Bratman on belief, intention, and intentional action

Jeff Speaks

January 27, 2009

1 Causal theories and the assumption of tight fit

So far, we have discussed two causal theories of intentional action.

The first was Davidson's early theory, and analyzed intentional action via causation by belief/desire pairs:

> Belief-desire theory
>
> A ϕs intentionally $\equiv_{df}$ (for some property F, (i) A has some pro-attitude toward actions which are F, (ii) A believes that her ϕing is F, and (iii) this belief and desire cause A in the right way to ϕ)

The second was suggested by Davidson's paper "Intending," in which pure intentions replace belief/desire pairs:

> Simple intention theory
>
> A intentionally ϕs $\equiv_{df}$ A's ϕing is caused by A's intention to ϕ.

As we saw, this sort of view might be accompanied by an analysis of pure intentions, such as Davidson's view of intentions as all-out evaluative judgements. If we add that to the simple intention theory above, we'd get something like this:

<u>Evaluative judgement theory</u>

A intentionally ϕs $\equiv_{df}$ A's ϕing is caused by A's judgement that ϕing is, all things considered, the best thing to do.

Each of these theories involves something like what Bratman calls "the assumption of tight fit": the idea that what makes ϕing an intentional action must be some mental state directed specifically toward the act of ϕing:

Assumption of *tight fit*	If intentionally ϕing must involve a distinctive pro-attitude, it must be a pro-attitude specifically in favor of ϕing.

You can think of Bratman's paper as having two goals: to argue against causal theories which involve this assumption, and to show how to begin to sketch a causal theory which does not involve this assumption.

2 Bratman's arguments against the assumption of tight fit

2.1 Intentions and intentional action

In the context of a simple intention theory, the assumption of tight fit entails the following claim about the relationship between intentional actions and intentions to act:

The Simple View	If A intentionally ϕs, then A intends to ϕ

Bratman's strategy is to argue against the simple view, and replace it with the following weaker view:

The Single *Phenomenon View*	Intentional action and intentions to act both involve a common state, which is what distinguishes intentional from unintentional action

2.2 Intentions, belief, and rationality

Bratman's argumentative strategy is to first argue that there are certain important connections between intention and belief, and then to argue that if these connections obtain, the simple view must be false.

There are three important principles about intention, belief, and rationality worth keeping in mind:

Intention-Belief *Principle*	If A intends to ϕ, then A believes that she will ϕ.

This is a very strong instance of the general idea that there is some kind of belief constraint on intentions to act. Bratman formulates two weaker versions of this kind of belief constraint, which also capture the idea that having a certain intention to act entails something about the status of one's beliefs. He advocates both of the following:

Strong Consistency Principle	If A intends to ϕ, then, if A is not criticizably irrational, this intention must be able to be put together with the rest of my intentions into a plan which is consistent with my beliefs.
Weak Consistency Principle	If A intends to ϕ, then, if A is not criticizably irrational, this intention must be consistent with the rest of my intentions.

It is important to note that these Principles do not entail the impossibility of certain kinds of belief/intention conflicts. Rather, they entail that in the case of such conflicts, the agent will be criticizably irrational.

Let's consider some of the arguments for and against this Intention-Belief:

Arguments for Intention-Belief

1. *Argument from Strong Consistency.* One crucial difference between desire and intention is that I am 'rationally criticizable' if I have two intentions which are inconsistent with my beliefs, and hence which I cannot jointly execute (given the truth of my beliefs). Desires do not have this property; as Bratman notes, there is nothing criticizable in my wanting two incompatible things. It is natural to try to explain this difference in terms of the claim that intention, but not desire, involves belief, and that it having intentions which are not jointly satisfiable is criticizable for just the same reasons that having inconsistent beliefs is criticizable.

2. *Argument from oddness of cases of avowing intention but not belief.* Grice's example (from "Intention and Certainty"): A: 'I intend to go to the party tomorrow night.' B: 'Well, then I'll see you there.' A: 'I may not go.' These are explained by the claim that having an intention entails something about one's beliefs, since (to use the Intention-Belief Principle) the same kind of oddness results from the dialogue: A: 'I believe I will go to the party tomorrow night.' B: 'Well, then I'll see you there.' A: 'I may not go.'

3. *Argument from the role of intentions in means-end reasoning.* Once we form the intention to ϕ, we often make further plans for action using the fact that we will ϕ as a premise in deciding what to do.

Arguments against Intention-Belief

1. *Intuitive counterexamples.* Bratman himself offers an argument against the strongest belief constraint. He considers a case in which an agent has the intention to stop by the store on the way home, but does not believe that he will (or will not), knowing his own absent-mindedness. This is similar to Davidson's case in "Intending" of the man who intends to provide for his children via his will. In both cases it seems plausible to say that the agent lacks the belief that would be entailed out by Intention-Belief.

2. *The analogy with hope.* It is clearly not true that hope involves belief, in the sense that hoping for p entails that one also believe p. But, as Harman points out in 'Practical Reasoning', it is striking that it is irrational to hope for both p and q if these are inconsistent (p. 437). This fact cannnot be explained in terms of the fact that hope implies belief; so why should we think that parallel

3

facts about Weak Consistency in the case of intention should be explained in terms of intention implying belief? Two replies to this point: (i) This does not remove the force of the point about conversational weirdness in saying that you intend to ϕ but do not believe that you will; no parallel effect happens with claims about what one hopes. (ii) It is important to note that only the Weak Consistency Constraint plausibly applies to hopes, not the Strong. It may be irrational to hope for two things which are inconsistent, but it is not irrational to hope for something which you believe will not happen. This is a contrast with intention where, as Bratman argues, the Strong Consistency Constraint does have some plausibility.

The two Consistency principles might at first seem to be on much stronger ground. But these also face a pair of apparent counterexamples which have received a lot of discussion in the literature on belief:

1. Contradictory beliefs which the subject does not know to be contradictory. Puzzling Pierre and catsup/ketchup.

2. The paradox of the preface.

We will return to the question of how important these sorts of counterexamples might be to Bratman's argument.

2.3   The video game example

Bratman's main initial argument against the Simple View is based on his example involving the two video games. Imagine that A is simultaneously playing two video games, one with each hand. In each game, the goal is to shoot a certain target. His aim (we can imagine that he has a bet on the game) is just to win one of the games. But the two video games are linked in such a way that the agent's shooting both of the targets will shut both games down, causing the agent to lose both games. Nonetheless, A knows that hitting the targets is hard, and figures that his best bet is to try to hit both targets. Now suppose that A hits target 1. It seems that he does this intentionally.

But did he intend to hit target 1? Not if the Strong Consistency Principle is correct. For, by parity of reasoning, A intended to hit target 1 iff he intended to hit target 2. But he knows that these are incompatible goals. Further, A does not seem to be criticizably irrational. Hence, despite the fact that A intentionally hit target 1, he did not intend to hit target 1, and the Simple View (in both its forms) is false. The argument can be summarized like this:

1. A intentionally hit target 1.
2. A intended to hit target 1 iff A intended to hit target 2.
3. The Strong Consistency Principle.
4. $\neg$ (A intended to hit target 1) (2,3)
C. $\neg$ (The Simple View) (1,4)

4

The argument could also be recast using Intention-Belief rather than Strong Consistency.

Bratman considers a few objections to his argument. The most plausible seems to be the one that claims that the Strong Consistency Principle is a prima facie principle about rationality which may be overridden when the agent's purposes are best served by his having contradictory intentions. Bratman replies that the agent's purposes do not require him to *intend* to hit target 1 and target 2, but only that he *try* to hit each target.

But I think that there is a more important objection to this argument. It is tempting to say that an agent is criticizably irrational if she has contradictory beliefs which she knows to be contradictory. But the paradox of the preface casts some doubt on this claim. In that case, an agent's aim of maximizing true beliefs comes into conflict with the maxim that one should not believe contradictory propositions, and it is plausible to think that the former overrides the latter. Why should we not regard Bratman's example in just this way?

2.4 The distinction between acts an agent intends to do and acts an agent does in the course of fulfilling an intention

Most discussion of Bratman's case against the Simple View has focused on cases like the video game example. But some points he makes later in the paper seem to me to make a more convincing case against the Simple View.

Bratman argues (see on this point also Harman, 'Practical Reasoning') that one of the key roles played by intentions is as premises in practical reasoning. If I settle on the intention to ϕ, I will plan further actions around the fact that I am going to ϕ. But now consider acts I perform in the course of performing other actions, such as, in Bratman's example, my wearing down my sneakers when running a marathon. I knowingly do this, and it is not implausible to say that I do it intentionally. But I do not plan around this act in quite the same way I plan around acts I intend to do, like running the marathon. I may, as Bratman notes, engaged in reasoning aimed at finding a way to run the marathon. But I do not engage in reasoning aimed at finding a way to wear down my sneakers (399-400).

This does seem like an important intuitive distinction, and it is natural to express it by saying that I don't have to intend to do the things which are necessary parts of my carrying out a plan to execute some intention.

3   Bratman's replacement view

The Simple View is not without intuitive appeal. It is the burden of the last part of Bratman's paper to show that we can give an intuitively satisfying account of the relationship between intentions to act and intentional action despite rejecting the Simple View.

This is the point of the Single Phenomenon View. Bratman's idea is that there can be a single state of intention involved in both intentions to act and intentional action, despite the fact that not every act of intentionally ϕing is accompanied by an intention to ϕ. The way to formulate such a view is to reject the assumption of tight fit.

Conjoining the Single Phenomenon View with the rejection of this assumption gives us the view that every act of intentionally ϕing is to be explained in terms of some intention, but that this intention may not be the intention to ϕ.

We can make this more concrete by following Bratman in defining the useful notion of *motivational potential*. Intuitively, ϕing will be within the motivational potential of ψing (for an agent at a time) iff it is possible for the agent to ϕ in the course of executing his intention to ψ. So the idea is that an act can be intentional if either caused in the right way by an intention to perform that act, or done as part of an agent's execution of an intention to perform some other act.

It is important to note that this does not give us sufficient conditions for acting intentionally, even if it does give us necessary conditions. Some 'sub-intentional' acts can be done as part of the execution of an intention without being done intentionally. Given this, Bratman's way of laying out the issues suggests a causal theory of action of the following rough form:

> A intentionally ϕs $\equiv_{df}$ $\exists \psi$ (i) A intends to ψ, (ii) this intention causes A's ϕing (in the right way), (iii) A's ϕing is in the motivational potential of his ψing, and (iv) _____.

where (iv) spells out some condition to rule out non-intentional components of intentional actions.

How could (iv) be spelled out? Could we add a condition requiring something of the agent's beliefs?

4   Are intention-based analyses of intentional action circular?

Wilson (1989) and Ginet (1990) have suggested that intentions to ϕ are always intentions to ϕ intentionally. If so, this suggests that analyses, like Bratman's above analysis and the strong version of the simple view, are circular. The following is one way of laying out an argument like this:

1. A intends to ϕ $\equiv$ A intends to ϕ intentionally.
2. So if we analyze intentional action in terms of (appropriate causation by) intentions to act, we will thereby have covertly analyzed intentional action in terms of intentions to perform intentional actions.
3. Analyzing intentional action in terms of intentions to act intentionally is objectionably circular.

Each of the moves in this argument can be questioned:

Premise 1. The intuitive argument for premise 1 is not unconvincing: if I intend to call you on the phone, and then accidentally hit the numbers of your phone number while the phone is in my pocket, it is natural to say that the intention has not been fulfilled. This does suggest that intentions to act are only fulfilled by intentional actions, which suggests

that intentional action must be a part of the content of the intentions. But perhaps this only holds of some cases. Mele (1999), e.g., suggests that it may be possible for an agent who takes a bet that he can do something unintentionally to intend to win the bet, by intending to perform the act in question without intending to perform it intentionally.

Premise 2. It does not follow from the truth of the biconditional that an analysis of intentional action in terms of intentions is an analysis of intentional action in terms of intentions to act intentionally. In effect, this transition assumes not only the truth of the biconditional, but a certain kind of explanation of its truth. On this explanation, the truth of (1) is explained by the fact that ascriptions of intentions to act are elliptical versions of ascriptions of intentions to act intentionally. But it could be that there is some other explanation of the biconditional. E.g., it is true that the belief that there is water in my glass is only true if there is H2O in my glass; but it does not follow that the ascription of this belief to me is an elliptical ascription of a belief about hydrogen and oxygen.

Premise 3. The argument does not show that the analysis of intentional action in terms of intention is strictly circular, since in the analysans the concept of intentional action only occurs within the propositional attitude of the agent. There may be no obvious circularity in analyzing, e.g., evaluative properties in terms of beliefs involving those evaluative properties; often such analyses are proposed by theorists who wish to treat some property as 'response-dependent.' But even so, this does not on the face of it seem like a very plausible view about intentional action. It is unlikely that intentional action should be analyzed in terms of intentions concerning intentional action.

References

Carl Ginet, 1990. *On Action.* Cambridge: Cambridge University Press.

Alfred Mele, 1999. Is There a Place for Intention in an Analysis of Intentional Action? *Philosophia* .

George Wilson, 1989. *The Intentionality of Human Action.* Stanford, CA: Stanford University Press.