an action that is normally automatic but which can be done intentionally with a lower chance of success? If one is aware of all this, one will be, on balance, less confident that one is performing that action when one is doing it intentionally—but still more confident than if one were not doing it intentionally and one's automatic system were shut down. (391)

Setiya's suggestion for how to treat the counterfactual should be rejected. One's automatic systems keep running as usual in the absence of an intention to produce their behavior, on any ordinary conception of what otherwise would be. It's hard to see how to modify his formulation to explicitly shut down automatic processes without rendering it trivial or *ad hoc*.

The desire-belief account doesn't imply that Bridget would form the belief about doing. When her desire to survive drives her to reason about her situation, she concludes that she's probably not bouncing, while she'd know she was bouncing if she hadn't formed the intention. The second premise of the Gricean reasoning is reversed—she believes that her intention makes her less likely to bounce. Belief about doing is widespread because it's very rare that someone intentionally φs while believing that intending to φ makes φing less likely. But when this happens, φing intentionally can reduce our confidence that we're φing. This case supports the inferentialist picture and shows that the desire-belief view is right to make belief about doing a merely contingent correlate of intentional action done for reasons.

3. Reason-Choosing

Now I'll provide a desire-belief explanation of reason-choosing. I'll explain how agents who see two reasons for some action, one of which comes from a weaker desire and one of which comes from a stronger desire, can act solely on the reason grounded in their weaker desire. These agents need self-knowledge about their motivations and a strong second-order volition that keeps the stronger desire in check. These elements are absent from many cases of intentional action, so we don't usually choose which reason to act on, and our attempts to choose may fail.

Many philosophers hold that we can choose which of several reasons to act on. In *Rationality in Action*, John Searle writes, "when one has several reasons for performing an action, one may act on only one of them; one may select which reason one acts on" (65).[21] He offers a case in which someone has several reasons for voting for a particular candidate, and claims that one can vote for the candidate for one of these reasons, but not for his other reasons. Christine Korsgaard seems to express a similar view in her *Locke Lectures*, claiming that "the *aim* is included in the description of the action, and that it is the action as a whole, *including the aim*, that the agent chooses" (1.2.4).[22] Setiya writes that "A second defect [of the desire-belief view] is its failure to accommodate the active and reflective character of the attitude we take to reasons, in acting on them. Our reasons are in some sense "up to us"—we decide why to do something, as well as what to do—and we seem to recognize our reasons, as such" (39). Writing that "we *choose* the reasons on which we act," he offers an example:

> There are many reasons for which I might decide to write a book: personal satisfaction, a fragment of immortality, professional ambition. I am not passive in the face of this: even if I believe that books give their authors a kind of immortality, and even if I think that this is a reason—a good reason—to write a book, it may not be *my* reason for doing so. That is up to me. (39–40)

We can develop Setiya's example in a way that makes it especially challenging for the desire-belief view. Even if the desire that most strongly supported writing the book was the desire for a fragment of immortality, Setiya need not have written the book for this reason. Perhaps he really liked the idea of writing simply for the satisfaction of writing, and his desire for personal satisfaction was in fact strong enough to motivate the act of writing by itself. Then even if his desire for a fragment of immortality was the strongest one supporting writing, it may not have been his reason for writing.

It may be mysterious how desire-belief views can explain how someone may choose which desire to act on. The notion of reasons involved here is that of motivating reasons, the reasons for which someone acted. These, on a desire-belief view, are something like the agent's desires and beliefs.[23] So it seems that agents can only choose their reasons by choosing which desires and beliefs motivate them. Traditional desire-belief views claim that we can't create desires just by choosing to have them. It's widely accepted that we can't create beliefs by mere choice either. So it's hard, at first glance, to see how we can choose our reasons.

The desire-belief view allows reason-choosing by allowing a second-order voli-tion to block the strongest desire supporting some action from motivating it, so that a weaker desire is the sole motivator. Suppose one has a desire concerning which desire will motivate a particular action—in Frankfurt's terms, a second-order vo-lition.[24] Let's call the first-order desire that one desires to be effective the favored desire, and the other ones disfavored desires. Now suppose someone with a strong second-order volition sees that a disfavored desire is about to motivate his action. As long as he believes that it's possible to act on the favored desire, the second-order volition will exert motivational force against acting on a disfavored desire. After all, doing the action out of a disfavored desire prevents him from doing it solely out of the favored desire. If his second-order volition and his belief that he can act on the favored desire are strong enough, the second-order volition will block all other ways of acting, so that the favored desire alone causes action, and provides his reason.

While I don't know which reason Setiya chose for writing his book, I can explain how he chose. He had a number of desires which pushed in favor of writing the book, and because of which he saw a number of reasons for writing. He desired to write the book for some reasons rather than others. If this desire about which reason to act on was strong enough, it could exert motivational force blocking any decision to write which he believed arose from the wrong reasons. Then only the reasons that he desired to motivate his action would be his.

Successful reason-choosing can be hard. It requires three things. First, one needs a sufficiently strong desire not to act for the wrong reasons. If this desire is too weak, it can be overpowered by a conflicting desire. Then one will act on reasons one doesn't want to act on, possibly feeling an unpleasant emotion like guilt or shame as one acts. Second, one's favored desire must be strong enough to overcome whatever desires may be blocking action. Otherwise, even though one's disfavored desires are being held at bay, the favored desire won't motivate action. Third, one needs an impressive sort of self-knowledge. Sometimes people desire not to act for a particular sort of reason, and believe that they aren't acting for that reason when in fact they are. Sexism and racism often work this way. Even those who strongly desire not to make decisions in a sexist or racist way may act in a way that is driven by sexist or racist attitudes. They may not know that such attitudes are driving their decisions. Since they lack the belief that could generate a second-order volition blocking action, their desire not to make sexist or racist decisions won't stop them from acting. If they had greater self-knowledge, and could tell what motivational forces were about to drive their decision, the desire not to act for these pernicious reasons would spring into effect, and they'd act differently.

Our ability to choose our reasons is quite restricted. Suppose Andy is drinking whiskey. His reason might be that it tastes good, or to get drunk. Unless he has unusual desires or beliefs, his reason for drinking can't be to reduce the amount of drinkable liquid in town, or to save the whales. The former reasons are plausible because he might desire these things and believe that drinking the whiskey will bring them about. But as long as he doesn't desire to reduce the amount of drinkable liquid in town, and doesn't believe that drinking will save the whales, he can't act

for the latter reasons. Even if his frat brothers offered him a hundred dollars to drink whiskey for the reason that it would reduce the amount of drinkable liquid in town, he couldn't drink for that reason.

Andy's case illustrates how the desire-belief view explains which considerations are eligible to be reasons for action. Searle might respond that we can only choose reasons that we already accept at some level, but that we are free to choose any of these reasons. So I'll offer a case to show how even someone with perfect self-knowledge who takes something as a reason can fail to choose it as her reason for action.

Jane has received a marriage proposal from the King. She is poor, and she knows his wealth gives her a good reason to marry him. She also knows he is very kind, and she takes that as a good reason to marry him too. But he is old and grey, and had he been a commoner, she would've politely turned him down. She knows this, and sighs as she thinks about it. For it means a lot to her that her reason for marrying be something about her husband himself, and not his money. Something like his kindness! What a wonderful girl she would be if she could marry him for that reason! She wishes she could find a witch to cast a spell on her increasing her love for kind men, so she could marry him for his kindness, but all the witches were burned long ago.

Jane can't make the King's kindness her reason to marry him because she doesn't desire him for his kindness strongly enough. Her desire to marry a kind man isn't part of a sufficiently powerful desire-belief pair to cause her action alone, as she realizes when she imagines what she'd do if he were a commoner. So her desire to marry a man for his personal qualities can't be satisfied. (It's a sad fact of life that desiring to desire φ doesn't directly increase one's desire for φ. Otherwise we'd be more motivated at work, and happier in love.) As it stands, the King's kindness can be part of her reason for marrying him, but only a part. Having self-knowledge, Jane can't say that she would have married him for his kindness alone. Sometimes we want to act solely on our favored desires, and that's what Jane can't do.

The psychological machinery that allows successful reason-choosing is complex. Usually it won't all be in place. This fits the data. We choose our reasons only in a small fraction of our everyday actions. For the most part, we cook, cross streets, and casually converse without concern about which reason is motivating us. This is because we lack second-order volitions that play any important role in preventing unsavory first-order desires from providing our reasons. The desire-belief account explains how reason-choosing is possible, how we can fail at it, and why we don't attempt it in most cases of intentional action.

4. Bratman and Deliberation

Now I'll consider objections to the desire-belief model from Michael Bratman, who offers an account on which A intends to φ just in case:

(a) A has a "conduct-controlling" disposition to φ, (b) A is disposed not to deliberate any more about whether to φ unless new and relevant information comes to light, and

(c) A is disposed to deliberate about intended means or preliminary steps to φ as well as about more specific intentions, as when one reasons from an intention to take a bus to the library and information on a bus schedule to the intention to take a particular bus.

Bratman thinks desire-belief accounts can't explain these features of intention. I'll argue that (a) is easily explained by desire's most famous functional property, its ability to motivate action. Bratman's discussion suggests that (b) is really two separate things—first, that we tend not to reconsider our intentions without new information, and second, that forming one intention excludes the formation of an inconsistent intention. The former doesn't require a robust explanation, while the latter is explained by beliefs about our future intentions. The desire-belief view explains why both phenomena generally obtain. It also explains reconsideration without new information in cases where desire strengths fluctuate, and for rare inconsistent intentions. The ability of desire to direct attention explains (c).

Bratman claims that the desire-belief view can't explain (a), the conduct-controlling nature of intentions. In the section titled "The Initial Challenge to the Desire-Belief Model," Bratman writes that there is a "problem with the reduction of intention to predominant desire. An intention to A is a conduct controlling pro-attitude. But my predominant desire to go to Tanner at noon does not guarantee that when I see it is noon my desire will control my conduct. I might still be disposed to deliberate about what to do; for I might still not see the issue as settled" (19). Here the desire-belief theorist can appeal to desire's most well-known feature, its ability to motivate action when combined with an appropriate means-end belief.[25] While a desire alone might not motivate, Bratman isn't attacking a view on having a predominant desire is sufficient for intending. He's attacking a view on which desire-belief pairs are sufficient for intending. These are sufficient for motivation, and thus conduct-controlling. If there's some problem with desire-belief pairs controlling conduct, Bratman hasn't told us what it is.

It's easier to see why Bratman thinks desire-belief accounts can't explain (b), which concerns the role of intention in deliberation. My account of intention doesn't directly imply anything about deliberation. This problem is especially severe if, like Bratman, one regards desires as "merely *potential influencers* of action" (16). Bratman gives us an example of (b), again involving a desire to go to Tanner library:

> . . .even if I now have a predominant desire to go to Tanner. . .I might still not see the issue as *settled*: I might be disposed to continue to give serious consideration to the possibility of taking the afternoon off and going to a concert. But if I were to intend to go to Tanner, I would be disposed not to continue to deliberate in this way: this is what is involved in the resistance to reconsideration characteristic of intention. (Bratman 1987: 18–19)

And earlier: "My intention resists reconsideration: it has a characteristic stability or inertia" (17).

In "Humean Intentions," Michael Ridge explains deliberative stability by including an extra element in his account of intention: "(d) A has a desire not to deliberate

any more about whether to φ unless new, relevant information comes to light."[26] While adding an additional element makes the desire-belief account less elegant, it explains the phenomena Bratman cites while grounding all motivation in desire. As Ridge notes, Bratman gives the desire-belief theorist something to work with when he says that intention involves a disposition not to deliberate further. Ridge claims that this disposition operates in combination with a means-end belief as desires do when they motivate action: "this disposition is presumably one that is potentially belief-mediated; if I come to believe that I am deliberating about something I have formed an intention to do when no new, relevant information has come to light, I typically will stop deliberating."

Unfortunately, Ridge's proposal combined with the richer conception of desire I've invoked here would produce false phenomenological predictions. On this conception, desire directs our attention towards things relevant to its satisfaction as we deliberate, driving us to reason about how its object might be achieved. So a desire not to deliberate any further would focus my attention on my own mind as it operated, making me watch myself to see whether I was going to deliberate or not. After all, if deliberating further is the kind of thing I'd be likely to do, desiring not to deliberate further would make me worry that it was going to happen. The idea that these dispositions are belief-mediated deepens the difficulties. If I saw myself starting to deliberate again, I'd have to think of means to stop deliberating, such as directing my attention elsewhere or resolving not to act on the deliberation and thus making it pointless. These mental gymnastics don't need to occur when I cease deliberating after forming an intention—the way I cease deliberating typically is much more automatic.[27] There certainly are some cases, often involving time constraints that make further deliberation costly, where one might desire not to deliberate further, be frustrated with oneself for deliberating too much, and engage in some of these mental gymnastics to stop deliberation. In situations like this, the agent clearly has a desire not to deliberate further, and thoughts like "no more second-guessing myself!" are natural. But when time constraints are absent, it's not clear why the agent would desire not to deliberate further. The agent might have nothing better to do, and be indifferent about further deliberation. Ridge's view has the unfortunate consequence that we couldn't then attribute intentions to the agent, as (d) would be false. Furthermore, it would be best to preserve the simplicity of the desire-belief view by finding something already within it that explains the phenomena Bratman cites. I'll try to do this.

Let's distinguish two phenomena that (b) picks out. The first phenomenon, expressed in Bratman's initial formulation of (b), is that we don't constantly reconsider our intentions. I'll call this "nonreconsideration." It doesn't require a robust explanation. We usually don't reconsider our beliefs or other mental states unless new evidence or an interesting event makes us do so. What requires a robust explanation is the reconsideration of an attitude, not nonreconsideration. The second and more interesting phenomenon, which comes up when he describes how he doesn't think about taking the afternoon off and going to a concert, is that we don't usually form new intentions that conflict with existing ones. That having one intention usually excludes seriously considering or forming a conflicting one deserves a robust

explanation. I'll call this phenomenon "exclusion" and explain it. I'll also discuss unusual cases where nonreconsideration and exclusion don't obtain, and show how the desire-belief view handles these cases too.

Beliefs about our future actions explain exclusion. These beliefs sometimes accompany current intentions, but they can generate exclusion even without current intentions. Suppose that at 8 AM Mike believes that at noon he'll intend to go to Tanner, and that he'll be able to go there. Even if he has no desire to spend noon that way at present, his belief about his future intention will prevent him from seriously deliberating now about whether to do things at noon that he can't do if he's at Tanner. Mike believes that he won't be able to do these things after his strange new intention emerges and causes him to go to Tanner. He may consider the things he'd do elsewhere at noon in the unlikely event that the new intention doesn't arise. But since he believes that the situations in which he'd do these things won't obtain, because his intention will in fact take him to Tanner, he can only form conditional intentions to do them. All Mike needs for his future desire to give an option deliberative stability here is a belief about his future actions. If he has a belief about what he'll do, he won't make conflicting plans. The way his beliefs about what he'll do constrain his intentions is continuous with the way his beliefs about other things constrain his intentions. If he believes he won't be Pope at noon, he can't intend to issue a papal bull at noon. Beliefs constrain intentions by telling us that we won't be in the situations we'd need to be in to act on them, and beliefs about our future actions are no different.

Some theorists take intentions to be beliefs about what one will do. Bratman considers explaining exclusion by building such a belief requirement into intention, but rejects it. I sympathize with his rejection. Any account of intention can explain exclusion if it lets us easily generate the right sort of beliefs about our future actions. These beliefs need not be the intentions themselves, or part of the intentions. All that matters for exclusion is that when we intend to do something in the future, we form the belief that we'll do it. The explanation of belief about doing presented earlier in this paper doubles as an explanation of how these beliefs about future actions are formed. My desires focus my attention on future events that might promote or prevent their satisfaction, including my own actions. When I intend to perform some action, I usually believe that I'll do it.

One might wonder how beliefs about future actions can explain exclusion, given that these beliefs are easily revised in some cases. The answer is that in these cases, intentions are similarly easily revised, supporting the hypothesis that beliefs accompanying intentions explain exclusion.[28] Tanya's belief that she'll get vodka may be revised to a belief that she'll get tequila and then revised back several times between her leaving the table and reaching the bar, as her intentions about which drink to get change. In what order do her beliefs change in such a case? Suppose she intends and believes that she'll only get one drink. Then she won't at any one time believe both that she'll get vodka and that she'll get tequila, or else she'd have contradictory beliefs. So the belief that she'll get tequila can't rise before the belief that she'll get vodka falls. This is as we'd expect if beliefs about future actions, formed along with intentions, explain exclusion. The fall of her

belief that she'll get vodka allows her to form the intention for tequila that it excluded and the belief that she'll get tequila. When conflicting desire-belief pairs produce approximately the same motivational force, slight fluctuations in the force of our desires (possibly triggered by environmental cues) can cause us to abandon old intentions and the beliefs about future actions that we had, allowing us to form new intentions, and with them, new beliefs. Of course, in many cases our intentions and the accompanying beliefs about future actions are firmer. Tanya's firm intention to accept the excellent tenure-track job she's been offered will be accompanied by a firm belief that she'll accept the job, in line with the hypothesis that beliefs accompanying intentions explain exclusion. As she firmly intends to accept the job, she firmly believes that she'll accept it. This firm belief excludes conflicting intentions more firmly than her less firm belief that she will get vodka rather than tequila.

Sometimes exclusion fails dramatically and we form inconsistent intentions. Tim may intend to stay at the movie from 6 to 8, but also intend to be at the restaurant at 7. The inconsistency of Tim's intentions is the inconsistency of his beliefs about what he'll do. Tim believes that he'll both go to the restaurant and not go to the restaurant at 7. Inconsistent intentions have the same rarity that inconsistent beliefs do because the forces preventing them are the same forces preventing inconsistent beliefs. A theory of intention should allow for the possibility of inconsistent intentions while explaining why they're weird. Explaining the possibility but weirdness of inconsistent intention in terms of the possibility but weirdness of inconsistent belief accomplishes this.

Now I'll return to nonreconsideration, and consider a case where it fails even without new information. Susie has thought things through, and she intends to break up with Robert tonight. She intends this knowing exactly how he's going to plead with her, and how blue his eyes are, and how his hands are going to feel in her hair. But knowing is different from seeing, and hearing, and feeling. Despite her intention and her knowledge, his rakish charm may cause her to violate her intention in a classic case of weakness of will. And—here's the counterexample—it may be even worse. Perhaps, as he predictably gives her a confident grin and invites her into the bedroom, she'll reconsider her intention and say to herself, "even if he steals your money twice a week and spends it on cocaine, he's worth it!"

Full information about what will happen (at least until she decides to follow him into the bedroom) won't keep Susie's intention stable when she has vivid sensations of Robert's many charming features. Without new information, she can engage in full-fledged intentional action against her prior intention. Neil Sinhababu describes a feature of desire that explains this:]

> Intensification by Vivid Images: When agents are presented with vivid images they associate with a state of affairs they desire, either in imagination or by their senses, that will strengthen the desire's causal powers. The desire's phenomenal effects increase greatly, and its motivational powers increase substantially as well. (471)

Susie's desire for Robert grows much more motivationally powerful—in Hume's words, more violent—when she's presented with more vivid visual, auditory, and

tactile representations of his charming features. Meanwhile, the motivational force of Susie's desire to keep her money rather than having it repeatedly stolen and spent on cocaine stays constant. If intentions are desire-belief pairs poised to cause action, the rising motivational strength of a new desire-belief pair due to vivid images can change one's intentions even without new information.[29]

Some desires' strengths change differently than others under different conditions, explaining why some intentions are more stable than others. Some desires, like the desire for food that is part of hunger, rapidly change in their intrinsic strength because of internal biological conditions like the fullness of our bellies. Some, like Susie's desire for Robert, change in motivational potency because their objects are closely associated with vivid images that we occasionally encounter. Some, like Nancy Pelosi's desire for America to have universal health care, don't vary with internal states or connect so tightly with particular vivid images. Stable intentions are made of stable desires. A hungry man's intention to eat three pies at the pie-eating contest may be reconsidered without new information as the first pie fills his belly, weakening his desire. Susie's newfound intention to stay with Robert may change without new information once they're apart and her desire for him loses its violence. Pelosi's intention to establish universal health care remained stable under a variety of demoralizing circumstances because her desire was stable. The desire-belief view explains why intentions made of different desires differ in their stability. It's unclear how Bratman's view can do so.

Bratman's feature (c) is that intentions act as inputs to deliberation. He writes,

> I will frequently reason from intended end to intended means or preliminary steps: as when I reason from my intention to go to Tanner to intentions concerning how to get there. And I will frequently reason from more general to more specific intentions: as when I reason from an intention to take a bus to Tanner, and my reflections on the bus schedule, to an intention to take a particular bus.

Here the desire-belief theorist need only note that desires themselves are inputs to deliberation. As Hume says, desire causes us to cast our view on every side of its object, comprehending whatever else is associated with it by cause and effect. This is what happens when we deliberate about various ways of filling out our intentions. The intention to go to Tanner is made of a desire to go to Tanner, so it focuses the agent's attention on the causes and effects of going to Tanner. Riding an appropriately timed bus is an available cause, so the agent's attention focuses on this option. The desire to go to Tanner then combines with the means-end belief that riding the bus raises the probability of going to Tanner, forming an intention to ride the bus.

As we deliberate, desires direct our attention, pointing out potential means and combining with means-end beliefs to form new intentions. If having an intention is having an appropriately related desire and belief, intentions will be inputs to deliberation. Desires are inputs, and intentions are inputs because they're made of desires.

5. Better Explanations

The desire-belief view explains why intentional action almost always involves belief about doing and less often involves reason-choosing, why nonreconsideration and exclusion usually hold but sometimes fail, and why intentions control conduct and provide inputs to further deliberation. I'll conclude by calling attention to an attractive feature of the desire-belief view appearing in these explanations—its simplicity.

The desire-belief view of intending offers sufficient conditions for intending in terms of desire and belief, and builds its explanations using their properties (as well as the properties of other mental processes like imagination and attention-direction). It doesn't appeal to further unexplained features of intention in explaining actual phenomena. This is an advantage. While its opponents don't make use of desire's ability to direct attention or the way a vivid representation of its object increases its violence, they'd be foolish to deny that desire has these properties. The desire-belief view's explanations thus use conceptual resources that both sides must admit, while leaving out any unexplained properties of intention. Thus it allows us to simplify our total psychological theory.

Simple theories that don't explain the data should be rejected. Previous desire-belief theorists didn't try to explain a wide range of psychological phenomena, leading many philosophers to seek more powerful explanations from more complex theories. Appreciating the features of desire allows the desire-belief account to simply and accurately explain a wide range of phenomena surrounding intentional action. I hope this will persuade those who gave up on the desire-belief account to consider it afresh.[30]

Notes

[1] Kieran Setiya, "Practical Knowledge," *Ethics* 118: 388–409, 2008.

[2] Kieran Setiya, *Reasons Without Rationalism*, Princeton NJ: Princeton University Press, 2007.

[3] Michael Bratman, *Intention, Plans, and Practical Reason*, Cambridge MA: Harvard University Press, 1987.

[4] Donald Davidson, "Actions, Reasons, Causes," *Journal of Philosophy* 60: 685–99, 1963. Robert Audi, "Intending," *Journal of Philosophy* 70: 387–403, 1973.

[5] The relation between φing intentionally and intention/intending is too complex to treat here. See, for example, Joshua Knobe, "Intentional Action and Side Effects in Ordinary Language," *Analysis* 63, 2003, 190–193.

[6] John Searle, *Intentionality: An Essay in the Philosophy of Mind*, New York: Cambridge University Press, 1983.

[7] People have traditionally used Davidson's carbon-copy case to show this. My case involves a lower probability of success, making the point more forcefully.

[8] (3) assumes a solution to the problem of deviant causal chains, which afflicts all causal theories. As my focus is on the mental states involved in action and not whether their role is causal, I won't discuss this problem here. Those who despair of solving this problem are welcome to replace causation with whatever relation they think intentions bear to action.

[9] It would be good if the conditional in 3 could be rephrased as a dispositional claim to avoid the conditional fallacy. However, I don't know how to do this while guaranteeing the same consequences in cases where the conditional fallacy doesn't arise.

[10] On my view it's a law of human psychology, but not of the psychology of all possible creatures, that if a mental state is an intention, it's composed this way. It's a psychological law about all possible creatures that if a mental state is composed this way, it's an intention.

[11] J. David Velleman, *The Possibility of Practical Reason*, Oxford: Oxford University Press, 2000.

[12] G. E. M. Anscombe, *Intention*, Cambridge MA: Harvard University Press, 2000.

[13] Making belief about doing part of intention risks putting dubious processes of belief-formation that resemble wishful thinking into intention-formation. See Rae Langton, 2004, "Intention as Faith," in John Hyman and Helen Steward (eds.) *Action and Agency*, Cambridge: Cambridge University Press, 243–58.

[14] Paul Grice, "Intention and Uncertainty," *Proceedings of the British Academy* 57: 263–79, 1971. There are many ways to set up an inferential view, as there are many sets of premises from which one might plausibly infer what one is doing. Whatever one takes the premises to be, the ability of desire to drive reasoning can explain why we put them together, as I'll describe.

[15] David Hume, *Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1888. This feature of desire is also noted by T. M. Scanlon, *What We Owe To Each Other*, Cambridge MA: Harvard University Press, 1998.

[16] If this view is correct, experienced agents (perhaps including all humans old enough to talk) will be so used to this process that they won't have any experience of drawing interesting inferences that they'll act. Instead, the process will in normal cases run automatically, as happens with language processing in experienced speakers.

[17] Sarah Paul, "How We Know What We're Doing," *Philosophers' Imprint* 9:4, 1–29, 2009. Paul follows Grice in developing an inferentialist view.

[18] In her exchange with Setiya in *Ethics*, Paul offers a counterexample to this effect. Someone whose Alien Hand Syndrome is about to make him unintentionally button up his shirt does so intentionally instead with a lower probability of success, because his lack of control infuriates him. Setiya responds that as he buttons his shirt, he isn't less confident that he's buttoning up his shirt, even if he might have a lower probability of success. Setiya's response can only work by pointing to the satisfaction of the first disjunct in the consequent of his principle—the buttoner believes that he's buttoning. If the buttoner is equally confident that he's engaged in buttoning with or without the intention, as may happen, the second disjunct won't be satisfied. While it helps against this counterexample, having both disjuncts seems redundant. Setiya should've noted that in this case, the buttoner intentionally buttons his shirt intentionally. (We don't usually act this way, but here part of the buttoner's goal is to avoid buttoning unintentionally.) If he buttons by buttoning intentionally, and is more confident that he's intentionally buttoning than he'd otherwise be, he buttons by doing something else for which the condition is satisfied. In the upcoming case, Bridget doesn't intentionally bounce intentionally or believe that she's bouncing, so neither response is available. See Sarah Paul, "Intention, Belief, and Wishful Thinking: Setiya on 'Practical Knowledge,'" *Ethics* 119, 546–557, 2009; and Kieran Setiya, "Practical Knowledge Revisited," *Ethics* 120, 128–137, 2009.

[19] Could Bridget by deliberation succeed in keeping her bouncing wholly unintentional? I think not. Even an intentional act of aborting deliberation before it produced any motivational force causing bouncing would be performed out of a causally effective desire to bounce, and thus involve an intention to bounce, though one that achieves its object in a strange way. If you disagree, you can still get into the example—just imagine how her terror might make her fail at the delicate task of keeping her bouncing unintentional.

[20] Setiya returns to this response in his exchange with Sarah Paul.

[21] John Searle, *Rationality in Action*, Cambridge, MA: MIT Press, 2001.

[22] Christine Korsgaard, *Locke Lectures*.

[23] There is disagreement about whether the reasons are the mental states themselves, the contents of the mental states, or something else tied to the mental states. The problem arises in any case.

[24] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68, 5–20, 1971. Someone who accepts both Frankfurt's account of free will and my account of reason-choosing will see all attempts at reason-choosing at the doings of persons, with successful attempts being free and unsuccessful attempts being unfree. Of course, I remain neutral as to whether Frankfurt's view of free will is correct. I invoke second-order volitions only to explain how and when it's psychologically

possible to make a certain kind of choice, leaving it open whether such choices are free, or whether reason-choosing or second-order volitions have any interesting connection to free will. For criticisms of Frankfurt's account of free will, see Gary Watson, "Free Action and Free Will," *Mind* 96, 145–72, 1987.

[25] Sinhababu characterizes the "Motivational Aspect" of desire as follows: "If agents occurrently desire D, and they occurrently believe that they can bring about D by doing A, they will be motivated to do A. The strength of their motivation will increase with the strength of the desire and the subjective probability that they can bring about D by doing A. If at any time there is some action that they are the most motivated to do, they will initiate that action." Neil Sinhababu, "The Humean Theory of Motivation Reformulated and Defended," *Philosophical Review* 118:4, 465–500, 2009.

[26] Michael Ridge, "Humean Intentions," *American Philosophical Quarterly*, 157–178, 1998.

[27] Claiming that the desire not to deliberate more is unconscious would better fit the phenomenology, but doing so would be *ad hoc* without an explanation of why it's unconscious. While desire-belief pairs that we've incorporated into a routine sometimes lack a robust phenomenology of conscious deliberation, different means for ending deliberation might be appropriate in different circumstances, preventing any particular means from becoming routine enough to mute the phenomenology.

[28] Cases like this make Bratman's description of exclusion and Velleman's claim that forming an intention "settles the practical question, leaving us nothing further to deliberate" (111) seem too strong. David Velleman, *Practical Reflection*, Princeton: Princeton University Press, 1999.

[29] This is an example of "judgment shift" (97) as discussed in Richard Holton, *Willing, Wanting, Waiting*, Oxford: Oxford University Press, 2009.

[30] This paper was improved by comments at the National University of Singapore, King's College London, Tufts University, the University of Illinois, the University of Chicago Practical Philosophy Workshop, Illinois State University, Georgetown University, Wayne State University, the University of Missouri at St. Louis, the Princeton Workshop in Normative Philosophy, the University of Maryland, the University of Delaware, and Florida State University. An anonymous referee for this journal helped me make the arguments clearer and stronger. Conversations with Brian Leiter and Errol Lord made my views bolder and better. My NUS colleagues, especially Ben Blumson and Weng Hong Tang, were always helpful. A conversation with Warren Tusk and Elisabeth Cohen improved my account of conditional intention. Kieran Setiya and Michael Bratman kindly helped me understand their views and pointed me to useful literature. John Maier's understanding of the intention debate explains the structure of this paper, and his encouragement explains its existence.