

# Belief, Intention, Intentional Action, and the Simple View

Jeff Speaks

October 11, 2004

1	Linking principles . . . . .	1
1.1	The reducibility of intention to belief and desire . . . . .	2
1.2	Belief constraints on intentions to act . . . . .	3
1.3	Belief constraints on intentional action . . . . .	5
1.4	Intentions and intentional action . . . . .	5
2	Bratman's arguments against the simple view . . . . .	5
2.1	The video game example . . . . .	5
2.2	The distinction between acts an agent intends to do and acts an agent does in the course of fulfilling an intention . . . . .	6
3	Bratman's replacement view . . . . .	7
4	Are intention-based analyses of intentional action circular? . . . . .	7

Bratman's "Two Faces of Intention" raises at least three issues that have already come up in this course:

- The relation between intentions to act and belief and desire.
- The relation between intentions to act and intentional action.
- The relation between belief and intentional action.

Bratman's article presents some convincing arguments that some natural combinations of views on these three topics must be rejected; our aim will be to see what the most plausible theory which survives his discussion is.

## 1 Linking principles

But first we will begin with an outline of some different available views on the relationships between belief, intention, and intentional action. It is worth noting that any true principle like the following that we can find will play a role in testing a theory of intentional action.

After all, we want a theory of the nature of intentional action not only to single out the right class of actions, but also to explain necessary truths involving intentional action. Inasmuch as any theory of intentional action will have to say something about the relations between intentional action, intention, and belief, the following proposed necessary truths will, if true, need some explanation.

### 1.1 *The reducibility of intention to belief and desire*

*Reducibility of intention*     $A$  intends to  $\phi \equiv_{df}$   $A$  has some cluster of beliefs and desires which bear some relation  $R$  (to be spelled out by the theory) to  $\phi$ ing

#### Arguments for the reducibility thesis:

1. *Argument from the priority of belief and desire.* Many philosophers of mind find it attractive to think of the mind as a kind of information processing machine, which receives information as input from the world, and causes actions as output. On this picture, it is very natural to think of belief and desire as the two most fundamental kinds of mental states: beliefs register information received as input from the world, and desires are those states which, when added to beliefs, yield certain behavioral outputs. On this view, it is natural to try to explain other states as certain kinds of combinations of, or constructions out of, belief and desire. So on this view, it will be natural to think of intentions to act as reducible to constellations of beliefs and desires.

2. *Reply to objections.* Arguments against the reducibility thesis typically turn on differences between intention and belief and intention and desire. But it is not obvious how such objections generalize to views which take intentions to be certain, perhaps complex, combinations of beliefs and desires.

#### Arguments against the reducibility thesis:

1. *Davidson on reasons for intending.* Davidson (in “Intending”) points out that the reasons for intending to  $\phi$  seem different in kind than reasons for believing that one will  $\phi$ . But they also seem different from reasons for desiring; in many cases we might say that we have no reason for desiring something, but that we just do; this reply is less apt in the case of intention.

2. *No version seems to work.* Harman, e.g., objects to the reducibility thesis that “no plausible analysis of this sort has ever been suggested, and it is, therefore, much more reasonable to take intentions to be a real part of the causal and explanatory order” (441). See our discussion of Davidson’s “Intending” for arguments against some of the most obvious candidates. Of course, these kinds of case-by-case arguments do not show that we haven’t overlooked some unobvious possibility.

3. *Reply to functionalist objection.* It is wrong to link functionalist views of the mind with the idea that belief and desire are the most fundamental mental states. Indeed,

## 1.2 Belief constraints on intentions to act

Many have found the following belief-intention principle plausible:

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

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  $\phi$ , 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  $\phi$ , 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 *either* the absence of such conflicts, or the 'criticizable irrationality' of the agent in question.

### Arguments for their being some belief constraint on intention

1. *Argument from rational criticizability.* 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  $\phi$ , we often make further plans for action using the fact that we will  $\phi$  as a premise in deciding what to do.

### Arguments against belief constraints on intentions to act

1. *The Intention-Belief principle is too strong.* 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 has the intention that would be ruled out by Intention-Belief (the intention to stop at the store, and the intention to provide for his children).

2. *Contradictory beliefs and rational criticizability.* Bratman’s consistency principles do not say that it is impossible to have contradictory intentions, or intentions which contradict one’s own beliefs; they say that it is impossible to have such intentions (or intentions and beliefs) without being rationally criticizable. One problem with this is that it seems possible in at least some cases to have contradictory beliefs without being rationally criticizable. In some cases, like ‘puzzling Pierre,’ this is because one does not realize that one’s beliefs are contradictory. In other cases, like the paradox of the preface, one does realize that the belief are contradictory, but is (arguably) still not rationally criticizable. But this means that the inference from (1) the claim that *A* is not rationally criticizable to (2) *A* does not have contradictory beliefs is a very questionable one. If so, then one worries that the inference from (1) *A* is not rationally criticizable to (2\*) *A*’s beliefs and intentions are not contradictory should be, if anything, even less plausible. This is enough to call into question both the Strong and Weak Consistency Principles.

3. *Futile actions.* Anna and Noah raised in seminar the possibility of futile actions: cases in which one does some action even though one knows (believes) that the action will not have its intended effect. Two examples: (i) Intending in attending a demonstration to influence public opinion, even though one believes that demonstrations never really influence public opinion. (ii) Intending to hit one’s sparring partner in karate exercises in which one’s partner knows that the punch is coming, and so will certainly get out of the way. In (i), it seems plausible to me to say that to the extent that you believe that your attending the demonstration will have no effect on public opinion, you do not really intend to affect public opinion. In (ii), I wonder whether there is a kind of pretense involved. In one frame of mind, you may believe that your partner will get out of the way, and not really intend to hit him. But in actually performing the exercise you may accept the fiction that this is a real sparring match, and intend to hit your partner; but there is an intuitive sense in which engaging in this pretense involves giving up the belief that your partner knows what’s coming and will get out of the way of the punch.

4. *Hoping.* 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 cannot be explained in terms of the fact that hope implies belief; so why should we think that parallel facts about 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  $\phi$  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.

### 1.3 Belief constraints on intentional action

*Action-Belief Principle* If  $A$   $\phi$ s intentionally, then  $A$  acts in the knowledge that he is  $\phi$ ing

*Weak Action-Belief Principle* If  $A$   $\phi$ s intentionally, then there is some intentional act  $\psi$  (which may be identical to  $\phi$ ing) such that  $A$   $\phi$ s in the knowledge that he is  $\psi$ ing

### 1.4 Intentions and intentional action

*The Simple View (Weak)* If  $A$  intentionally  $\phi$ s, then  $A$  intends to  $\phi$

*The Simple View (Strong)*  $A$  intentionally  $\phi$ s  $\equiv_{df}$   $A$ 's  $\phi$ ing is caused (in the right way) by  $A$ 's intention to  $\phi$

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

(The Simple View (Weak) = The Single Phenomenon View + the view that this common state is the intention to act.)

## 2 Bratman's arguments against the simple view

### 2.1 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.
- 
- C1.  $\neg (A \text{ intended to hit target 1})$  (2,3)  
C2.  $\neg (\text{The Simple View})$  (1,C1)

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.

Three points about this: (i) The objection does have some intuitive plausibility, partly because of a parallel with cases of contradictory beliefs. 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? (ii) Someone who thinks that intentionally  $\phi$ ing requires the intention to  $\phi$  is also likely to think that trying to  $\phi$  requires the intention to  $\phi$ . So the claim that the agent need not intend to hit target 1, but only to try to hit target 1, may not have all that much force. (iii) Some philosophers (e.g. Adams) think that there are belief constraints not only on intention, but also on trying. If these constraints were strong enough, this might make the move from describing the case in terms of intending to describing it in terms of trying less obviously helpful. But it does not seem plausible that any principle as strong as the Strong Consistency Principle could apply to trying.

## 2.2 *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  $\phi$ , I will plan further actions around the fact that I am going to  $\phi$ . 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  $\phi$ ing is accompanied by an intention to  $\phi$ . The way to formulate such a view is to reject what he calls the 'assumption of tight fit':

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

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

We can make this more concrete by following Bratman in defining the useful notion of *motivational potential*. Intuitively,  $\phi$ ing will be within the motivational potential of  $\psi$ ing (for an agent at a time) iff it is possible for the agent to  $\phi$  in the course of executing his intention to  $\psi$ . 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  $\phi$ s  $\equiv_{df}$   $\exists \psi$  (i)  $A$  intends to  $\psi$ , (ii) this intention causes  $A$ 's  $\phi$ ing (in the right way), (iii)  $A$ 's  $\phi$ ing is in the motivational potential of his  $\psi$ ing, and (iv) \_\_\_\_\_.

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

### 4 Are intention-based analyses of intentional action circular?

Recently, Wilson, *The Intentionality of Human Action* and Ginet, *On Action* have suggested that intentions to  $\phi$  are always intentions to  $\phi$  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  $\phi \equiv A$  intends to  $\phi$  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. This 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, ‘Is there a place?’, 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 H<sub>2</sub>O 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 analysis 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

- Adams, Frederick, ‘Trying: You’ve Got to Believe’, *Journal of Philosophical Research*, (1995), pp. 549–561.
- Bratman, Michael, ‘Two Faces of Intention’, *Philosophical Review*, 99 (1984), pp. 375–405.
- Davidson, Donald, ‘Intending’, (1978), pp. 83–102.
- Ginet, Carl, *On Action*, (Cambridge: Cambridge University Press, 1990).
- Harman, Gilbert, ‘Practical Reasoning’, *Review of Metaphysics*, 79 (1976), pp. 431–463.
- Mele, Alfred, ‘Is There a Place for Intention in an Analysis of Intentional Action?’ *Philosophia* (1999).



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